

UNIVERSIDADE PAULISTA
PROGRAMA DE DOUTORADO EM ENGENHARIA DE PRODUÇÃO

**MINERAÇÃO DE DADOS PARA DESCOBERTA DE FATORES PREDITIVOS DE
SUCESSO EM AVALIAÇÕES DE LARGA ESCALA DOS ALUNOS DO ENSINO
MÉDIO.**

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Paulista - UNIP, para obtenção do título de Doutor em Engenharia de Produção.

IVONALDO VICENTE DA SILVA

São Paulo

2021

UNIVERSIDADE PAULISTA
PROGRAMA DE DOUTORADO EM ENGENHARIA DE PRODUÇÃO

**MINERAÇÃO DE DADOS PARA DESCOBERTA DE FATORES
PREDITIVOS DE SUCESSO EM AVALIAÇÕES DE LARGA ESCALA
DOS ALUNOS DO ENSINO MÉDIO.**

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Paulista - UNIP, para obtenção do título de Doutor em Engenharia de Produção.

Área de concentração: Gestão de Sistemas de Operação

Linha de pesquisa: Redes de empresas e planejamento da produção

Projeto de pesquisa: Gestão de operações de serviços - formas organizacionais, métodos e ferramentas para a gestão.

IVONALDO VICENTE DA SILVA

São Paulo

2021

IVONALDO VICENTE DA SILVA

**MINERAÇÃO DE DADOS PARA DESCOBERTA DE FATORES
PREDITIVOS DE SUCESSO EM AVALIAÇÕES DE LARGA ESCALA
DOS ALUNOS DO ENSINO MÉDIO.**

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Paulista - UNIP, para obtenção do título de Doutor em Engenharia de Produção.

Aprovado em: ____/____/____

Banca Examinadora:

Dra. Márcia Terra da Silva

(Orientadora)

Universidade Paulista - UNIP

Prof. Dr. Pedro Luiz Oliveira Costa Neto

Universidade Paulista - UNIP

Prof. Dr. Rodrigo Franco Gonçalves

Universidade Paulista - UNIP

Prof. Dr. Luis Naito Mendes Bezerra

Universidade Cruzeiro do Sul - UNICSUL

Prof. Dr. Mauro Zilbovicius

Universidade de São Paulo - USP

DEDICATÓRIA

À minha esposa, Elaine Caprioli Silva, pela paciência, motivação e companheirismo, e por todos os momentos que estive ao meu lado, contribuindo para que eu pudesse superar todos os obstáculos e para que minha fé não se abalasse.

AGRADECIMENTOS

Em uma jornada acadêmica, muitos são os desafios a serem enfrentados, e esses só podem ser superados com o apoio e compreensão de todos aqueles que direta ou indiretamente caminham juntos na empreitada e torcem para que o sucesso seja alcançado. Nessa empreitada tive muito apoio e colaboração, e muito a agradecer.

Inicialmente agradeço a Deus pela força, saúde e paz a mim concedidas, pois sem o equilíbrio espiritual não seria possível finalizar esse desafio.

Aos meus pais, Ivanilda Maria da Conceição Silva e Cícero Vicente da Silva (*in memoriam*), por todos os esforços dispendidos para que eu me tornasse uma pessoa que preza pela estrutura familiar, honestidade e compromisso profissional.

Aos meus patrões, Sr. José Deuzimar Dantas e Sra. Deusa Maria Dantas, por todo incentivo e colaboração.

Ao corpo docente do curso Doutorado em Engenharia da Produção na figura dos professores doutores, Irenilza de Alencar Nääs, João Gilberto Mendes dos Reis, José Benedito Sacomano, Oduvaldo Vendrametto, Pedro Luiz de Oliveira Costa Neto, Rodrigo Franco Gonçalves e Silvia Helena Bonilla.

Um agradecimento super especial à minha orientadora, Profa. Dra. Márcia Terra da Silva, pela paciência, dedicação, companheirismo e direcionamento para que o sucesso dessa empreitada fosse alcançado. Sem o vosso empenho e dedicação, de certo que eu não teria conseguido superar esse desafio.

Quero agradecer também aos funcionários da UNIP, em especial à secretária Márcia Nunes por toda ajuda e atenção dispendida na resolução de assuntos relacionados ao curso.

Agradeço à UNIP pelo privilégio da bolsa concedida, para realização do curso.

A todos os amigos que conquistei durante o curso, que com certeza estarão sempre presentes, em especial à Saturnina Alves da Silva Martins.

RESUMO

A qualidade da educação vem se tornando tema de destaque em publicações acadêmicas, o que demonstra que esse tema pode ser de grande interesse dos pesquisadores e da sociedade. O governo tem tentado ofertar um ensino de qualidade no Brasil, no entanto os esforços dispendidos não têm alcançado resultados positivos. Para verificar o nível de aprendizado dos alunos são utilizadas as avaliações em larga escala do Sistema de Avaliação da Educação Básica-SAEB. Os resultados da edição do SAEB do ano de 2017, demonstraram que os desempenhos dos alunos estão fracos, visto que mais de 90% ficaram abaixo do nível médio de desempenho, Nível 5, em uma escala de 0 a 10. Diante desse cenário, o objetivo dessa tese foi analisar, por meio da Mineração de Dados Educacionais, quais fatores presentes no ambiente educacional puderam ser preditores para que os alunos obtivessem bons desempenhos nas avaliações de larga escala. Os dados coletados diretamente no website do governo foram referentes às edições de 2015 e 2017 do SAEB, para os egressos do Ensino Médio. Nos pacotes de dados coletados, além dos desempenhos dos alunos, também foram coletados os resultados dos formulários direcionados aos Alunos, Professores, Diretores e Representantes de Escola. Após a fase de tratamento de dados, com as bases prontas, foram realizadas as análises por meio das técnicas da Mineração de Dados Educacionais e os resultados encontrados, por meio dos algoritmos de *Árvore de Decisão*, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes*, indicaram que os fatores que estão associados aos melhores desempenhos dos alunos em avaliações do SAEB são: nível de escolaridade dos pais ou responsáveis pelos Alunos, tempo de experiência do Professor no exercício da função, além da atualização em habilidades e competências no exercício da sua função, tempo de experiência do Diretor de Escola no exercício da função, o nível de formação em curso de pós-graduação do Diretor de Escola, e o tipo de ações relacionadas à prática de gestão escolar, e finalmente, já observando a instituição de ensino, o nível de acesso a computadores e internet por alunos e professores e se a escola tem uma internet com boa qualidade de tráfego de dados (banda larga). A descoberta desses fatores pode contribuir para proporcionar debates envolvendo profissionais da educação, no intuito de desenvolver planos de ação que possam propor melhorias no ambiente escolar, inclusive auxiliar nos processos de tomada de decisão pelos gestores escolares.

Palavras-Chave: Desempenho escolar; Avaliação em larga escala; Mineração de dados Educacionais; SAEB.

ABSTRACT

The quality of education has become a prominent topic in academic publications, which shows that this topic can be of great interest to researchers and society. The government has tried to offer quality education in Brazil, however the efforts expended have not achieved positive results. Large-scale assessments of the Basic Education Assessment System-SAEB are used to verify the students' learning level. The results of the 2017 SAEB edition, showed that the students' performances are weak, since more than 90% were below the average level of performance, Level 5, on a scale of 0 to 10. Given this scenario, the objective of this thesis was to analyze, through Educational Data Mining, which factors present in the educational environment could be predictors for students to obtain good performances in large scale assessments. The data collected directly on the government website were for the 2015 and 2017 editions of SAEB, for high school graduates. In the data packages collected, in addition to the students' performances, the results of the forms directed to Students, Teachers, Principals and School Representatives were also collected. After the data treatment phase, with the bases ready, the analyzes were carried out using the Educational Data Mining techniques and the results found, using the Decision Tree, Random Forest, Gradient Boosted Tree and Naive Bayes algorithms, indicated that the factors that are associated with the best performance of students in SAEB assessments are: level of education of parents or guardians of students, length of experience of the teacher in the exercise of the function, in addition to updating skills and competences in the exercise of their function , time of experience of the School Director in the exercise of the function, the level of training in the postgraduate course of the School Director, and the type of actions related to the practice of school management, and finally, already observing the educational institution, the level of access to computers and internet by students and teachers and whether the school has an internet with good data traffic quality (broadband). The discovery of these factors can contribute to provide debates involving education professionals, in order to develop action plans that can propose improvements in the school environment, including assisting the decision-making processes by school managers.

Keywords: School performance; Large-scale evaluation; Educational data mining; SAEB.

LISTA DE FIGURAS

Figura 1. Estrutura do Desenvolvimento da Tese.....	22
Figura 2. Níveis e Modalidades da Educação/Ensino	44
Figura 3. Descoberta de Conhecimento em Bases de Dados - KDD.....	53
Figura 4. Exemplo de Árvore de Decisão.	60
Figura 5. Interface do Software RapidMiner.	62
Figura 6 Sequência dos Artigos que compõem a Tese.	73
Figura 7. Fatores preditores do desempenho escolar	214

LISTA DE QUADROS

Tabela 1. Resultado SAEB 2017 para o Ensino Médio.....	18
Quadro 1. Habilidades e conhecimentos observados referentes aos Níveis 1 e 2 das Escalas de Desempenho.	18
Quadro 2. Pesquisa Bibliométrica relativa às palavras-chave do estudo.....	21
Quadro 3. Relação dos fatores preditores do desempenho escolar.	30
Quadro 4. Cinco países melhores classificados conforme resultados das avaliações do PISA 2018.....	39
Quadro 5. Características do Sistema SAEB a partir de 2019.....	41
Quadro 6. Questões dos formulários SAEB	42
Quadro 7 Órgãos responsáveis pela Educação	43
Quadro 8. Percentual de escolas por nível do indicador de complexidade de gestão e dependência administrativa – Brasil e Regiões Geográficas - 2019	45
Quadro 9 Total de escolas e de alunos matriculados, no Brasil.	46
Quadro 10. Médias estaduais para o Ensino Médio referente à proficiência Língua Portuguesa.	47
Quadro 11. Relação de competências em Língua Portuguesa, para os níveis de 1 a 3 na escala de desempenho	49
Quadro 12. Médias Estaduais para o Ensino Médio referente a proficiência Matemática.	50
Quadro 13. Relação de competências em Matemática, para os níveis de 1 a 3 na escala de desempenho	52
Quadro 14. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes aos Alunos.	66
Quadro 15. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes aos Professores	67
Quadro 16. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes aos Diretores.....	68
Quadro 17. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes os Representantes de Escola.....	70
Quadro 18. Relação de artigos que compõem a Tese.	71

LISTA DE SIGLAS E ABREVIações

ANA	Avaliação Nacional da Alfabetização
ANEB	Avaliação Nacional de Educação Básica
ANRESC	Avaliação Nacional do Rendimento Escolar.
IDEB	Índice de Desenvolvimento da Educação Básica
IDH	Índice de Desenvolvimento Humano
KDD	<i>knowledge-discovery in databases</i>
MEC	Ministério da Educação
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ONU	Organização das Nações Unidas
PIB	Produto Interno Bruto
PISA	<i>Programme for International Student Assessment</i>
PNE	Plano Nacional de Educação
SAEB	Sistema de Avaliação da Educação Básica.
LDB	Lei de Diretrizes e Base

SUMÁRIO

LISTA DE FIGURAS.....	viii
LISTA DE QUADROS.....	ix
LISTA DE SIGLAS E ABREVIACÕES.....	x
1 CONSIDERAÇÕES INICIAIS.....	13
1.1 Introdução.....	13
1.2 Justificativa.....	16
1.3 Objetivos	20
1.3.1 Objetivo Geral	20
1.3.2 Objetivos Específicos.....	20
1.4 Pesquisa Bibliográfica.....	20
1.5 Organização deste trabalho.....	22
2 REFERENCIAL TEÓRICO.....	24
2.1 Publicações selecionadas referentes ao tema.....	24
2.2 Publicações relacionadas com a Mineração de Dados Educacionais.....	28
2.3 Proposições.....	32
2.4 Plano Nacional de Educação – PNE.....	33
2.5 Índice de Desenvolvimento da Educação Básica – IDEB.....	34
2.6 Avaliações em Larga Escala.....	37
2.6.1 PISA.....	37
2.6.2 Sistema SAEB.....	40
3 ESTRUTURA DA EDUCAÇÃO NO BRASIL.....	43
3.1 As instituições de ensino no país.....	45
3.2 As avaliações SAEB nos últimos anos.....	47
4 O PROCESSO DE KDD E A MINERAÇÃO DE DADOS.....	53
4.1 Pré-Processamento.....	54
4.2 Mineração de Dados.....	55
4.2.1 Mineração de Dados Educacionais.....	56
4.2.2 Pós-Processamento.....	58
4.3 Técnicas de Mineração de Dados.....	59
4.3.1 Árvores de Decisão.....	59
4.3.2 Naive Bayes.....	61
4.4 Softwares para Mineração de Dados	62
5 METODOLOGIA.....	64
5.1 Coleta e preparação dos Dados.....	64

6 RESULTADOS EM ARTIGOS	74
6.1 ARTIGO 1. Análise do Desempenho Escolar na Avaliação SAEB 2015.....	77
6.2 ARTIGO 2. Predictive Success Factors in School Performance: An analysis of the large-scale assessment in Brazil.....	92
6.3 ARTIGO 3. Ensino Médio e as Avaliações em Larga Escala do SAEB: Uma análise do desempenho apresentado no período 2005 a 2017.....	103
6.4 ARTIGO 4. Teaching practices and school performance: discovering predictors of good performance using Data Mining.....	134
6.5 ARTIGO 5. Fatores preditivos de desempenho escolar em avaliações do SAEB: influência da Gestão Escolar.....	171
6.6 ARTIGO 6. Use of Data Mining to identify the technological resources that contribute to school performance in large-scale evaluations of Brazilian high school.....	203
7 DISCUSSÃO DOS RESULTADOS	210
8 CONSIDERAÇÕES FINAIS	215
REFERÊNCIAS	218
ANEXOS	229
Anexo 1 – Plano Nacional da Educação – Metas.....	230
Anexo 2 – Escalas de Desempenhos para as proficiências Língua Portuguesa e Matemática.....	232

1 CONSIDERAÇÕES INICIAIS

1.1 Introdução

A Educação se tornou um tema de destaque em nível nacional, à medida que pesquisadores e profissionais da área procuram descobrir formas e alternativas para que o ambiente de ensino-aprendizagem possa contribuir para que os alunos desenvolvam e aprimorem suas habilidades e conhecimentos.

No entanto, o nível de qualidade da educação que o sistema de ensino brasileiro tem demonstrado reforça a necessidade de debates com intuito de identificar como as metodologias de ensino, as práticas pedagógicas e a formação do corpo docente podem reverter essa situação.

Os resultados das avaliações em larga escala, tanto em nível nacional quanto internacional, têm demonstrado que o modelo de ensino utilizado atualmente pelas escolas brasileiras não tem surtido efeitos positivos em relação ao nível de aprendizado, pois os alunos não conseguem resolver exercícios básicos propostos nessas avaliações. Diante dessas dificuldades de aprendizado, os alunos apresentam desempenhos compatíveis com os níveis mais inferiores das escalas de desempenho nas avaliações nacionais, e internacionalmente também ocupam as posições mais inferiores no ranking.

Não só o governo brasileiro, mas também vários países têm demonstrado que existe a necessidade de desenvolvimento de ações de melhorias nos sistemas de ensino, exemplo disso são os objetivos propostos para o plano Agenda 2030, que além de propor a erradicação da pobreza e melhoria da qualidade de vida dos indivíduos, também reforça que os governos possam proporcionar uma educação de qualidade para todos (ONU, 2019).

Em relação ao sistema educacional brasileiro, o governo federal alterou, por meio da Lei nº 13005/14, a maneira como o Plano Nacional de Educação - PNE era conduzido, passando, desde então, a ser uma exigência constitucional. O PNE determina as diretrizes, metas e estratégias para um determinado período, por exemplo, o plano em vigência durante o desenvolvimento desse estudo compreende o período de 2014 a 2024 (BRASIL, 2014).

Uma das propostas do PNE é avaliar a qualidade da educação do sistema educacional brasileiro, criando para isso o Indicador de Desempenho da Educação

Básica – IDEB (MEC, 2014). O PNE 2014-2024 estabelece que esse índice deve alcançar 6,0 pontos, até o ano de 2022, para os Anos Iniciais do Ensino Fundamental, de 5,5 para os Anos Finais do Ensino Fundamental e de 5,2 para o Ensino Médio (INEP, 2020a).

O índice IDEB é composto pelos dados oriundos do Censo Escolar, (aprovação), e pelo desempenho na avaliação em larga escala do Sistema de Avaliação da Educação Básica – SAEB. Esse sistema avalia os alunos egressos do Ensino Fundamental e Médio, nas proficiências de Língua Portuguesa e Matemática (INEP, 2020a).

Avaliações em larga escala têm proporcionado informações e conhecimentos, para o desenvolvimento de ações de políticas públicas com objetivo de propor melhorias no sistema de educação básica, e têm sido consideradas como uma alternativa de estudo desde a década de 1960. Neste ano, nos Estados Unidos, o governo solicitou um estudo que ficou conhecido como Relatório *Coleman*, para que se observasse a questão da igualdade de acesso ao sistema de ensino, ao mesmo tempo que na Inglaterra, o Relatório *Plowden* teve por objetivo observar a qualidade de ensino da escola primária (BROOKE e SOARES, 2011)

No entanto, apenas a avaliação do conhecimento não é suficiente para entender o processo de desempenho escolar, se faz necessário também observar o contexto em que os alunos estão inseridos. Alguns autores salientam, por exemplo, a necessidade de boa infraestrutura na escola (SCHNEIDER e GOUVEIA, 2011); outros atentam para questões envolvendo o modelo de gestão e do corpo docente (BARBOSA e WILHELM, 2009; FINI, 2009; SOARES et al., 2010) como contexto adequado para o desenvolvimento do aprendizado.

O sistema educacional brasileiro passa por diversas dificuldades, sejam elas oriundas da estrutura escolar, da gestão escolar, do corpo docente, material didático, etc., fatores esses já apontados por Campos e Jorge (2010), Pinto (2012), Castro e Carvalho (2013), Alves e Silva (2013), Garcia et al. (2014) e Ciliato e Sartori (2015).

Investimentos na educação são necessários, mas investir desordenadamente e sem foco nos objetivos de qualidade pode não gerar resultados positivos no desempenho dos alunos, na qualidade de ensino e na melhora de toda a estrutura do sistema educacional brasileiro (REZENDE PINTO, AMARAL e CASTRO, 2011; GIVISIEZ e OLIVEIRA, 2011; AMARAL, 2015; TARDA e RODRIGUES, 2015).

Percebe-se que são vários os fatores ou variáveis presentes no contexto da educação, e analisar todos individualmente seria muito desgastante e oneroso. Além das dimensões relacionadas aos Alunos, Professores, Diretores de Escola e a própria Instituição Escolar, também há a presença da sociedade, comunidade local, *stakeholders* da área, e inclusive o governo. Portanto, torna-se difícil identificar qual ou quais são os mais importantes e urgentes para o desenvolvimento de estudos e análise em prol da criação de planos de melhorias.

No Brasil, uma alternativa para análise de dados relacionados à qualidade da educação são os resultados das avaliações em larga escala do Sistema de Avaliação da Educação Básica – SAEB. Por meio dos desempenhos dos alunos e das respostas dos questionários aplicados aos Alunos, Professores, Diretores e Representantes de Escola, é possível observar informações importantes relacionadas ao ambiente escolar que podem servir de apoio para o desenvolvimento de planos de melhorias na qualidade de ensino.

No entanto, como as avaliações em larga escala são aplicadas em nível nacional, a análise dos resultados sem o auxílio de sistemas de informação pode se tornar uma grande dificuldade. Então, uma alternativa para a investigação e desenvolvimento de análises utilizando grandes volumes de dados pode ser a Mineração de Dados Educacionais, que tem como um dos principais objetivos encontrar padrões, anomalias e correlações em grandes conjuntos de dados, na tentativa de prever resultados que possam auxiliar nos processos de tomada de decisão.

Dessa forma, por meio da Mineração de Dados Educacionais é possível identificar nos resultados das avaliações em larga escala do SAEB quais são as variáveis relacionadas ao ambiente escolar que apresentam correlação mais forte com o desempenho escolar, e, ainda, a partir dos resultados dos alunos nessas avaliações, entender como essas variáveis estão associadas com os melhores desempenhos. Com isso, além de identificar as variáveis que podem estar associadas aos melhores desempenhos, também pode-se indicar, dentre elas, quais são as mais relevantes.

A descoberta das variáveis, relevantes para o desempenho escolar, pode ser considerada um passo importante para elaboração de planos de ação que visem melhorias no ambiente escolar. No entanto, se torna necessária a participação efetiva do governo, dos profissionais da educação, das famílias e da sociedade para

que, após implantados, esses planos alcancem os objetivos propostos em relação à melhoria da educação.

1.2 Justificativa

A qualidade da educação é um tema que está presente em diversas publicações no mundo todo, inclusive faz parte de debates e projetos envolvendo governos e organizações que visam propor melhorias no nível de qualidade de vida da sociedade. Como exemplo, a Organização das Nações Unidas - ONU, com base em dados coletados em vários países, desenvolveu um documento denominado Agenda 2030, com dezessete objetivos considerados essenciais para o desenvolvimento sustentável de um país. Esses objetivos abrangem vários aspectos econômicos e sociais de uma nação, inclusive a “EDUCAÇÃO DE QUALIDADE”, que no documento proposto é descrito como objetivo nº 4 “Assegurar a educação inclusiva e equitativa e de qualidade, e promover oportunidades de aprendizagem ao longo da vida para todas e todos” (ONU, 2019).

Seguindo uma lógica parecida com a Agenda 2030 da Organização das Nações Unidas – ONU, o Plano Nacional de Educação apresenta várias metas relacionadas com a qualidade da educação, desde a universalização do ensino até a melhoria da formação dos profissionais em curso de nível superior. Uma das metas que se destaca e que será analisada no desenvolvimento desse estudo é a Meta 7, que propõe:

Meta 7: fomentar a qualidade da educação básica em todas as etapas e modalidades, com melhoria do fluxo escolar e da aprendizagem, de modo a atingir as seguintes médias nacionais para o IDEB: 6,0 nos anos iniciais do ensino fundamental; 5,5 nos anos finais do ensino fundamental; 5,2 no ensino médio (MEC, 2014).

O governo brasileiro também tem se alinhado com as propostas da ONU em relação à qualidade de ensino, pois, ao propor um índice de qualidade que observará um valor mínimo para cada nível do ensino do ciclo educacional brasileiro, força uma atenção para as metodologias de ensino e práticas educacionais utilizadas em território brasileiro. Cada estado é responsável por sua estrutura educacional, mas todos se reportam ao Governo Federal e dessa forma devem atender as prerrogativas do Plano nacional de Educação – PNE.

No Brasil, a maior oferta de vagas em cursos da Educação Básica é feita pelo governo. Segundo o Censo 2019, do total de 47.874.246 de matrículas efetuadas na rede de ensino, 9.134.785 foram realizadas em escolas particulares, e ainda do total geral de matrículas, 6.842.713 são de alunos do Ensino Médio (INEP, 2020b). Esses dados enfatizam que se torna relevante observar como estão os níveis de conhecimento desses alunos, principalmente em proficiências básicas como Língua Portuguesa e Matemática.

De forma a observar a qualidade da educação, como já descrito anteriormente, a proposta do governo para o IDEB é que os índices alcancem o valor de 6,0 pontos até 2022. Diante dessa proposta, observa-se que na edição de 2019 apenas a meta para os Anos Iniciais do Ensino Fundamental, que era de 5,7, foi alcançada e superada, nesta edição o índice IDEB ficou em 5,9. A meta para os Anos Finais do Ensino Fundamental era de 5,2, no entanto ficou em 4,9 e, finalmente, a meta para o Ensino Médio que era de 5,0 ficou em 4,2.

As metas para os Anos Finais do Ensino Fundamental e do Ensino Médio só foram superadas na edição de 2011, desta edição em diante não foram superadas as demais metas. Observa-se inicialmente que o problema pode estar concentrado na fase que compreende os Anos Finais do Ensino Fundamental e que se estende para o Ensino Médio. Embora possa haver algum problema na fase dos Anos Iniciais, o foco desse estudo está nos resultados apresentados pelos alunos do Ensino Médio, pois são os indivíduos que estão finalizando o ciclo básico da educação, e que estarão habilitados para o ensino superior.

Então, como forma de identificar o nível de desempenho nas avaliações do SAEB, e para ilustrar o cenário da educação brasileira relativo ao Ensino Médio, a Tabela 1 ilustra os rendimentos dos alunos nas proficiências de Língua Portuguesa e Matemática da edição do ano de 2017. Os resultados são apresentados já separados por nível da escala de proficiência que para Língua Portuguesa é de zero a oito (0-8) e para Matemática vai de zero a dez (0-10). Os conteúdos utilizados como base para avaliação em cada proficiência são disponibilizados pelo governo e estão nos anexos desse estudo, já separados por nível da escala de avaliação.

De acordo com a documentação da avaliação do Sistema de Avaliação da Educação Básica – SAEB, os alunos que estão no nível zero da escala de proficiência, “requerem atenção especial, pois ainda não demonstram habilidades muito elementares que deveriam apresentar nessa etapa escolar” (INEP, 2020c).

Tabela 1. Resultado SAEB 2017 para o Ensino Médio

Escalas		Ensino Médio 2017							
		Feminino				Masculino			
		LP	%LP	MAT	%MAT	LP	%LP	MAT	%MAT
Nível0	< 225	150.162	20,5	174.202	23,8	175.535	28,9	133.558	22,0
Nível1	225 < 250	109.909	15,0	146.261	20,0	93.852	15,5	95.411	15,7
Nível2	250 < 275	132.177	18,1	148.927	20,4	100.573	16,6	104.788	17,3
Nível3	275 < 300	135.254	18,5	112.474	15,4	96.755	15,9	94.371	15,5
Nível4	300 < 325	107.385	14,7	74.889	10,2	74.372	12,2	78.040	12,9
Nível5	325 < 350	62.897	8,6	42.870	5,9	42.725	7,0	54.439	9,0
Nível6	350 < 375	26.430	3,6	19.220	2,6	17.962	3,0	26.881	4,4
Nível7	375 < 400	6.928	0,9	8.112	1,1	5.006	0,8	11.690	1,9
Nível8	400 < 425	645	0,1	3.624	0,5	484	0,1	5.600	0,9
Nível9	425 < 450			1.195	0,2			2.454	0,4
Nível10	>=450			13	0,0			32	0,0
Totais...:		731787	100	731787	100	607264	100	607264	100

Fonte: (INEP, 2020c)

Inicialmente, de acordo com os resultados apresentados na Tabela 1, se pode observar que o percentual de alunos que estão classificados no nível zero para cada proficiência se torna preocupante, de acordo com a Estatística Descritiva a MODA está nesse nível. Para o governo, esses alunos ainda não desenvolveram habilidades elementares que deveriam dominar nesse estágio de avaliação (INEP, 2020a). Se os resultados forem observados utilizando a MEDIANA, percebe-se que os valores para cada proficiência são compatíveis com o Nível 2 da tabela. Esse nível da escala de desempenho contempla conhecimentos básicos de cada proficiência, e podem ser conferidos no Quadro 1.

Quadro 1. Habilidades e conhecimentos observados referentes aos Níveis 1 e 2 das Escalas de Desempenho.

Língua Portuguesa	
<p>Nível 1 Desempenho maior ou igual a 225 e menor que 250</p>	<p>Os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Identificar elementos da narrativa em história em quadrinhos. • Reconhecer a finalidade de recurso gráfico em artigos. • Reconhecer a relação de causa e consequência em lendas. • Inferir o sentido de palavra em letras de música e reportagens.
<p>Nível 2 Desempenho maior ou igual a 250 e menor que 275</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Reconhecer a ideia comum entre textos de gêneros diferentes e a ironia em tirinhas. • Reconhecer relações de sentido estabelecidas por conjunções ou locuções conjuntivas em letras de música e crônicas. • Reconhecer o uso de expressões características da linguagem (científica, profissional etc.) e a relação entre pronome e seu referente em artigos e

	reportagens. • Inferir o efeito de sentido da linguagem verbal e não verbal em notícias e charges.
Matemática	
Nível 1 Desempenho maior ou igual a 225 e menor que 250	Os estudantes provavelmente são capazes de: TRATAMENTO DE INFORMAÇÕES Associar uma tabela de até duas entradas a informações apresentadas textualmente ou em um gráfico de barras ou de linhas.
Nível 2 Desempenho maior ou igual a 250 e menor que 275	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: ESPAÇO E FORMA Reconhecer as coordenadas de pontos representados em um plano cartesiano localizados no primeiro quadrante. NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES Reconhecer os zeros de uma função dada graficamente. Determinar o valor de uma função afim, dada sua lei de formação. Determinar resultado utilizando o conceito de progressão aritmética. TRATAMENTO DE INFORMAÇÕES Associar um gráfico de setores a dados percentuais apresentados textualmente ou em uma tabela.

Fonte: (INEP, 2020c)

Os desempenhos apresentados na Tabela 1 reforçam a necessidade da construção de um modelo de ensino aprendizagem que possa propor melhorias no processo de desenvolvimento do conhecimento e aprendizado, durante a fase da educação escolar dos indivíduos. Com foco na Avaliação SAEB, os resultados das pesquisas de Filho, Nuñez e Ribeiro (2009), Schneider e Gouveia (2011) e Fonseca e Namen (2016) convergiram também para uma necessidade de melhorias do sistema de ensino, observando inclusive a necessidade da participação mais efetiva não apenas dos docentes, mas também dos gestores escolares, dos alunos e seus responsáveis (pais) e do governo de um modo geral, no desenvolvimento de planos de ação.

As avaliações em larga escala são aplicadas em nível nacional, com isso fica impossível analisar todos os resultados manualmente. Nesse ponto surgem como alternativa as técnicas de Mineração de Dados Educacionais, que têm por objetivo analisar grandes volumes de dados para encontrar padrões, anomalias e correlações para prever resultados, contribuindo, com isso, para o processo de tomada de decisão.

Essa pesquisa se torna importante à medida que pode revelar, por meio da Mineração de Dados Educacionais, os fatores que estão associados aos melhores desempenhos nas avaliações em larga escala. Com isso, os profissionais da educação podem elaborar planos de ação para implementar melhorias, não só em relação aos fatores descobertos, mas também em todo ambiente educacional.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo central desse estudo é analisar, por meio da Mineração de Dados Educacionais, quais fatores presentes no ambiente educacional podem ser preditores para que os alunos obtenham um bom desempenho nas avaliações de larga escala.

1.3.2 Objetivos Específicos

De modo a alcançar o objetivo geral estabelecido para esse estudo, há a necessidade também de responder os seguintes objetivos específicos:

- a)** analisar os resultados das avaliações em larga escala das edições de 2015 e 2017 do SAEB, de modo a identificar possíveis padrões de desempenho dos alunos do Ensino Médio;
- b)** identificar, por meio de análise bibliográfica, os fatores preditores de desempenho escolar dos alunos do Ensino Médio, que estejam relacionados com os Professores, Diretores e ambiente escolar;
- c)** Descobrir por meio da mineração de dados, os fatores preditores para que os alunos possam obter um melhor desempenho nas avaliações em larga escala.

1.4 Pesquisa Bibliográfica

Para compor uma primeira visão sobre o tema desse trabalho, foi realizada pesquisa bibliográfica nas bases de dados de periódicos da CAPES, Web of Science, Science Direct (acessados pelo portal da CAPES www.periódicos.capes.gov.br), Google Acadêmico (www.scholar.google.com.br), e Scielo (<http://www.scielo.org/php/index.php>), constando de duas fases principais: (1) coleta, filtragem e normalização dos dados, e (2) análise e síntese dos dados coletados para a elaboração do documento final.

O Quadro 1 contém os resultados das pesquisas efetuadas. Utilizou-se como filtro de busca as palavras-chave desse estudo, que estavam presentes nos Títulos, Resumos e nas Palavras-Chave dos artigos e trabalhos publicados. Em síntese, os resultados apresentados no Quadro 1 contêm a quantidade de publicações

encontradas que possuem uma ou mais palavras chaves contidas em seus Títulos ou Resumos.

Quadro 2. Pesquisa Bibliométrica relativa às palavras-chave do estudo.

Palavras-Chave/Keywords	CAPES	Scholar Google	Web of Science	Science Direct	SciELO
“Dados Educacionais”	44	4.170	0	2	7
“Dados Educacionais” AND “SAEB”	4	901	0	0	0
“Desempenho Escolar”	1.162	71.400	3	94	417
“Desempenho Escolar” AND “SAEB”	63	8.080	0	3	8
“Mineração de Dados Educacionais”	10	775	0	1	1
“Mineração de Dados Educacionais” AND “SAEB”	1	47	0	0	0
“Avaliação em Larga Escala”	118	5.990	0	1	30
“Avaliação em Larga Escala” AND “SAEB”	16	3.360	0	1	1
“School Performance”	41.337	1.070.000	5.752	16.305	492
“School Performance” AND “SAEB”	25	1.280	66	13	8
“Educational Data”	5.556	56.200	1.851	1.431	23
“Educational Data” AND “SAEB”	4	148	2	2	0
“Educational Data Mining”	2.133	25.300	1.186	383	6
“Educational Data Mining” AND “SAEB”	0	43	0	0	0
“Large-scale Assessment”	3.860	33.000	810	2.053	30
“Large-scale Assessment” AND “SAEB”	6	262	1	2	2

Fonte: Desenvolvido pelo Autor e atualizado em outubro/2020 (2020).

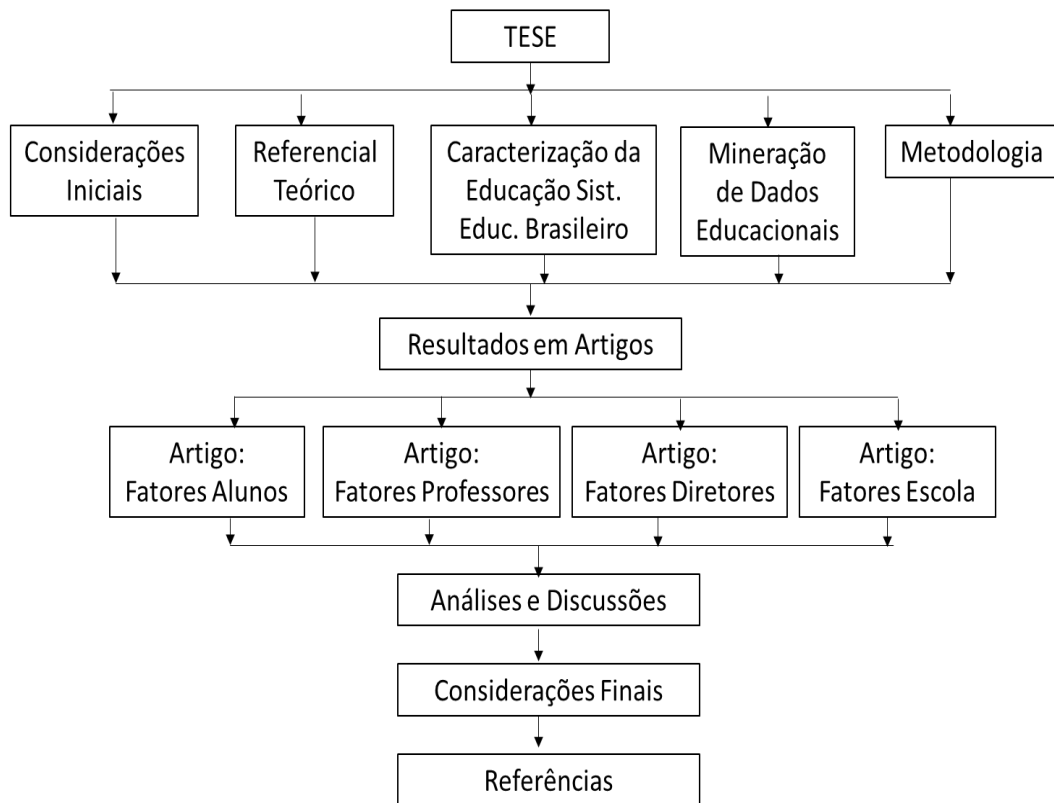
De acordo com os resultados descritos no Quadro 1, existe uma grande quantidade de trabalhos desenvolvidos e publicados sobre o tema *Educational Data Mining*. No entanto, o total de publicações relativas a esse tema com foco nas avaliações do SAEB ainda é pequeno, demonstrando com isso um campo que necessita de maiores investigações e aprofundamentos.

Após a conclusão da pesquisa bibliométrica, do total de artigos verificados, alguns foram escolhidos para compor o debate teórico inicial sobre o tema, reforçando não só a temática do estudo, mas também os pontos que ainda merecem atenção e investigação.

1.5 Organização deste trabalho.

Esse estudo terá sua configuração estipulada em sete capítulos para uma melhor compreensão dos assuntos expostos. De acordo com a Figura 1, é possível observar a ordenação dos capítulos e o conteúdo de cada um deles.

Figura 1. Estrutura do Desenvolvimento da Tese



Fonte: Desenvolvido pelo autor (2020)

No Capítulo 1 foram descritas as Considerações Iniciais; o Capítulo 2 – Referencial Teórico – apresenta a análise de um conjunto de artigos coletados em bases de dados científicas. Com os resultados dessa análise foi possível observar quais atributos estão mais presentes ou associados com os melhores desempenhos em avaliações de larga escala.

Já o Capítulo 3 apresenta a estrutura e caracterização do sistema educacional brasileiro e o Capítulo 4 os conceitos e técnicas referentes ao uso da Mineração de Dados Educacionais. Em seguida, o Capítulo 5 apresenta os passos metodológicos necessários para o desenvolvimento dessa pesquisa.

O Capítulo 6 – Resultados em Artigos, apresenta os artigos que foram desenvolvidos a partir das análises dos dados. Os artigos são descritos na íntegra e

estão diretamente relacionados com os objetivos específicos e proposições desse estudo.

No Capítulo 7 os resultados encontrados são analisados e confrontados com a literatura e no Capítulo 8 é realizada uma análise geral da tese, bem como as contribuições e sugestões para a realização de novos estudos a partir dos resultados encontrados.

Como último capítulo, as Referências são agrupadas e apresentadas para posteriores consultas.

2 REFERENCIAL TEÓRICO

O presente capítulo tem como objetivo delimitar o referencial teórico que será utilizado para conduzir esse trabalho. Inicialmente, uma visão geral do volume de publicações sobre os tópicos dessa investigação mostra a importância do tema na academia. Em seguida, além das proposições apresentadas para esse estudo, também serão abordados os seguintes tópicos: Plano Nacional da Educação – PNE, Estrutura da Educação Brasileira, Avaliações em Larga Escala, Índice de Desenvolvimento da Educação Básica – IDEB.

2.1 Publicações selecionadas referentes ao tema.

Do total de publicações descritas no Quadro 2, item 1.4 da Introdução, foram selecionados artigos que contribuíram com o desenvolvimento dessa pesquisa. A escolha desses artigos teve como critério a importância dos atributos para o desempenho escolar, segundo os resultados apresentados em cada artigo. A ordem de leitura da publicação selecionada foi: primeiro o resumo, logo após os resultados e conclusões e finalmente a metodologia utilizada. Foram observados nos resultados das publicações os fatores que estão associados aos melhores desempenhos escolares, e que, de alguma forma, convergiam para o objetivo dessa pesquisa.

Com base nos resultados encontrados, um fator importante é o nível socioeconômico dos alunos. Alunos com nível socioeconômico mais baixo possuem mais dificuldades de aprendizado, necessitando com isso de maior atenção da escola, atenção essa que deve ser direcionada não só para o seu aprendizado, mas também para suas condições básicas de bem-estar (alimentação, segurança e apoio emocional, entre outras) (SIMÕES e FERRÃO, 2005; BEZERRA e KASSOUF, 2006; GONÇALVES e FRANÇA, 2008; CUNHA et al., 2009; RODRIGUES, RIOS-NETO e PINTO, 2011; LAROS, MARCIANO e ANDRADE, 2012; BEZERRA et al., 2016; TRAVITZKI, FERRÃO e COUTO, 2016) . Complementando, Brooke e Soares (2011) indicam ainda que, além do nível socioeconômico, o nível cultural das famílias também influencia o desempenho dos alunos, já que Luz (2006) aponta que no Brasil existe, entre esses dois fatores, uma relação acentuada.

Em relação à renda familiar, pode-se observar que em comunidades mais carentes, em alguns casos, os jovens param de estudar para realizar trabalhos informais, com a finalidade de ajudar na renda familiar. Com isso surge o atraso

escolar, o trabalho infantil e os problemas oriundos das condições socioeconômicas (BEZERRA e KASSOUF, 2006)

Após analisarem os fatores do desempenho escolar no SAEB 2011, Laros, Marciano e Andrade (2010) observaram que os alunos do sexo masculino obtêm melhores resultados na avaliação de matemática e que a questão do atraso escolar, ou seja quando o aluno é repetente, influencia negativamente o desempenho nas avaliações de um modo geral (BEZERRA e KASSOUF, 2006; ANDRADE e LAROS, 2007; BRITO, 2017; FERNANDES et al., 2018).

Além dos fatores socioeconômicos dos alunos, da comunidade e da escola, é importante salientar que os alunos também devem ter iniciativa, habilidades sociais e desejo pela aprendizagem, pois alunos que não estão motivados para o desenvolvimento do conhecimento, tendem a ter mais dificuldades no ciclo educacional e apresentam desempenhos ruins nas avaliações (SIMÕES e FERRÃO, 2005; BEZERRA e KASSOUF, 2006; MARTURANO e PIZATO, 2015; BRITO, 2017; JOVARINI, LEME e CORREIA-ZANINI, 2018).

Outro fator que pode influenciar no desempenho dos alunos é a questão do cansaço ou desgaste por conta da rotina escolar. Dessa forma, cabe ao professor entender essas condições e propor atividades que possam conduzir os alunos ao nível desejado de conhecimento, utilizando para isso alternativas de práticas pedagógicas em suas aulas. Fatores relacionados aos conteúdos de aulas e do clima em sala de aula, notas baixas, necessidade de apoio para realização de tarefas já foram analisados por Correia-Zanini (2013), Marturano e Pizato (2015) e Jovarini, Leme e Correia-Zanini (2018).

Para Sebanc, Guimond e Lutgen (2016), existe correlação positiva entre qualidade do relacionamento com amigos na escola e o desempenho. Na mesma direção, Fernandes et al. (2018) sinalizam que os alunos podem superar as dificuldades relacionadas às evoluções dos ciclos escolares se tiverem facilidade de manter e criar novas amizades

Em relação à participação do professor no processo de ensino-aprendizagem, Dourado, de Oliveira e Santos (2007) e Ortigão (2008) propõem que não se pode penalizar o professor pelo baixo desempenho dos alunos, pois existe a necessidade de se analisar em conjunto as condições objetivas de ensino, as desigualdades sociais, econômicas e culturais dos alunos, bem como a desvalorização profissional e a possibilidade restrita de atualização permanente dos profissionais da educação.

Nos estudos de Soares et al. (2010) e de Fonseca e Namen (2016), os resultados apontaram para a questão da expectativa do professor em relação ao ambiente escolar e à condição sociodemográfica dos alunos. Os autores concluíram que uma maior expectativa por parte dos docentes influencia positivamente nos resultados dos alunos, nas proficiências analisadas. Corroborando essa argumentação, Pereira e Mori (2012) salientam que é de responsabilidade do professor desenvolver atividades de ensino que despertem o interesse do estudante, pois essa é uma das principais ações do corpo docente no processo de ensino-aprendizagem.

Para Amâncio-Vieira et al. (2015) e Vidal e Vieira (2017), o professor é peça fundamental no desenvolvimento educacional do aluno, sem desconsiderar a questão do nível socioeconômico dos discentes. O corpo docente deve estar sempre preparado para novos desafios e temáticas atuais, entre outros assuntos. Dessa forma, na visão de Mesquita (2018), o corpo docente deve possuir especializações acadêmicas (cursos de extensão universitária) para ter condições suficientes para desenvolver conteúdos mais orientados com a realidade social e mercado de trabalho. Já para Filho, Nuñez e Ribeiro (2009) o número de faltas dos professores afetam significativamente o desempenho dos alunos.

A questão da remuneração dos professores pode ser fator motivador para o magistério. Os salários do corpo docente, atrelado a uma carga horária semanal considerada relativamente alta, de certo não contribuem para o desenvolvimento de novas metodologias de ensino, bem como para o desenvolvimento de novos estudos envolvendo os currículos educacionais básicos (FINI, 2009; FONSECA e NAMEN, 2016; SIMIELLI, 2017).

Com foco no SAEB 2003, os resultados alcançados por Gonçalves e França (2008) indicaram que uma maior participação de pais e mestres, e uma boa infraestrutura escolar, contribuem na redução da necessidade de complementaridade da família. A escolaridade dos pais parece contribuir com o bom desempenho dos alunos, fato esse já observado nos estudos de Bezerra e Kassouf (2006) e de Barbosa e Wilhelm (2009).

Os resultados encontrados por Barbosa e Wilhelm (2009) e Filho, Nuñez e Ribeiro (2009) indicaram que possíveis falhas relacionadas à gestão escolar podem dificultar o aprendizado do aluno. Os autores, por meio dos resultados encontrados, ainda apresentam subsídios para que os Diretores de escolas possam tomar

decisões mais responsáveis em relação ao direcionamento de recursos e esforços, com o objetivo de que as escolas sejam mais eficientes. Inclusive Drabach (2018) salienta que a forma como o Diretor de escola é escolhido para o cargo também pode influenciar nos processos de gestão.

Já para Garcia et al. (2016) e Philippsen (2018), o processo de gestão eficiente faz com que os resultados positivos de uma fase da educação possa ser replicada para as outras. Dessa forma, para que isso aconteça é necessário que o gestor tenha uma visão mais ampla de todos os segmentos do ambiente escolar. Para Oliveira e Waldhelm (2016), o papel do gestor escolar é primordial para um bom desempenho educacional, pois o estilo de liderança pode ser fator preponderante para a construção de um clima organizacional favorável à execução do trabalho docente, incluindo a isso o tempo de experiência na função (FILHO, NUÑEZ e RIBEIRO, 2009).

Ainda em relação às atividades do gestor escolar, é imprescindível que exista na escola um planejamento que aborde as questões envolvendo o absenteísmo dos alunos, causas de repetência escolar e prevenção a incidências de indisciplina por parte dos alunos (LAROS e MARCIANO, 2010)

Outra condição importante para o bom desempenho escolar é a condição da estrutura das escolas. Escolas que possuem laboratórios de informática, ciências, bibliotecas, salas de leitura entre outros itens, podem ofertar condições suficientes no sentido de se criar um ambiente saudável para o desenvolvimento do ensino (FELÍCIO e FERNANDES, 2005; GATTI, 2009; SCHNEIDER e GOUVEIA, 2011; FONSECA e NAMEN, 2016; CALIXTO, SEGUNDO e GUSMÃO, 2017; BRITO, 2017).

Em relação ao uso de computadores por parte dos alunos, Dwyer et al. (2007) e Gonçalves e França (2008) salientam que se as atividades que necessitam de apoio da tecnologia não forem bem desenvolvidas e acompanhadas, podem gerar resultados considerados ruins no que se refere ao desempenho escolar.

Esses estudos descritos não utilizaram a Mineração de Dados Educacionais como metodologia, dessa forma se faz necessário também observar as publicações que utilizaram essa técnica para encontrar seus resultados.

2.2 Publicações relacionadas com a Mineração de Dados Educacionais.

O uso da Mineração de Dados Educacionais vem se tornando uma importante alternativa para o desenvolvimento de pesquisas na área da educação, principalmente quando é necessário observar grandes volumes de dados, como exemplo resultados das avaliações de larga escala. As informações apresentadas por meio das técnicas de mineração podem proporcionar melhores condições de análise, pois para produzir os resultados, são utilizados algoritmos e aprendizagem de máquina.

Como exemplo de estudos publicados envolvendo a Mineração de Dados Educacionais, Bezerra et al. (2016) utilizaram essa técnica para identificar os fatores que motivaram os alunos do ensino fundamental de escolas públicas do estado de Pernambuco a desistirem de seus estudos, se evadindo das escolas. Os resultados encontrados por meio de Árvores de Decisão, Indução de Regras e Regressão Logística indicaram que fatores como a idade, turno das aulas e ainda os locais onde as escolas estavam situadas, influenciavam fortemente a evasão escolar.

Já os resultados encontrados por Calixto, Segundo e Gusmão (2017), utilizando Mineração de Dados Educacionais nos dados dos Censos Escolares dos anos de 2014 a 2016, em escolas dos estados do Ceará e Sergipe, indicaram que alguns fatores que motivam a evasão a escolar são a idade e a localização da escola, e, além desses, os autores destacam ainda a etapa de ensino, modalidade de ensino e existência de laboratórios. As técnicas de mineração utilizadas foram Indução de Regras e Regressão Logística.

Utilizando a Mineração de Dados Educacionais no Exame Nacional do Ensino Médio – ENEM do ano de 2015, os resultados encontrados por Simon e Cazella (2017) indicaram que os fatores Tipo da Escola e Nível Socioeconômico dos alunos contribuem para melhores desempenhos e, ainda, que os alunos com nota superior a 550 pontos na proficiência Ciências da Natureza e suas Tecnologias apresentam no mínimo o fator Nível Socioeconômico médio alto, independente se a escola é privada ou pública.

Como objetivo de analisar fatores que motivam alunos do Ensino Médio a abandonarem os estudos, Colpani (2018) utilizou as técnicas de mineração, com base em correlação e regressão linear, nos dados do Censo Escolar do ano de 2017. Como resultado, a taxa de distorção idade-série foi a mais associada com a evasão escolar, indicando com isso que se torna necessário acompanhar com mais

atenção os alunos que estão nessa condição, pois podem estar mais propícios ao abandono dos estudos.

Com o propósito de observar o desempenho dos alunos do 5º ano do Ensino Fundamental das escolas públicas do estado de Pernambuco, Silva (2019), utilizando as técnicas de Mineração de Dados Educacionais, e por meio da regressão linear, chegou à conclusão de que o incentivo dos pais para o estudo, a resolução das atividades escolares e a mãe ser alfabetizada, são fatores importantes que influenciam o desenvolvimento educacional desses alunos.

Com o objetivo de identificar os atributos motivadores para a evasão escolar de alunos da educação básica matriculados na rede municipal da cidade de Juiz de Fora, Minas Gerais, Sales et al. (2019) encontraram como resultados três atributos, sendo eles a etnia do aluno, se o aluno possui um responsável e a ainda o turno que o aluno estuda. Os autores utilizaram a técnica denominada *Random Forest*, que é uma floresta de Árvores de Decisão.

De forma a contribuir para os processos decisórios dos gestores de escolas Júnior et al. (2019) utilizaram a mineração de dados para analisar o índice de Desenvolvimento da Educação Básica (IDEB) das escolas públicas de Maceió, no estado de Alagoas. Por meio dos algoritmos para Regressão Linear e Árvores de Decisão, os resultados encontrados indicaram que diversos fatores puderam influenciar o desempenho do aluno, sendo eles a escolaridade dos pais do aluno, o incentivo aos estudos e o compromisso do docente.

Além desses estudos descritos, outros com metodologia amparada pelas técnicas de Mineração de Dados Educacionais, e que apresentavam como objetivo observar principalmente os desempenhos dos alunos, seja do Ensino Fundamental, Ensino Médio ou do Ensino Superior, apresentavam em seus resultados contribuições para melhorias dos processos de gestão escolar e de ensino-aprendizagem. Esses estudos são analisados individualmente nos artigos que são frutos do desenvolvimento dessa pesquisa.

Com base nos resultados dos estudos coletados, observa-se que existem diversos fatores que estão associados ao desempenho escolar, fatores esses que são pertinentes das características do aluno e sua família, do professor e de suas ações pedagógicas, do diretor e dos modelos de gestão e da estrutura da escola. Os resultados das publicações indicaram que esses fatores podem influenciar o

processo de ensino-aprendizagem e com isso proporcionar melhores desempenhos dos alunos, nas avaliações em larga escala.

Após finalizada a etapa de análise dos artigos relacionados com o tema desse estudo, foi construído um quadro contendo todos os fatores preditores do desempenho escolar, localizados nesses artigos. O Quadro 3 ilustra esses fatores preditores, já separados por dimensão, ou seja, alunos, professores, diretores e escolas. Esse quadro foi utilizado como referencial para comparação com os dados dos formulários disponibilizados pelo INEP, provenientes da avaliação do SAEB.

Quadro 3. Relação dos fatores preditores do desempenho escolar.

	Autores	Fatores Observados
ALUNOS	Simões e Ferrão (2005); Bezerra e Kassouf (2006); Luz(2006); Gonçalves e França (2008); Brooke e Soares (2008); Cunha et al. (2009); Rodrigues, Rios-Neto e Pinto (2011); Laros, Marciano e Andrade (2012); Bezerra et al. (2016); Travitzki, Ferrão e Couto (2016); Gustafsson, Nilsen e Hansen (2018); Murphy (2019); Queiroz, Sampaio e Sampaio (2020)	Nível socioeconômico da família/aluno
	Bezerra e Kassouf (2006) ; Barbosa e Wilhelm (2009); Machebe, Ezegbe e Onuoha (2017); Castillo-Vergara et al. (2018)	Renda familiar
	Bezerra e Kassouf (2006); Lobo (2017); Fantinato e Cia (2017)	Escolaridade da mãe
	Barbosa e Wilhelm (2009); Lobo (2017); Fantinato e Cia (2017)	Escolaridade dos pais
	Barbosa e Wilhelm (2009)	Alunos que trabalham
	Simões e Ferrão (2005); Bezerra e Kassouf (2006); Marturano e Pizato (2015); Brito (2017); Jovarini, Leme e Correia-Zanini (2018)	Sempre fazer as lições escolares e motivação para aprender
	Correia-Zanini (2013); Marturano e Pizato (2015); Jovarini, Leme e Correia-Zanini (2018)	Fatores estressantes (clima sala de aula, notas baixas, apoio para realiza as tarefas, etc.)
	Sebanc, Guimond e Lutgen (2016); Fernandes et al. (2018).	Habilidades sociais dos alunos (relacionamento com colegas, empatia, etc.)
	Bezerra e Kassouf (2006) ; Andrade e Laros (2007); Gonçalves e França (2008); Laros, Marciano e Andrade (2010); Brito (2017); Fernandes et al. (2018); Paula, Franco e Silva (2018); Lacruz et al. (2019)	Atraso escolar
	Dwyer et al. (2007); Gonçalves e França (2008)	Uso do computador
Gonçalves e França (2008); Zafani e Omote (2016); Brito (2017)	Cobrança e incentivo dos pais do aluno	
PROF ESSO DES	Laros e Marciano (2010)	Estimular o trabalho colaborativo entre os professores

	Soares et al. (2010); Fonseca e Namen (2016); Lauer mann (2017); Schildkamp et al. (2020); Zakariya, Bjorkestol e Nilsen (2020)	Expectativas e Motivação do professor
	Bezerra e Kassouf (2006); Dourado, de Oliveira e Santos (2007); Ortigão (2008); Hanushek e Woessmann (2017); Lauer mann (2017); Mesquita (2018); Symeonidis (2019); Voinea (2019); Nascimento e Gomes (2020); Schildkamp et al. (2020); Zakariya, Bjorkestol e Nilsen (2020).	Escolaridade/formação dos professores
	Bezerra e Kassouf (2006); Fini (2009); Fonseca e Namen (2016); Simielli (2017).	Nível de salários professores
	Pereira e Mori (2012); Amâncio-Vieira et al. (2015); Fonseca e Namen (2016); Pettersson et al. (2016); Vidal e Vieira (2017); Jovarini, Leme e Correia-Zanini (2018); Schaik et al.(2018)	Atividades pedagógicas do corpo docente
DIRETOR DE ESCOLA	Gonçalves e França (2008); Barbosa e Wilhelm (2009)	Participação dos pais na administração das escolas
	Filho, Nuñez e Ribeiro (2009); Souza e Ribeiro (2017); Vieira e Bussolotti (2018); Barros et al. (2019); Pereira e Gasque (2019); Marangoni et al. (2019); Dallazem e Coelho (2019); Nascimento e Chiusoli (2019); Ribeiro, Síveres e Brito (2019); Zambello et al. (2019); Nichele e Mello (2020); Morais et al. (2020); Pereira (2020); Souza (2020).	Experiência do diretor, Liderança e formação educacional
	Filho, Nuñez e Ribeiro (2009)	Estabilidade do corpo docente
	Filho, Nuñez e Ribeiro (2009); Lacruz et al. (2019)	Frequente faltas por parte dos professores
	Filho, Nuñez e Ribeiro (2009); Laros e Marciano (2010)	Ocorrência de roubos e episódios de violência nas escolas
	Filho, Nuñez e Ribeiro (2009); Drabach (2013); Cária e Lambert-De-Andrade (2016); Back et al. (2016); Garcia et al. (2016); Oliveira e Waldhelm (2016); Philippsen (2018); Vieira e Bussolotti (2018); Aldrup et al. (2018); Masci, De Witte e Agasisti (2018); Freitas (2019); Riddle e Apple (2019)	Processos de Gestão Escolar
	Laros e Marciano (2010)	Aumentar o controle sobre a incidência de faltas na sala de aula
	Laros e Marciano (2010) ; Mohr e Naujorks (2017); Góis e Rocha (2019)	Sanar as causas da repetência do ano letivo
ESCOLA	Gonçalves e França (2008); Laros, Marciano e Andrade (2012); Amâncio-Vieira e Pereira (2016); Gustafsson, Nilsen e Hansen (2018); Murphy (2019);	Nível socioeconômico da escola

Soares e Alves (2003); Felício e Fernandes (2005); Bezerra e Kassouf (2006); Gonçalves e França (2008); Gatti (2009); Schneider e Gouveia (2011); Amâncio-Vieira e Pereira (2016); Fonseca e Namen (2016); Calixto, Segundo e Gusmão (2017); Brito (2017); Razak et al. (2018); Damon et al. (2019); Ford (2019); Lasater, Bengtson e Albiladi (2020); Vasconcelos et al. (2020)	A infraestrutura escolar (salas de aula, laboratórios; material didático; materiais audiovisuais, etc.)
--	---

Fonte: Desenvolvido pelo Autor (2019)

A partir da análise das publicações, dos fatores descritos no Quadro 3 e dos formulários do SAEB disponibilizados pelo governo, foi possível desenvolver as proposições que direcionaram esse estudo, bem como, separar todas as informações necessárias para compor a fase de Mineração de Dados Educacionais.

2.3 Proposições

As proposições que contribuíram para o direcionamento dessa pesquisa foram elaboradas a partir da análise dos fatores preditores do desempenho referentes a cada dimensão, descritos no Quadro 3, em conjunto com as publicações observadas e com os formulários SAEB. Nesse sentido, as proposições que norteiam este estudo são:

P1. Os alunos que são incentivados e orientados por seus pais ou responsáveis em relação ao compromisso com os estudos durante o ciclo escolar, podem apresentar melhor nível de aprendizado e de desempenho escolar.

P2. Professores comprometidos com o nível de aprendizagem dos alunos têm mais condições de desenvolver práticas pedagógicas que proporcionem melhorias no ambiente de ensino e no desempenho escolar.

P3. Diretores de escola que são comprometidos com as melhorias das práticas de gestão e com os processos de ensino-aprendizagem, conseguem desenvolver ambientes mais adequados para as práticas de ensino.

P4. Ambiente escolar adequado para as práticas de ensino e que possa proporcionar acesso a recursos tecnológicos e de Internet apresentam melhores condições para o desenvolvimento do nível de conhecimento e de aprendizado dos alunos.

Em relação às ações direcionadas para melhoria da qualidade da educação no Brasil, o governo federal desenvolveu o Plano Nacional de Educação – PNE, que fornece informações essenciais para que os estados, municípios e instituições do governo federal possam conduzir seus planejamentos relacionados ao segmento educacional.

2.4 Plano Nacional de Educação – PNE.

O governo federal brasileiro decidiu, em 2014, por meio da Lei nº 13.005 de 25/06/2014, que o Plano Nacional de Educação - PNE passasse a ser uma exigência constitucional. A periodicidade do PNE é decenal e os estados devem adaptar seus planos educacionais com base nas diretrizes tratadas neste plano, inclusive no que se refere ao percentual de investimento estadual na educação, com base no Produto Interno Bruto - PIB (MEC, 2014).

O Plano Nacional de Educação para o período 2014-2024 tem como objetivo propor vinte metas, descritas no Anexo 1 desse estudo, com foco no desenvolvimento educacional, desde a política de acesso até o acompanhamento do desempenho escolar, apresentando inclusive estratégias para o alcance dessas metas (MEC, 2014).

Todas as metas têm sua importância no que se refere à melhoria do sistema educacional brasileiro, no entanto a meta que objetiva acompanhar o desempenho de aprendizagem dos alunos é a META 7, pois fixa valores de referência para os Índices de Desenvolvimento da Educação Básica-IDEA separados por nível educacional, sendo de 6,0 para os Anos Iniciais do Ensino Fundamental, 5,5 para os Anos Finais do Ensino Fundamental e 5,2 para o Ensino Médio (INEP, 2020d).

Os estados, distritos e municípios devem apresentar soluções práticas em seus planos educacionais para que tenham condições de cumprir as metas propostas pelo PNE. Alguns tópicos devem ser observados para que o sucesso seja plenamente alcançado, situações como acesso ao sistema educacional por crianças de 4 a 5 anos, previstas na Constituição Federal, já é uma questão preocupante para algumas cidades, pois faltam creches e unidades de pré-escola suficientes para atender a população local.

Em relação ao investimento na educação, a META 20 do Plano Nacional de Educação salienta que o governo deve destinar no mínimo 7% do seu Produto Interno Bruto à educação. Os dados divulgados pelo Instituto de Pesquisa Anísio

Teixeira – INEP indicam que muito esforço ainda precisa ser feito para que a META 20 seja cumprida, pois no período de 2000 a 2017, o percentual de investimento na educação evoluiu apenas 1,7%, sendo de 4,6% no ano 2000 e de 6,3% em 2017 (INEP, 2020e).

O último relatório divulgado pelo Tesouro Nacional Brasileiro referente aos Aspectos Fiscais da Educação no Brasil indicou que em 2017 o Brasil gastou mais do que a média dos países membros da Organização para a Cooperação e Desenvolvimento Econômico – OCDE, que foi de 5,5% dos PIBs. Apenas como exemplos, a Argentina investiu na educação 5,3%, Colômbia 4,7%, México 5,3% e Estados Unidos 5,4%, dos seus respectivos PIBs (BRASIL, 2018).

Se for observada a questão do gasto anual por aluno, o Brasil está bem abaixo da média da OCDE. No relatório denominado “*Education at Glance 2020*” publicado pela OCDE, o Brasil apresentou em 2017 um gasto de US\$ 4.661 anuais por aluno, enquanto que a média dos países foi de US\$ 10.102 (OCDE, 2020).

Embora a meta proposta no PNE seja de que o governo destine 7% do PIB para a educação, e mesmo que essa meta esteja quase sendo alcançada, esses recursos parecem não produzir o efeito esperado para todos os níveis da educação, pois conforme os resultados apresentados pelos alunos do Ensino Fundamental 2 (5º e 9º anos), e do Ensino Médio nas avaliações em larga escala do SAEB, estão distantes do que o governo considera ideal, que é a meta 6.0 para o IDEB.

Os resultados dos alunos do Ensino Fundamental nas avaliações de larga escala do PISA também são semelhantes aos do SAEB, ou seja, internacionalmente o Brasil ocupa classificações muito baixas em relação aos demais países que participam desse tipo de avaliação. Os resultados do PISA serão analisados em item próprio desse estudo.

Os recursos físicos e financeiros disponibilizados pelos governos devem ser bem administrados para que possam, em um determinado período de tempo, produzir condições suficientes para que os resultados dos alunos nas avaliações tanto nacionais quanto internacionais evoluam positivamente.

2.5 Índice de Desenvolvimento da Educação Básica - IDEB

Criado em 2007 por meio do Decreto nº 6.094 de 24 de abril de 2007, o Índice de Desenvolvimento da Educação Básica – IDEB tem como principal objetivo acompanhar a qualidade da educação. Esse indicador basicamente observa os

resultados referentes ao fluxo escolar e as médias de desempenho nas avaliações do sistema SAEB.

A partir da criação, pelo Ministério da Educação-MEC, do Plano de Metas Compromisso Todos pela Educação, o governo passa a exercer um determinado controle sobre como os estados e municípios desenvolvem suas metodologias e práticas de ensino. O acompanhamento dos resultados dos índices IDEB estaduais se torna importante à medida que podem fornecer informações sobre o rendimento escolar, e é a partir desse conhecimento que o governo federal pode atuar com apoio financeiro, assistência técnica e de gestão, sempre que se fizer necessário.

Para que se obtenha o índice IDEB, é necessário reunir informações sobre promoção, repetência e evasão escolar dos alunos, informações essas, fornecidas pelo Censo Escolar (INEP, 2020b). O agrupamento dessas informações é definido como fluxo escolar. Além do fluxo escolar, também são coletados os resultados dos alunos nas proficiências avaliadas no sistema SAEB. A partir da combinação entre os dados do fluxo escolar e das médias de desempenho no SAEB, obtém-se uma média que varia de zero a dez (0 a 10). Essa média é aferida para os alunos do Ensino Fundamental (5º e 9º anos) e para a 3ª série do Ensino Médio (INEP, 2020a).

Por meio do índice IDEB o governo acredita que os profissionais da educação e a sociedade passarão a acompanhar a qualidade da educação, pois os resultados são divulgados pelo INEP e todos podem consultar. De certa forma, um dos grandes problemas da educação está relacionado com a qualidade oferecida pelas escolas que compõem a rede de ensino brasileira (CHIRINÉA e BARREIRO, 2009; SCHNEIDER e GOUVEIA, 2011; SANTOS, MONTANUCI e PACHECO, 2019).

Como já mencionado, a proposta do governo é que até o ano de 2022, o índice IDEB seja de 6,0 para os Anos Iniciais do Ensino Fundamental, 5,5 para os Anos Finais do Ensino Fundamental e 5,2 para o Ensino Médio. Esses valores são estabelecidos utilizando-se como base, o desenvolvimento educacional dos países desenvolvidos. Essa comparação com os países desenvolvidos só foi possível por que o governo adotou uma técnica de compatibilização entre as proficiências observadas na avaliação PISA e no SAEB (FERNANDES, 2007).

Cada estado e município brasileiro tem sua meta diferenciada, estipulada para cada edição do SAEB, que ocorre bianualmente. Embora a média da meta unificada para o sistema educacional brasileiro seja de 6,0 até o ano de 2022, o

objetivo é que esse índice seja 9,9 até o ano de 2096. Ao alcançar esse valor, o governo acredita que o Brasil terá reduzido significativamente as dificuldades de acesso à educação, além de proporcionar melhores controles relacionados com a qualidade de ensino.

A fórmula para o cálculo do IDEB de acordo com o exposto por Fernandes (2007) é:

$$\text{IDEB}_{ji} = N_{ji} P_{ji}, \quad (1)$$

onde:

i = ano do exame (Saeb e Prova Brasil) e do Censo Escolar;

N_{ji} = média da proficiência em Língua Portuguesa e Matemática, padronizada para um indicador entre 0 e 10, dos alunos da unidade j , obtida em determinada edição do exame realizado ao final da etapa de ensino;

P_{ji} = indicador de rendimento baseado na taxa de aprovação da etapa de ensino dos alunos da unidade j ;

Sendo que $0 \leq j \leq 10$ N; $0 \leq j \leq 1$ P e $0 \leq j \leq 10$.

Após a realização das avaliações do SAEB, os resultados das proficiências em Língua Portuguesa e Matemática são convertidos em uma escala de zero a dez. Logo após, a média referente aos dois testes é multiplicada pela taxa de aprovação, que pode variar de 0% a 100%. O resultado desse cálculo é o índice IDEB. Como exemplo:

a) Média das notas SAEB = 7,0

b) Taxa de aprovação = 75%, então

$$\text{IDEB} = 7 \times 0,75$$

$$\text{IDEB} = 5,2$$

Os resultados do IDEB podem regular e monitorar as políticas educacionais em território brasileiro, e, com isso, proporcionar motivação para o alcance das metas estipuladas pelo governo federal, além de proporcionar reflexões sobre as metodologias e práticas pedagógicas desenvolvidas no ambiente escolar (FERNANDES, 2007).

Como próximo passo, serão descritos os sistemas de avaliações de larga escala do SAEB e do PISA.

2.6 Avaliações em Larga Escala

No ambiente escolar existem as avaliações internas, que são indispensáveis para o desenvolvimento das práticas pedagógicas. Essas avaliações são contínuas e tem como objetivo principal acompanhar o desempenho dos alunos nas proficiências que compõem o currículo educacional da escola.

Além das avaliações internas existem também as avaliações externas. Essas avaliações, normalmente em larga escala, têm por objetivo, além acompanhar o desempenho dos alunos, também proporcionar metodologia de análise para entender o nível de qualidade da educação, utilizando para isso os desempenhos dos alunos nas proficiências analisadas nessas avaliações.

No Brasil, o Sistema de Avaliação da Educação Básica – SAEB contempla as avaliações aplicadas para os alunos matriculados no Ensino Fundamental ou Médio, e a partir de 2019 também contemplará a avaliação para o nível de Educação Infantil. Além do SAEB, que é desenvolvido pelo Governo Federal, existem ainda as avaliações em larga escala estaduais, que possuem objetivos semelhantes. No entanto, o desenvolvimento e controle são de responsabilidade de cada estado.

O Brasil também participa de ações internacionais relacionadas às avaliações em larga escala, sejam elas para alunos ou para os docentes. Como exemplos, o *Programme for International Student Assessment - PISA*, o *Teaching and Learning International Survey – TALIS* e os Estudos Regionais Comparativos – ERCE.

Importante salientar que todas as avaliações em larga escala que o Brasil disponibiliza ou participa têm por finalidade avaliar a qualidade da educação, e os resultados deverão ser utilizados para que o governo desenvolva políticas públicas que proporcionem condições suficientes de melhorias em todo o sistema de ensino.

2.6.1 PISA

O Brasil participa do Programa Internacional de Avaliação de Estudantes – PISA (em inglês *Programme for International Student Assessment*) desde sua criação em 1998 e, de certa forma, os resultados dessas avaliações contribuem como mais uma ferramenta de análise em relação à evolução da qualidade da educação brasileira. É considerado um modelo de avaliação em larga escala,

aplicado de forma amostral para os alunos matriculados a partir do 7º ano do Ensino Fundamental. A faixa etária base é de 15 anos, pois se pressupõe que com essa idade o aluno já tenha finalizado o nível escolar básico obrigatório (OCDE, 2020).

No Brasil, quem coordena as ações do PISA é o INEP. As avaliações do PISA são trienais e contemplam Leitura, Matemática e Ciências. Cada edição da avaliação do PISA tem ênfase em uma das áreas do conhecimento utilizadas; em 2009 a ênfase foi a Leitura, em 2012 Matemática, em 2015 Ciências e em 2018 foi novamente Leitura. Uma novidade na edição de 2015 foi a inclusão de questões relacionadas às áreas de Competência Financeira e Resolução Colaborativa de Problemas.

De forma semelhante ao sistema SAEB, o PISA tem por finalidade produzir conhecimento para promover debates relativos à qualidade da educação nos países membros da OCDE. Além da preocupação com a qualidade de ensino, existem também observações, por meio de questionários, sobre qual carreira profissional os jovens estão vislumbrando (INEP, 2020f; OCDE, 2020).

Também de forma análoga ao SAEB, são coletados por meio de formulários específicos dados dos alunos, professores e representantes das escolas. Esses dados servem para a confecção de indicadores que analisam as questões demográficas, socioeconômicas e educacionais das dimensões analisadas.

Na edição de 2018 participaram da avaliação PISA 37 países membros da OCDE e 42 países parceiros. Dos países da América do Sul o Brasil é o único que participa desde a primeira edição. (INEP, 2020g).

Os resultados do Brasil na última edição do PISA, em 2018, ficaram abaixo da média da OCDE. Os resultados apresentados no Quadro 4 indicam a necessidade do desenvolvimento de ações envolvendo os governos e os profissionais da educação, que possam promover melhorias em todo o sistema de ensino brasileiro. Apenas como efeito ilustrativo, o Quadro 4 apresenta 10 países que obtiveram as melhores classificações em cada proficiência analisada, além disso, os desempenhos referentes aos alunos brasileiros.

Quadro 4. Cinco países melhores classificados conforme resultados das avaliações do PISA 2018.

Ranking	Leitura		Matemática		Ciências	
	País	Pts	País	Pts	País	Pts
1	B-S-J-Z (China)	555	B-S-J-Z (China)	591	B-S-J-Z (China)	590
2	Singapore	549	Singapore	569	Singapore	551
3	Macao (China)	525	Hong Kong (China)	558	Macao (China)	544
4	Hong Kong (China)	524	Macao (China)	551	Estonia	530
5	Estonia	523	Chinese Taipei	531	Japan	529
6	Canada	520	Japan	527	Finlândia	522
7	Finlândia	520	Korea	526	Korea	519
8	Ireland	518	Estonia	523	Canada	518
9	Korea	514	Netherlands	519	Hong Kong (China)	517
10	Poland	512	Poland	516	Chinese Taipei	516
	Média OCDE	487	Média OCDE	489	Média OCDE	489
	Brasil (57 ^º)	413	Brasil (70 ^º)	384	Brasil (66 ^º)	404

Fonte: Adaptado do Relatório PISA (OCDE, 2020).

De acordo com os resultados apresentados no Quadro 4, pode-se verificar que B-S-J-Z (China) obteve a primeira colocação nas três proficiências analisadas no PISA de 2018, com 555, 590 e 591 pontos respectivamente para Leitura, Matemática e Ciências. Esta avaliação representa um forte progresso em relação à anterior, de 2015, em que os jovens obtiveram 494, 531 e 518 pontos nas mesmas matérias. A evolução dos resultados apresentados entre a edição de 2015 e 2018 podem indicar um nível maior de compromisso desse país com a educação (OCDE, 2020). Já no Brasil, nesse mesmo período, houve um retrocesso em Matemática e Ciências e um pequeno avanço em Leitura. Na edição de 2015, as posições alcançadas pelo Brasil foram, em Leitura 59^a, em Matemática 66^a e em Ciências a 63^a, respectivamente. Na edição de 2018 as classificações foram 57^a, 70^a e 66^a, mostrando com isso uma estagnação numa posição considerada como muito ruim.

Em relação ao desempenho dos alunos brasileiros, pode-se observar que as classificações alcançadas reforçam a necessidade de ações para a melhoria da qualidade da educação, e é nesse ponto que o governo precisa destinar mais atenção e investimentos para essa área.

O próximo item descreve como é constituído o sistema de educação no Brasil, pois se torna necessário entender quais são os cursos que compõem esse sistema e de quem responde pela manutenção da oferta de vagas nesses cursos.

2.6.2 Sistema SAEB

O governo brasileiro, por meio do INEP, iniciou em 1990 um processo de avaliação de aprendizagem com intuito de observar a qualidade do sistema de ensino, por meio de testes direcionados aos alunos matriculados em toda rede de ensino, envolvendo escolas públicas e particulares. Neste ano, criou o Sistema de Avaliação da Educação Básica - SAEB, que tinha como objetivo primordial avaliar como a educação básica brasileira estava contribuindo para a formação educacional dos indivíduos.

Em 2005, o governo decidiu reestruturar o sistema de avaliação de aprendizagem e criou dois sistemas de avaliação, um denominado Avaliação Nacional de Educação Básica - ANEB que manteve as características iniciais do SAEB, e outra avaliação denominada Avaliação Nacional do Rendimento Escolar - ANRESC, conhecida como PROVA BRASIL, que tem como objetivo avaliar a qualidade de ensino das escolas públicas, no que se refere aos desempenhos dos alunos nas proficiências avaliadas. Em 2013, outra avaliação foi incorporada ao sistema SAEB, Avaliação Nacional da Alfabetização - ANA, e desde a sua criação tem como objetivo medir o desempenho na alfabetização e letramento em Língua Portuguesa e Matemática. Para a edição de 2019 as siglas ANA, ANEB e ANRESC não serão mais utilizadas, com isso todos os níveis de avaliação serão identificados apenas como Avaliação SAEB (INEP, 2020a).

Esse sistema evoluiu desde 1990, principalmente no que se refere ao público-alvo para a realização dos testes, bem como na estrutura da avaliação (base do conhecimento). A cada edição do SAEB o Governo Federal vem propondo alterações que têm por finalidade melhorar os aspectos das avaliações, seja para ampliar o volume de alunos avaliados, seja para melhorar a coleta de dados por meio dos formulários aplicados ou, ainda, para estruturar os itens avaliados em cada nível de proficiência.

Uma alteração importante já na edição de 2017 foi a ampliação do rol de alunos avaliados no Ensino Médio, o que antes era amostral, a partir de 2017 passou a ser censitária. Dessa forma, todos os alunos de escolas públicas devem participar da avaliação SAEB. As características da Avaliação SAEB, a partir do ano de 2019, estão descritas no Quadro 5.

Quadro 5. Características do Sistema SAEB a partir de 2019.

Avaliação	Abrangência	Formulação dos Itens	Área do Conhecimento/Disciplinas Avaliadas
Creche e pré-escola da Educação Infantil	Escolas públicas (amostral) - Estudo piloto	BNCC	-
2º ano do Ensino Fundamental	Escolas públicas (amostral) Escolas privadas (amostral)	BNCC	Língua portuguesa e matemática
5º e 9º ano do Ensino Fundamental	Escolas públicas (censitário) Escolas privadas (amostral)	Matriz de referência	Língua portuguesa e matemática
9º ano do Ensino Fundamental	Escolas públicas (amostral) Escolas privadas (amostral)	BNCC	Ciências da natureza e ciências humanas
3ª e 4ª séries do Ensino Médio	Escolas públicas (censitário) Escolas privadas (amostral)	Matriz de referência	Língua portuguesa e matemática
<p>-A avaliação da alfabetização será realizada no 2º ano do Ensino Fundamental de forma amostral.</p> <p>-Começa a avaliação da Educação Infantil, em caráter de estudo piloto, com aplicação de questionários eletrônicos exclusivamente para professores e diretores.</p> <p>-Amostra de estudantes do 9º ano do Ensino Fundamental para avaliar Ciências da Natureza e Ciências Humanas</p> <p>-A Base Nacional Comum Curricular (BNCC) passa a ser referência na formulação dos itens do 2º ano (Língua Portuguesa e Matemática) e do 9º ano (Ciências da Natureza e Ciências Humanas)</p>			

Fonte: Adaptado pelo autor (INEP, 2020a)

O INEP integrou os resultados da Prova Brasil e do Censo Escolar para viabilizar o Indicador de Desenvolvimento da Educação Básica - IDEB. Esse indicador tem como um dos princípios básicos que o aluno aprenda e passe de ano, enquanto que o desempenho é medido por meio da Prova Brasil e a aprovação pelo Censo Escolar.

Os resultados das avaliações são divulgados no próprio site do INEP, (INEP, 2020c), e podem ser utilizados para análises e pesquisas acadêmicas, entre outros estudos. Como os resultados apontam as médias alcançadas pelos alunos em nível nacional, não é possível observar diretamente nesses dados os fatores ou dificuldades de cada região ou unidade escolar, o que pode comprometer de certa forma uma análise mais profunda sobre a evolução da qualidade da educação.

Além das avaliações descritas no Quadro 5, o SAEB também coleta dados por meio de formulários direcionados aos Alunos, aos Diretores de escola, aos Representantes de escola e também aos Professores. Esses dados coletados têm por finalidade observar questões relacionadas às percepções do ambiente escolar, além de proporcionar um entendimento sobre as questões relativas ao nível sócio

demográfico dos alunos. O Quadro 6 descreve o número de questões de cada formulário presente nas avaliações do SAEB.

Quadro 6. Questões dos formulários SAEB

Destino (Atores)	Questões
Alunos 5º ano Ensino Fundamental	51
Alunos 9º ano Ensino Fundamental	57
Alunos 3ª/4ª séries do Ensino Médio	60
Diretores de Escola	111
Professores	125
Escola (Preenchido por um representante)	74

Fonte: (INEP, 2020a), adaptado pelo autor

Os dados coletados por meio dos formulários do SAEB, descritos no Quadro 6, são disponibilizados juntamente com os resultados das avaliações, no portal do INEP. Os dados são para consulta pública, e de acordo com o governo, servem como informação e conhecimento sobre como está evoluindo a qualidade da educação brasileira (INEP, 2020a). O total de questões disponíveis para a avaliação do Ensino Médio é de 370, incluindo os formulários dos Alunos, Diretores, Professores e Escola.

Com os resultados das avaliações do SAEB, compostos pelo desempenho dos alunos e pelos dados dos formulários preenchidos por todos os envolvidos, o governo federal desenvolve relatórios de análises para verificar a evolução da qualidade da educação brasileira, os quais também são utilizados para compor planos de ação que tenham como objetivo promover melhorias em todo sistema educacional. Como exemplos desses relatórios são o Painel Educacional Municipal e Estadual, o Press Kit e o Indicadores Educacionais, sendo que esses e outros relatórios podem ser consultados diretamente do website do INEP (INEP, 2020h).

O próximo item ilustra a estrutura da educação em território brasileiro.

3 ESTRUTURA DA EDUCAÇÃO NO BRASIL

A Lei de nº 9.394 de Diretrizes e Bases da Educação Nacional - LDB, (BRASIL, 1996) é a que define toda estrutura do sistema de educação brasileiro. São apresentadas nessa lei a organização da educação, incluindo as finalidades dos órgãos administrativos, níveis e modalidades de ensino e a regulamentação do sistema educacional, tendo como texto base a Constituição Federal. O Quadro 7 ilustra as responsabilidades dos diferentes órgãos em cada nível governamental, de uma maneira geral.

Quadro 7 Órgãos responsáveis pela Educação

Nível	Órgãos	Níveis de Educação
Federal	Ministério da Educação – MEC Conselho Nacional de Educação - CNE	- Educação Infantil; - Ensino Fundamental;
Estadual	Secretaria Estadual de Educação – SEE Conselho Estadual de Educação - CEE Delegacia Regional de Educação – DRE	- Ensino Médio; -Educação Profissio- nalizante;
Municipal	Secretaria Municipal de Educação – SME Conselho Municipal de Educação - CME	- EJA - Ensino Superior.

Fonte: Lei de Diretrizes e Bases da Educação Nacional (BRASIL, 1996)

Cada órgão descrito no Quadro 7 possui atribuições inerentes à sua amplitude administrativa e deve ter como base a LDB, a Constituição Federal, as instruções normativas e ainda os planos e leis divulgados pelo Ministério da Educação, como exemplo o PNE, descrito no item 2.4, que é base para os planos estaduais.

No Brasil, segundo o art. 21 da LDB, a educação escolar básica está estruturada de forma a atender os indivíduos até a fase da educação superior, compreendendo os níveis de ensino referentes à Educação Infantil, Ensino Fundamental, Ensino Médio, Educação de Jovens e Adultos e Ensino Superior (BRASIL, 1996). A oferta de vagas para os níveis da educação brasileira pode ser feita pelos órgãos públicos responsáveis pela educação (governo) de forma gratuita ou pela iniciativa privada.

Em relação ao ingresso na educação e o progresso entre ciclos, normalmente no Brasil, a criança inicia na fase da Educação Infantil que vai até a pré-escola, em

seguida com idade de 7 anos é iniciado o ciclo do Ensino Fundamental, que compreende os cursos do 1º ao 9º ano. Na maioria dos casos, o jovem inicia o Ensino Médio, que é composto por 3 séries, com 15 anos de idade. Em alguns estados também é ofertada a 4ª. série do Ensino Médio.

Ainda de acordo com a LDB, o governo prevê o nível de educação profissional ou técnica, com o objetivo de preparar o jovem para o mercado de trabalho, e a educação especial, que atende especificamente indivíduos com deficiências, transtornos globais do desenvolvimento e altas habilidades ou superdotação.

Cada esfera governamental deve ser responsável pelas instituições de ensino que estão sob sua responsabilidade, seja de Educação Infantil, Ensino Fundamental, Ensino Médio ou Superior. A Figura 2 ilustra os níveis e modalidades previstas na LDB.

Figura 2. Níveis e Modalidades da Educação/Ensino



Fonte: Adaptado da LDB pelo autor (BRASIL, 1996).

Os custos para manter o sistema educacional funcionando perfeitamente devem estar previstos no orçamento de cada governo, seja ele municipal, estadual ou federal. Com isso, cada esfera governamental deve alinhar os custos da educação com a META 20 do Plano Nacional da Educação, que prevê investimentos na área de no mínimo 7% do PIB.

3.1 As instituições de ensino no país

Quanto maior for a escola em relação à sua capacidade de atendimento, maior é a complexidade de gestão. Existem escolas que atendem nos três turnos e com os diversos níveis de ensino (infantil, fundamental, médio, técnicos e EJA), portanto o corpo administrativo/pedagógico deve estar preparado para as especificidades de cada nível educacional.

Como exemplo de complexidade de gestão escolar, numa situação em que uma escola tenha mais de 500 matrículas, funcionando em 3 turnos e oferte 4 ou mais etapas de ensino, a complexidade de gestão pode ser classificada como nível 6, conforme descrito no Quadro 8. Com esse tipo de cenário, se torna necessário um modelo de gestão eficiente para cada etapa de ensino, pois os alunos possuem necessidades educacionais distintas, além de se fazer necessário diferentes planos pedagógicos, gestão de espaço contemplando as necessidades de cada faixa etária, capacitação de professores para suas atividades específicas, entre outros. Sendo assim, as escolas com complexidade mais alta impactam fortemente o trabalho do Diretor

O Quadro 8 ilustra a situação da complexidade de gestão escolar no território brasileiro no ano de 2019, incluindo todas as escolas presentes nas regiões brasileiras, sendo de administração pública ou privada.

Quadro 8. Percentual de escolas por nível do indicador de complexidade de gestão e dependência administrativa – Brasil e Regiões Geográficas - 2019

Unidade Geográfica	Percentual de escolas segundo nível do indicador complexidade de gestão					
	Nível 1	Nível 2	Nível 3	Nível 4	Nível 5	Nível 6
Brasil	26,2	30,7	19,2	12,8	8,5	2,6
Norte	24,4	29,6	25,1	8,9	9,6	2,4
Nordeste	24	31,4	17,9	11,5	12,9	2,3
Sudeste	29,2	31,6	16,5	14,4	5,6	2,7
Sul	29,5	27,5	22,1	14	4,4	2,5
Centro-Oeste	18,3	32	22,7	15,6	7,2	4,2

Os níveis de complexidade de gestão são descritos abaixo de acordo com as características predominantes das escolas pertencentes a cada um deles:

- Nível 1 - Escolas que, em geral, possuem porte inferior a 50 matrículas, funcionam em único turno, ofertam uma única etapa de ensino e apresentam a Educação Infantil ou os Anos Iniciais como etapa mais elevada.

- Nível 2 - Escolas que, em geral, possuem porte entre 50 e 300 matrículas, funcionam em 2 turnos, com oferta de até 2 etapas de ensino e apresentam a Educação Infantil ou os Anos Iniciais como etapa mais elevada.
 - Nível 3 - Escolas que, em geral, possuem porte entre 50 e 500 matrículas, funcionam em 2 turnos, com oferta de 2 ou 3 etapas de ensino e apresentam os Anos Finais como etapa mais elevada*.
 - Nível 4 - Escolas que, em geral, possuem porte entre 150 e 1000 matrículas, funcionam em 2 ou 3 turnos, com oferta de 2 ou 3 etapas de ensino e apresentam o Ensino Médio, a Educação Profissional ou a EJA como etapa mais elevada.
 - Nível 5 - Escolas que, em geral, possuem porte entre 150 e 1000 matrículas, funcionam em 3 turnos, com oferta de 2 ou 3 etapas de ensino e apresentam a EJA como etapa mais elevada.
 - Nível 6 - Escolas que, em geral, possuem porte superior a 500 matrículas, funcionam em 3 turnos, com oferta de 4 ou mais etapas de ensino e apresentam a EJA como etapa mais elevada.
- Considerou-se como a etapa mais elevada ofertada pela escola aquela que atenderia, teoricamente, alunos com idade mais elevada.

Fonte: (INEP, 2020i)

Observar a qualidade da educação em escolas públicas e privadas em todo território nacional passa a ser uma tarefa desafiadora, pois a quantidade de instituições de ensino públicas existentes exige um sistema de controle eficiente para que os recursos destinados às escolas não se percam ou sejam utilizados para outras finalidades. O Brasil conta hoje com um total de 47.874.246 alunos matriculados em 180.610 escolas, distribuídos nos vários níveis da educação básica. Como ilustração para a quantidade de escolas e de alunos matriculados nos cursos referentes à Educação Infantil, Ensino Fundamental e Ensino Médio, em território nacional, o Quadro 9 descreve os totais já separados por curso.

Quadro 9 Total de escolas e de alunos matriculados, no Brasil.

2019	Ed. Infantil		Ens. Fundamental		Ensino Médio		
	Creche	Pré - Escola	Anos Iniciais	Anos Finais	Propedêutico	Magistério	Técnico Integrado
Total de Escolas	71.403	102.335	109.644	61.765	27.481	692	2.134
Total de Matrículas	3.755.092	5.217.686	15.018.498	11.905.232	6.842.713	64.222	558.956

Fonte: Adaptado pelo autor com base nos dados do INEP(INEP, 2020j)

Alunos dessas instituições devem participar das avaliações em larga escala ao finalizarem seus ciclos da educação, ou seja, do 5º e 9º anos do Ensino Fundamental e 3ª ou 4ª séries do Ensino Médio. Além das avaliações em larga escala nacional, cada estado também tem seu modelo de avaliação, e os alunos são convocados para essas provas mediante calendários estaduais.

De acordo com o Quadro 9 pode-se perceber que existe uma concentração de alunos nos cursos referentes aos Anos Iniciais e Finais do Ensino Fundamental, pois quando os alunos encerram esse ciclo, nem todos continuam seus estudos nos cursos do Ensino Médio. Esse abandono, algumas vezes, é causado pela necessidade que o jovem tem em exercer uma atividade funcional, para ajudar na renda familiar.

3.2 As avaliações SAEB nos últimos anos

Como já ressaltado, todos os alunos egressos dos cursos do 5º e 9º anos do Ensino Fundamental e da 3ª ou 4ª séries do Ensino Médio devem realizar as avaliações do SAEB. Os resultados apresentados pelos alunos nessas avaliações servem de parâmetro inicial para que os governos observem como o nível de qualidade da educação está evoluindo, além de fazer parte para a constituição do Índice de Desenvolvimento da Educação Básica (INEP, 2020a).

Para melhor entendimento das últimas avaliações SAEB, vale compreender os valores e cálculos desse sistema avaliativo. O governo federal calcula e atribui o Desempenho Médio Estadual a partir dos resultados do conjunto de alunos de cada estado. Os valores obedecem uma escala que varia entre 0 e 500 pontos, com isso, quanto maior a média, melhor foi o desempenho dos alunos nas avaliações do SAEB. Essas médias podem indicar como cada estado está evoluindo em nível de aprendizado dos seus alunos, tanto na proficiência Língua Portuguesa quanto em Matemática. O Quadro 10 ilustra os desempenhos para a proficiência Língua Portuguesa. Até o momento do encerramento desse estudo, em 30/11/2020, o governo só tinha disponibilizado os resultados referentes às médias estaduais da edição de 2019 do SAEB, descritos nos Quadros 10 e 12. Os resultados dos desempenhos dos alunos e das escolas ainda não tinham sido divulgados, por conta disso não foram analisados.

Quadro 10. Médias estaduais para o Ensino Médio referente à proficiência Língua Portuguesa.

Língua Portuguesa									
Estados	2005	2007	2009	2011	2013	2015	2017	2019	Dif. 2019- 2005
BRASIL	257,60	261,39	268,83	267,63	263,18	267,06	267,61	278,36	20,76
Rondônia	257,67	257,77	271,11	268,02	261,67	266,88	267,50	278,25	20,58
Acre	252,58	259,59	264,29	254,69	259,69	262,05	263,60	272,12	19,54

Amazonas	231,27	238,21	250,57	253,60	242,23	259,27	249,50	258,62	27,35
Roraima	256,87	251,33	260,81	257,48	253,91	260,49	267,50	268,01	11,14
Pará	243,17	245,41	257,25	251,09	245,56	254,92	245,10	259,94	16,77
Amapá	253,16	249,01	254,93	257,92	250,32	255,74	251,20	261,16	8,00
Tocantins	240,45	244,08	249,92	258,37	249,05	253,81	260,10	266,71	26,26
Maranhão	231,66	241,24	247,33	245,97	239,11	246,34	250,60	259,73	28,07
Piauí	250,38	242,97	250,91	253,38	249,42	255,05	256,50	272,17	21,79
Ceará	257,79	256,79	263,06	261,75	255,46	256,63	265,90	275,73	17,94
Rio Grande do Norte	241,67	249,68	256,92	254,79	245,52	247,31	253,20	265,46	23,79
Paraíba	239,52	250,28	260,82	257,79	253,52	258,01	258,40	276,18	36,66
Pernambuco	250,17	246,49	255,27	253,09	261,37	270,23	268,70	282,45	32,28
Alagoas	243,28	243,72	251,16	246,93	241,23	250,61	256,00	267,95	24,67
Sergipe	260,18	242,09	261,85	255,46	252,12	257,10	265,40	272,57	12,39
Bahia	243,14	250,13	264,80	253,92	238,75	250,37	248,90	263,91	20,77
Minas Gerais	268,71	276,71	274,78	275,94	268,38	268,67	275,50	284,27	15,56
Espírito Santo	267,72	261,67	278,39	265,74	265,28	277,44	283,70	293,14	25,42
Rio de Janeiro	257,03	260,51	268,33	274,57	277,24	276,11	274,10	280,97	23,94
São Paulo	261,57	268,77	273,68	278,61	275,58	274,65	274,00	286,54	24,97
Paraná	267,16	271,94	280,92	271,72	262,55	273,31	274,30	290,85	23,69
Santa Catarina	266,99	268,90	275,17	279,86	272,91	276,58	275,10	284,32	17,33
Rio Grande do Sul	282,47	281,49	289,57	280,12	278,27	272,78	278,50	289,67	7,20
Mato Grosso do Sul	272,56	274,43	285,47	283,70	273,16	279,10	272,10	284,71	12,15
Mato Grosso	255,73	253,80	260,38	261,39	255,05	264,12	262,00	271,73	16,00
Goiás	253,08	251,95	265,59	269,17	267,20	269,73	276,20	290,95	37,87
Distrito Federal	278,22	288,41	278,85	282,72	278,89	284,20	278,30	292,65	14,43

Fonte: Desenvolvido pelo autor com base nos dados do INEP (INEP, 2020c).

Verifica-se que todos os estados apresentaram evoluções positivas para o período analisado, com evolução importante de 2005 a 2009, um período de estagnação de 2009 a 2017 e finalmente resultados positivos entre 2017 e 2019. Entre os estados, a maior evolução para o período 2005 a 2019 foi apresentada pelo estado de Goiás, com valor de 37,87. E com menor evolução está o estado do Rio Grande do Sul, com valor de 7,20.

A média nacional apresentou, em 2019, o valor de 278,36. Essas escalas variam entre 0 e 500 pontos e estão divididas em 8 níveis para Língua Portuguesa. O nível 1 compreende os valores de desempenhos entre 225 e 250 pontos. A partir desse nível as escalas vão evoluindo a cada 25 pontos até completar 400 pontos, no caso da Língua Portuguesa. Para Matemática, a escala vai até 500 pontos, como será visto adiante.

Dessa forma, vale ressaltar que a escala de desempenho para a proficiência Língua Portuguesa tem 8 níveis, sendo que o último compreende os desempenhos

iguais ou superiores a 400 pontos. Já a escala para Matemática evolui sequencialmente até o nível 10, e nesse nível são incluídos os alunos com desempenhos iguais ou superiores a 450 pontos.

Os desempenhos apresentados pelos alunos que compreendem entre 0 e 225 pontos são classificados pelo governo como Nível Zero. O governo ainda salienta que os alunos classificados nesse nível requerem uma maior atenção, pois ainda não demonstram habilidades que deveriam apresentar nesse nível de formação. As escalas de desempenhos estão incluídas como Anexo 2 desse estudo.

Assim, apenas para contextualização do Quadro 10, as médias indicam que, em 2019, 14 estados estariam com desempenhos compatíveis com o Nível 3 da escala e 13 estão compatíveis com o Nível 2. Os níveis correspondentes de conhecimento e aprendizado apresentados pelos alunos nesses níveis são apresentados no Quadro 11.

Quadro 11. Relação de competências em Língua Portuguesa, para os níveis de 1 a 3 na escala de desempenho

<p>Nível 1 Desempenho maior ou igual a 225 e menor que 250</p>	<p>Os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Identificar elementos da narrativa em história em quadrinhos. • Reconhecer a finalidade de recurso gráfico em artigos. • Reconhecer a relação de causa e consequência em lendas. • Inferir o sentido de palavra em letras de música e reportagens.
<p>Nível 2 Desempenho maior ou igual a 250 e menor que 275</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Reconhecer a ideia comum entre textos de gêneros diferentes e a ironia em tirinhas. • Reconhecer relações de sentido estabelecidas por conjunções ou locuções conjuntivas em letras de música e crônicas. • Reconhecer o uso de expressões características da linguagem (científica, profissional etc.) e a relação entre pronome e seu referente em artigos e reportagens. • Inferir o efeito de sentido da linguagem verbal e não verbal em notícias e charges.
<p>Nível 3 Desempenho maior ou igual a 275 e menor que 300</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Localizar informação explícita em artigos de opinião. • Identificar a finalidade de relatórios científicos. • Reconhecer relações de sentido marcadas por conjunções, a relação de causa e consequência e a relação entre o pronome e seu referente em fragmentos de romances. • Reconhecer o tema de uma crônica. • Reconhecer variantes linguísticas em artigos. • Reconhecer o sentido e o efeito de sentido produzido pelo uso de recursos morfosintáticos em contos, artigos e crônicas. • Reconhecer opiniões divergentes sobre o mesmo tema em diferentes textos. • Inferir informação, o sentido e o efeito de sentido produzido por expressão em reportagens e tirinhas.

Fonte: Desenvolvido pelo autor com base nos dados do INEP (INEP, 2020c).

Com base nos resultados dos desempenhos apresentados pelos alunos em comparação com a escala de desempenho, os governos podem desenvolver ações

que envolvam práticas pedagógicas direcionadas aos itens com maior necessidade de reforço, de forma a proporcionar melhores condições de aprendizado.

Já para a proficiência Matemática, o Quadro 12 mostra que não houve uma evolução de desempenhos semelhante à proficiência Português. Enquanto a evolução média nacional em Português foi de 20,75 pontos, em Matemática as avaliações melhoraram somente 6,05 pontos.

Quadro 12. Médias Estaduais para o Ensino Médio referente a proficiência Matemática.

Matemática									
Estados	2005	2007	2009	2011	2013	2015	2017	2019	Dif. 2019-2005
BRASIL	271,29	272,89	274,72	273,86	269,25	267,00	269,74	277,34	6,05
Rondônia	271,13	271,16	275,46	275,81	270,22	265,26	271,20	278,81	7,68
Acre	257,10	264,85	261,53	257,12	258,03	256,97	262,60	270,18	13,08
Amazonas	241,99	247,43	251,40	253,49	245,00	257,11	246,80	251,70	9,71
Roraima	268,49	259,60	263,87	262,36	255,98	259,66	271,20	265,66	-2,83
Pará	248,70	252,64	256,76	249,55	247,41	254,30	245,50	254,49	5,79
Amapá	260,28	252,21	254,61	252,84	250,14	252,15	247,20	254,30	-5,98
Tocantins	253,56	252,07	251,98	263,43	254,73	256,07	264,70	269,19	15,63
Maranhão	239,90	246,19	247,00	244,27	241,99	247,60	248,60	253,47	13,57
Piauí	260,49	253,78	256,11	257,93	256,88	256,14	261,70	273,09	12,60
Ceará	265,95	265,31	267,31	264,97	260,34	260,64	267,10	272,82	6,87
Rio Grande do Norte	254,90	257,90	259,67	258,70	250,48	251,40	254,50	262,55	7,65
Paraíba	253,58	261,03	264,16	258,83	257,89	257,67	260,60	274,75	21,17
Pernambuco	255,42	257,75	258,81	256,58	266,17	267,77	270,90	281,15	25,73
Alagoas	258,99	251,76	254,81	248,23	247,79	252,67	257,20	263,83	4,84
Sergipe	271,20	253,29	266,21	260,04	259,28	258,19	268,90	271,86	0,66
Bahia	262,05	261,30	266,89	257,12	245,95	251,10	250,60	261,36	-0,69
Minas Gerais	290,38	289,35	285,71	288,57	279,29	272,07	281,10	287,54	-2,84
Espírito Santo	282,39	277,45	286,96	281,98	278,84	281,04	291,60	298,49	16,10
Rio de Janeiro	268,91	271,87	277,01	286,45	282,98	273,97	276,70	280,56	11,65
São Paulo	272,92	279,43	278,10	282,90	278,07	272,98	274,10	283,93	11,01
Paraná	284,77	290,15	290,77	283,16	272,87	273,22	279,00	293,30	8,53
Santa Catarina	286,55	287,77	286,99	295,31	286,59	277,97	283,60	289,56	3,01
Rio Grande do Sul	306,12	296,83	304,81	294,89	290,00	273,31	282,70	291,41	-14,71
Mato Grosso do Sul	283,13	288,58	291,40	290,76	279,08	275,85	274,90	285,41	2,28
Mato Grosso	263,56	264,57	263,69	266,80	261,59	263,23	264,90	271,33	7,77
Goiás	266,76	265,60	271,39	276,17	274,56	270,18	278,90	291,79	25,03
Distrito Federal	297,83	300,31	285,65	289,73	287,02	280,13	286,20	296,45	-1,38

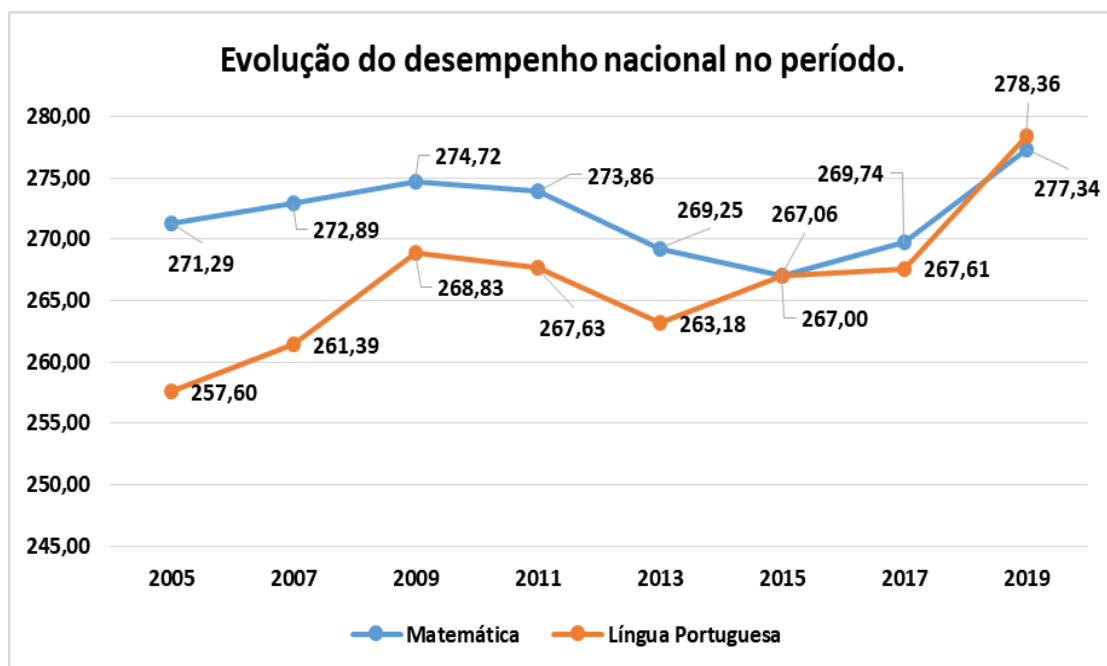
Fonte: Desenvolvido pelo autor com base nos dados do INEP (INEP, 2020c).

Diferente dos Desempenhos Médios para Língua Portuguesa, para a proficiência Matemática há estados que apresentaram evoluções negativas para o

período de 2005 a 2019. O maior valor de evolução negativa apresentado foi o do estado do Rio Grande do Sul, com -14,71. Já a maior evolução positiva foi a do estado de Pernambuco com 25,73.

Os valores de evolução positiva para a proficiência Matemática são menores do que os apresentados para Língua Portuguesa, isso reflete que pode haver dificuldades de aprendizado pelos alunos em disciplinas relacionadas à área de exatas, necessitando de mais investigações nesse tema. Com isso, os governos estaduais devem propor ações que possam proporcionar melhorias de aprendizado dos alunos nessas disciplinas e, com isso, inclusive, incentivar a formação educacional dos jovens em áreas relacionadas com o STEM (Ciências, Tecnologia, Engenharia e Matemática). O Gráfico 1 ilustra a evolução dos desempenhos nacionais no período de 2005 a 2019, para Língua Portuguesa e Matemática.

Gráfico 1 Evolução do Desempenho nacional no período de 2005 a 2019, para Língua Portuguesa e Matemática.



Fonte: Elaborado pelo Autor (2020).

Utilizando a mesma comparação com as escalas de desempenhos realizada para a proficiência Língua Portuguesa, a quantidade de estados classificados em 2019, como Nível 3 foi 12 e como Nível 2, o total foi de 15 estados. O Quadro 13 mostra as competências referentes a esses níveis da escala para Matemática.

Quadro 13. Relação de competências em Matemática, para os níveis de 1 a 3 na escala de desempenho

<p>Nível 1 Desempenho maior ou igual a 225 e menor que 250</p>	<p>Os estudantes provavelmente são capazes de: TRATAMENTO DE INFORMAÇÕES Associar uma tabela de até duas entradas a informações apresentadas textualmente ou em um gráfico de barras ou de linhas.</p>
<p>Nível 2 Desempenho maior ou igual a 250 e menor que 275</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: ESPAÇO E FORMA Reconhecer as coordenadas de pontos representados em um plano cartesiano localizados no primeiro quadrante. NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES Reconhecer os zeros de uma função dada graficamente. Determinar o valor de uma função afim, dada sua lei de formação. Determinar resultado utilizando o conceito de progressão aritmética. TRATAMENTO DE INFORMAÇÕES Associar um gráfico de setores a dados percentuais apresentados textualmente ou em uma tabela.</p>
<p>Nível 3 Desempenho maior ou igual a 275 e menor que 300</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES Reconhecer o valor máximo de uma função quadrática representada graficamente. Reconhecer, em um gráfico, o intervalo no qual a função assume valor máximo. Determinar, por meio de proporcionalidade, o gráfico de setores que representa uma situação com dados fornecidos textualmente. Determinar o quarto valor em uma relação de proporcionalidade direta a partir de três valores fornecidos em uma situação do cotidiano. Determinar um valor reajustado de uma quantia a partir de seu valor inicial e do percentual de reajuste. Resolver problemas utilizando operações fundamentais com números naturais.</p>

Fonte: Desenvolvido pelo autor com base nos dados do INEP (INEP, 2020c).

Com base nos resultados apresentados nos Quadros 10 e 12, fica evidente que os níveis de conhecimento dos alunos brasileiros estão ainda nos níveis iniciais das escalas, tanto para Língua Portuguesa quanto para Matemática. Percebe-se que os estados com melhores desempenhos não ultrapassaram o Nível 3.

Esses resultados apresentados podem proporcionar reflexões sobre como os planos de ensino estão sendo desenvolvidos e estruturados por cada estado brasileiro, como os conteúdos programáticos estão sendo desenvolvidos e ainda se todo o planejamento foi cumprido durante o ano escolar, se os professores estão recebendo orientações e treinamentos para trabalhar com casos mais específicos e se os gestores e as escolas possuem recursos necessários para promover um ambiente de ensino adequado para todos os alunos.

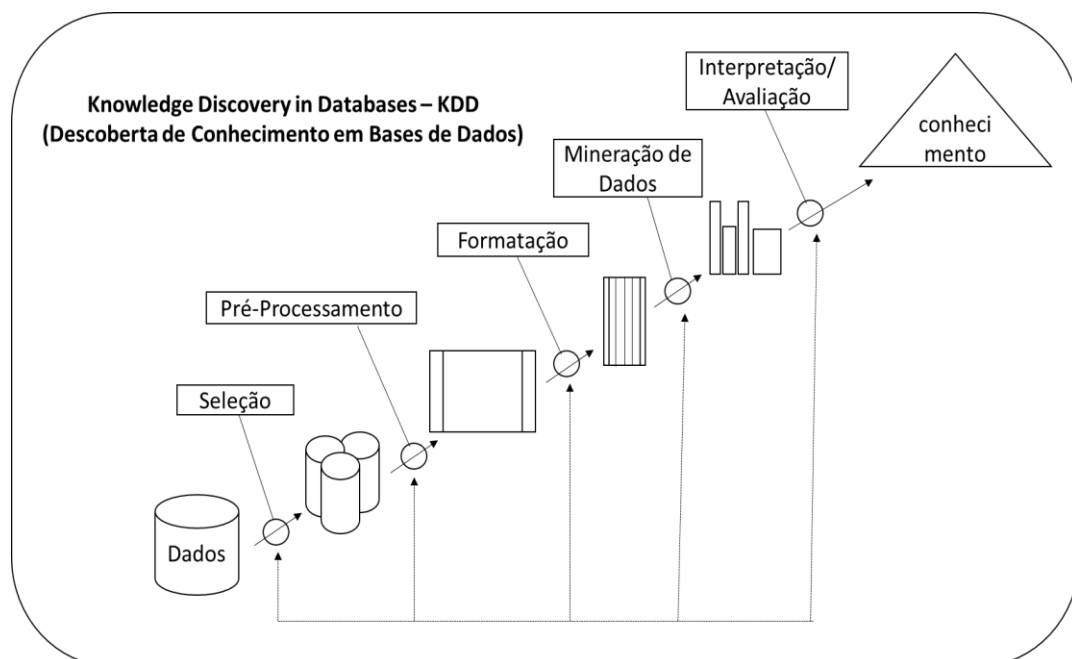
4 O PROCESSO DE KDD E A MINERAÇÃO DE DADOS.

Com o advento do avanço tecnológico e da necessidade do tratamento de grandes volumes de dados, surgem barreiras relacionadas ao tempo que podem dificultar a criação de conhecimento a partir desses volumes de dados. Praticamente, seria necessário um dispêndio enorme de energia e tempo para que grandes volumes de dados sejam devidamente analisados.

Como forma de auxiliar os processos de análise que envolvem grandes volumes de dados, surge a Mineração de Dados, que utiliza algoritmos e aprendizado de máquina para descobrir padrões e gerar conhecimentos, inclusive pode ser utilizada como apoio nos processos que envolvem tomada de decisão (PROVOST e FAWCETT, 2016; SILVA, PERES e BOSCARIOLI, 2016).

A mineração de dados é um dos processos de *Knowledge Discovery in Databases* – KDD, (Descoberta de Conhecimento em Bases de Dados), Figura 3, termo formalizado em 1989 com destaque para o conceito de procurar conhecimento em bases de dados, e que, de acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), pode ser entendido como “processo de várias etapas, não trivial, iterativo e interativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis de grandes conjunto de dados”.

Figura 3. Descoberta de Conhecimento em Bases de Dados - KDD



Fonte: Adaptado de Fayyad, Piatetsky-Shapiro e Smith (1996)

Sinteticamente, o processo para descoberta do conhecimento tem como início a análise da base de dados, uma fase denominada pré-processamento, a fase de mineração de dados, a fase de pós-processamento e finalmente o resultado, que é a informação (GOLDSCHMIDT, PASSOS e BEZERRA, 2015; SILVA, PERES e BOSCARIOLI, 2016).

Antes do processo de mineração de dados, deve-se realizar os processos de SELEÇÃO, PRÉ-PROCESSAMENTO e FORMATAÇÃO, que consistem em observar as bases de dados que serão utilizadas para a descoberta do conhecimento, separação dos dados necessários e úteis, e a formatação desses dados. Ainda como parte desses processos, são descartados os dados nulos, em branco e preenchidos erroneamente, e, após essa fase, o próximo passo é a Mineração de Dados. Nessa etapa deve-se definir quais técnicas ou algoritmos serão utilizados para que os objetivos sejam alcançados, como exemplos de técnicas de mineração podem ser a Redes Neurais, Algoritmos Genéticos e Modelos Estatísticos e Probabilísticos, entre outras (SILVA, PERES e BOSCARIOLI, 2016; BEZERRA, 2017; CASTRO e FERRARI, 2017).

De acordo com Holfmann e Klinkenberg (2013) e Castro e Ferrari (2017), a mineração de dados pode tornar mais prático e eficiente o apoio aos processos de tomada de decisão, fator importante na construção de ferramentas de *business intelligence*. A utilização da técnica de mineração de dados como ferramenta de apoio à tomada de decisão pode ser um grande diferencial para os processos de gestão, pois a partir dos resultados apresentados, os gestores podem entender como surgem determinados comportamentos, e, com isso, rapidamente adaptar seus planejamentos para solução de problemas no ambiente da empresa.

As etapas do processo de descoberta de conhecimento serão explanadas individualmente.

4.1 Pré-Processamento

A etapa de pré-processamento envolve as fases de Limpeza e Seleção dos dados. A finalidade dessa etapa é proporcionar uma melhor qualidade na transformação de dados, pois pode evitar distorções na produção do conhecimento. Segundo Han, Kamber e Pei (2011), as fases de pré-processamento compreendem:

- a) **Limpeza de Dados:** fase inicial que tem por objetivo remover os dados inconsistentes ou considerados incorretos para que se possa produzir um conjunto de dados com uma melhor qualidade.
- b) **Seleção e Transformação de Dados:** nessa fase são selecionados os dados considerados relevantes para que sejam utilizados na etapa de mineração de dados. Os dados selecionados podem ser transformados e consolidados em um formato considerado apropriado para a utilização na etapa de mineração.

Essa etapa deve ser executada até que se tenha condições de produzir um conjunto de dados que possa ser trabalhado na etapa de mineração de dados, e, com isso, gerar o conhecimento necessário para a tomada de decisão.

4.2 Mineração de Dados

A Mineração de Dados pode ser considerada uma atividade multidisciplinar pois envolve conhecimentos de diversas áreas, como exemplo pode se observar o uso de conceitos de banco de dados, estatística, aprendizado de máquina, inteligência artificial entre outras (PROVOST; FAWCETT, 2016). Segundo Han, Kamber e Pei (2011), o uso da Mineração de Dados pode proporcionar a descoberta de conhecimento que esteja escondido em um conjunto de dados ou repositórios de informações.

Os conceitos para Mineração de Dados podem variar de acordo com autores de cada área de conhecimento. Para Carvalho (2005) é a descoberta de novos padrões em grandes quantidade de dados por meio de técnicas automáticas de exploração de dados; para Witten, Frank e Hall (2011) e Provost e Fawcett (2016), é definida como um processo de descoberta de padrões nos dados; já para Castro e Ferrari (2017) é um processo que utiliza algoritmos adequados para se obter conhecimentos.

A técnica de mineração de dados tem sido utilizada por empresas do segmento varejista, indústrias, bancos, dentre outras, para descobrir relações entre preços, comportamentos da concorrência, promoções, entre outros fatores que possam contribuir com um melhor posicionamento da empresa no mercado em que atua (CARVALHO, 2005; PROVOST e FAWCETT, 2016).

Importante observar que o uso da Mineração de Dados pode ser motivado por sua capacidade em realizar várias tarefas, sendo que as mais comuns são Descrição, Classificação, Estimação, Predição, Agrupamento e Associação (PROVOST e FAWCETT, 2016; CASTRO e FERRARI, 2017).

Além dessas tarefas, a Mineração de Dados pode também utilizar técnicas oriundas do campo da estatística, e, com isso, possibilitar a análise em conjunto com os resultados obtidos pelas técnicas de mineração.

Uma vertente que está se ampliando dentro da Mineração de Dados é o uso das tarefas e técnicas de mineração para o campo educacional, e alguns estudos já têm comprovado a eficácia do uso dessa alternativa.

4.2.1 Mineração de Dados Educacionais

As empresas da área da educação podem utilizar a técnica de Mineração de Dados Educacionais (em inglês *Educational Data Mining – EDM*) para tentar descobrir padrões ou comportamentos relacionados aos atores envolvidos no ambiente educacional. Esse termo foi oficializado a partir do primeiro *Workshop on Educational Data Mining* em 2005, na cidade de Pittsburgh no Estados Unidos. Vários outros eventos foram organizados a partir deste, e em 2008 foi organizada a primeira conferência em EDM com o nome de *First International Conference on Educational Data Mining*. A partir desse evento o conceito se estabeleceu internacionalmente, em 2009 foi publicado o primeiro volume do *JEDM - Journal of Educational Data Mining* e em 2011 constitui-se a *International Educational Data Mining Society*, uma comunidade científica com a finalidade de realizar estudos nessa área (BAKER e INVENTADO, 2014).

Internacionalmente, o uso da Mineração de Dados Educacionais vem se consolidando mais a cada dia, e algumas definições para o uso da EDM são utilizadas para identificar o foco e objetivo dessas técnicas, como exemplos Baker e Yacef (2009) e Bezerra e Silva (2020) salientam que a EDM pode ser considerada um campo de pesquisa que procura desenvolver métodos de mineração de dados para explorar dados educacionais na tentativa de entender melhor os estudantes e o ambiente em que estão inseridos. Já para Romero e Ventura (2013), a Mineração de Dados Educacionais é uma área que tem como foco o desenvolvimento, pesquisa e aplicação de métodos computadorizados, com intuito de descobrir padrões em grandes volumes de dados, do segmento educacional.

No segmento educacional podem ser utilizadas muitas técnicas de Mineração de Dados Educacionais como exemplos a previsão, classificação e agrupamentos. Mesmo assim, pesquisadores utilizam outras técnicas oriundas da computação, estatística, psicométrica e educação, como forma de agregar mais recursos para a descoberta de padrões que possam gerar conhecimentos (PEÑA-AYALA, 2017).

Como forma de categorizar o método para Mineração de Dados Educacionais Baker e Yacef (2009) propõem uma taxonomia e indicam que as principais subáreas de pesquisa na Mineração de Dados Educacionais são Predição, Agrupamento, Mineração de relações, Destilação de dados para facilitar decisões humanas e Descoberta com modelos.

Na área educacional a Predição pode ser utilizada para, por exemplo, prever com antecedência o risco de reprovação de alunos de determinado curso utilizando informações relacionadas à nota, frequência, participação em aula, etc. Já a técnica de agrupamento pode ser utilizada para identificar um conjunto de alunos que possuem características similares, formando um grupo que pode exercer determinado comportamento no ambiente escolar.

A técnicas que estão relacionadas com a Mineração de relações normalmente são utilizadas para identificar se uma variável está relacionada a outra. Como exemplo, se a evasão escolar está diretamente relacionada com o desempenho em alguma disciplina. O uso das técnicas de destilação de dados para facilitar decisões humanas permite que informações de um determinado contexto, no ambiente escolar, sejam refinadas ao ponto de promover entendimento de pontos cruciais que garanta um melhor nível de segurança no processo de tomada de decisão.

Finalmente, o uso das técnicas que envolvem a descoberta com modelos, pode proporcionar a identificação de informações importantes dentro de um determinado contexto, por exemplo um ambiente escolar, utilizando, para isso, modelos preparados com uma ou mais técnicas de mineração, que já foram testadas e que foram consideradas por seus usuários com uma boa alternativa na descoberta de informações em grandes volumes de dados. Para uma melhor visualização da taxonomia de Baker e Yacef (2009), a classificação está descrita assim:

- a) Predição
 - ✓ Classificação
 - ✓ Regressão
 - ✓ Estimação de densidade
- b) Agrupamento
- c) Mineração de relações
 - ✓ mineração de regras de associação
 - ✓ mineração de correlações
 - ✓ mineração de padrões sequenciais
 - ✓ mineração de causas
- d) Destilação de dados para facilitar decisões humanas
- e) Descobertas com modelos

O uso da Mineração de Dados no ambiente educacional pode proporcionar um ganho importante, tanto em termos de tempo quanto financeiros. Com o uso de tecnologia e ferramentas de mineração de dados, a análise de grandes volumes de dados pode ser realizada em um curto espaço de tempo, proporcionando agilidade nos processos de tomada de decisão e redução de custos operacionais para obtenção do conhecimento.

4.2.2 Pós-Processamento

As etapas do pós-processamento envolvem as fases de divulgação e Análise dos resultados. De acordo com Han, Kamber e Pei (2011), a etapa de pós-processamento envolve apresentação dos resultados e Avaliações dos Padrões dos Dados, que consiste em realizar a identificação e interpretação dos padrões considerados importantes para o objeto de pesquisa. Nessa etapa se observam as informações relevantes que podem ser obtidas a partir da análise dos dados na fase da mineração de dados.

Pode ser que em alguns casos seja necessário que um especialista em Mineração de Dados contribua com interpretação e análise dos resultados. Essa participação se torna importante para que o máximo de informação possa ser extraída dos resultados apresentados (GOLDSCHMIDT, PASSOS e BEZERRA, 2015). Da mesma forma, se torna necessário o apoio de um especialista da área que se está analisando, denominado especialista de domínio. Esse profissional, por ter

maiores conhecimentos sobre o ambiente foco da pesquisa, pode contribuir para que seja realizada uma análise mais profunda e consistente das informações encontradas por meio da mineração de dados.

4.3 Técnicas de Mineração de Dados

Vários são os métodos ou técnicas utilizadas nos processos de Mineração de Dados e normalmente são divididos em supervisionados e não supervisionados e a diferença entre eles é que para o método não supervisionado não é necessário que exista um atributo alvo pré-definido (PROVOST e FAWCETT, 2016; CASTRO e FERRARI, 2017). Para o desenvolvimento desse estudo foram utilizadas as técnicas de Classificação baseadas em Árvores de Decisão e Naive Bayes (Classificação Bayesiana). As definições e funcionalidades dessas técnicas são descritas a seguir.

4.3.1 Árvores de Decisão.

As Árvores de Decisão podem ser utilizadas em modelos de classificação e regressão, utilizando para isso algoritmos como J48, C.45, ID3, CART, *Random Forest*, *Gradient Boosted Tree*. As Árvores de Decisão é que podem utilizar métodos supervisionados ou não. Quando utilizado o método supervisionado, é necessário que exista um atributo alvo de predição para o desenvolvimento das árvores, e nesse modelo o algoritmo se encarrega de encontrar os melhores pontos de poda para montar a Árvore, com os atributos que apresentaram os melhores ganhos de informação.

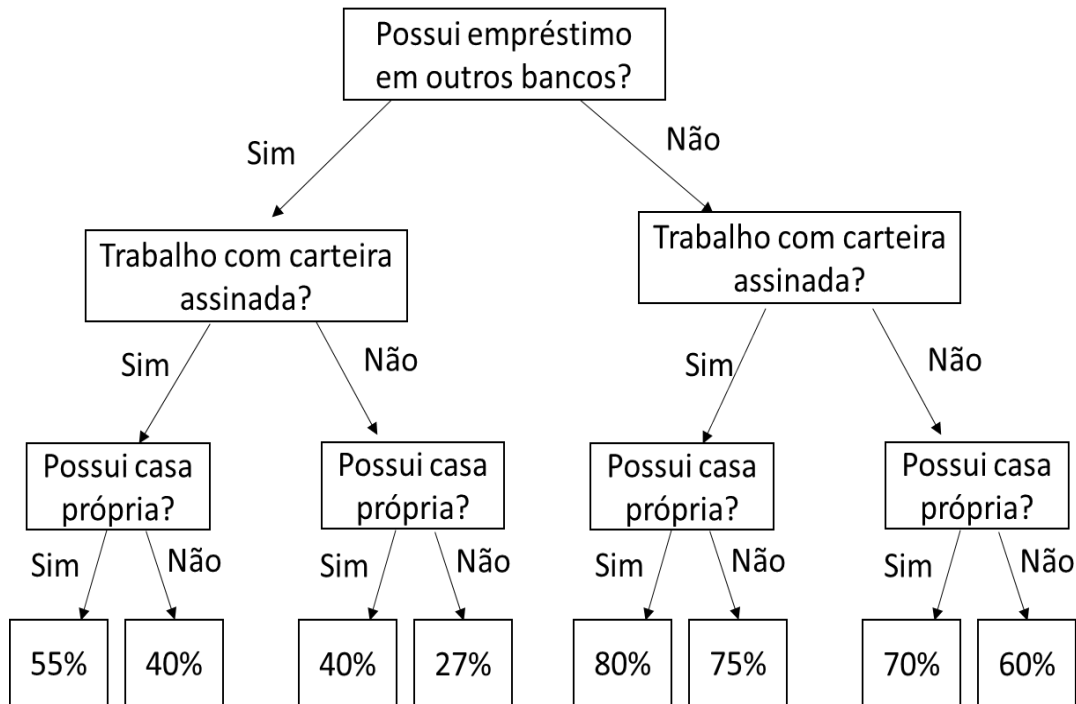
A partir de um conjunto de dados, o algoritmo utilizado para Árvore de Decisão executa os cálculos com base na Entropia, para encontrar o melhor ponto de corte para os nós. Assim que são finalizados todos os cálculos, é indicado como nó principal aquele que apresentou maior ganho de informação, e a partir desse nó todas as outras ramificações ou folhas são acrescentadas. Dessa forma pode-se entender que o atributo mais importante, com base na Entropia e no Ganho de Informação, está no topo da árvore. A fórmula utilizada para o cálculo da Entropia é:

$$Entropy(Y) = \sum_i p(c/Y) \log_2 p(c/Y) \quad (2)$$

Onde Y é o conjunto de casos e p(c / Y o valor de Y) é a proporção da classe c.

Após realizados todos os cálculos uma árvore é apresentada como resultado, e a análise pode ser realizada seguindo uma sequência entre o nó principal e as folhas subsequentes, até chegar no nó final. A Figura 4 ilustra um exemplo de Árvore de Decisão.

Figura 4. Exemplo de Árvore de Decisão.



Fonte: Adaptado de (BOUZADA; RIBEIRO; PEIXE, 2013)

Como alternativas de Árvores de Decisão, as *Random Forest* e as *Gradient Boosted Tree* são consideradas algoritmos mais robustos, pois os seus resultados normalmente são por meio de um conjunto de árvores que podem ser analisadas com base nos fatores de maior ganho de informação ou de importância da árvore.

As *Random Forest* ou Floresta de Árvores são consideradas um método de aprendizado em conjunto (*Ensemble*). O algoritmo utiliza para criar a Floresta de Árvores o método de *Bootstrap Agregating*, ou *Bagging*, que gera um conjunto de dados por amostragem *bootstrap* da base de dados utilizada no processo, e o conjunto de resultados pode proporcionar um melhor resultado geral. Como resultados são criadas várias pequenas árvores com baixo desempenho, mas que quando analisadas em conjunto se tornam uma boa alternativa para predição de atributos, pois utiliza a média entre as árvores para indicar os atributos de maior importância.

De forma semelhante, o *Gradient Boosted Tree* também utiliza a lógica de aprendizado em conjunto, podendo inclusive ser comparado com o *Random Forest*, mas esse algoritmo utiliza como critério para a criação do conjunto de árvores a otimização da função de perda. Com isso, sempre que ele cria uma nova árvore, ele tenta corrigir os erros da árvore anterior no conjunto, utilizando para isso a taxa de aprendizagem.

4.3.2 Naive Bayes

O algoritmo *Naive Bayes* é considerado um classificador probabilístico baseado no Teorema de Bayes criado pelo matemático Thomas Bayes, que tem como lógica a probabilidade condicional, ou seja, qual a probabilidade do evento X ocorrer dado o evento Y. A fórmula que esse algoritmo utiliza para prever os resultados é:

$$p\left(\frac{A}{B}\right) = P\left(\frac{B}{A}\right)P(A)/P(B) \quad 3)$$

O cálculo da densidade de probabilidade utilizado na representação dos resultados desse algoritmo é a soma das probabilidades de cada classe multiplicada por sua função densidade de probabilidade, sua fórmula é:

$$P(x) = \sum_i P(C_i)P\left(\frac{X}{C_i}\right) \quad 4)$$

Onde:

$P(x)$: densidade de probabilidade

$P(C_i)$: probabilidade a priori da classe C_i

$P(X/C_i)$: função densidade de probabilidade da classe C_i

Resultados alcançados por meio desse algoritmo são compatíveis com aqueles apresentados pelas Árvores de Decisão, e por conta da simplicidade no uso e alto desempenho, esse algoritmo vem ganhando notoriedade na literatura.

Os critérios de utilização das técnicas relacionadas à Árvore de Decisão, *Random Forest*, *Gradiente Boosted Tree* e *Naive Bayes* estão descritos nos artigos

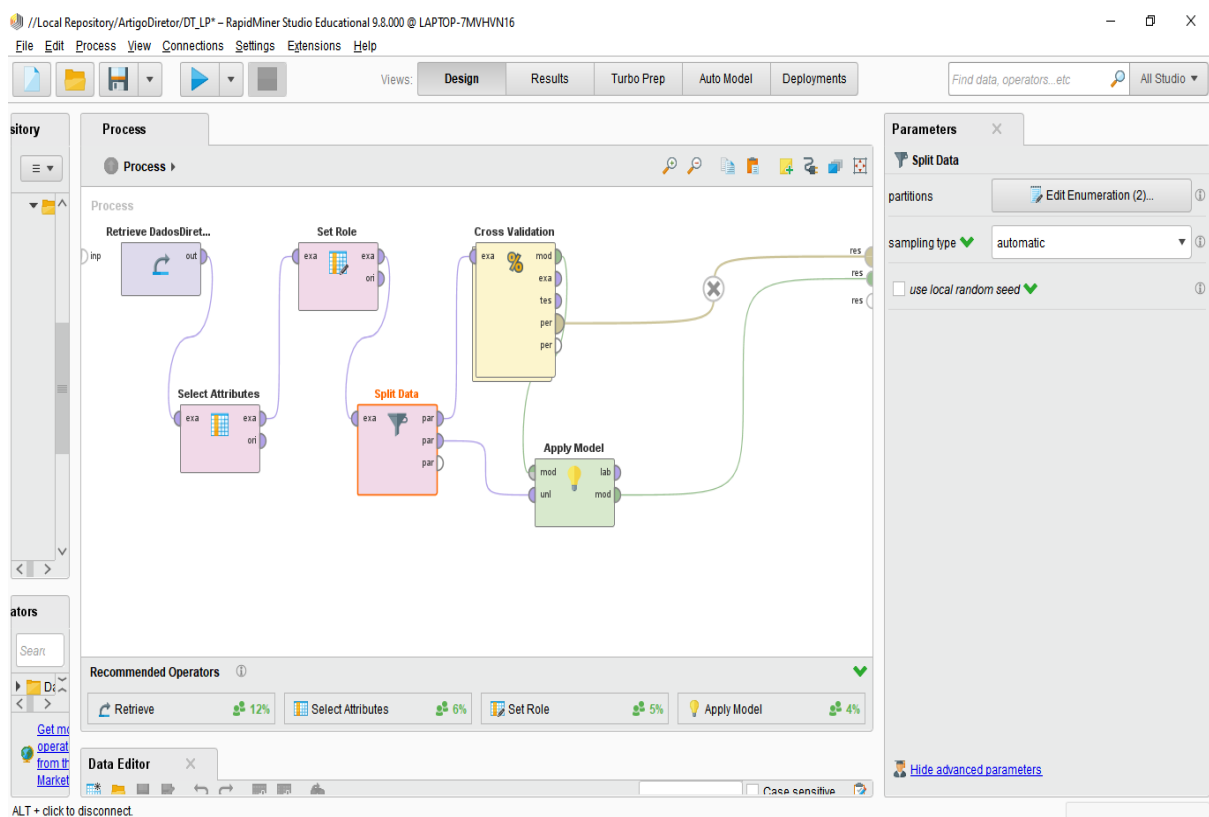
apresentados no Capítulo Resultados em Artigos, pois para cada artigo desenvolvido foi utilizado um conjunto de critérios específicos.

4.4 Softwares para Mineração de Dados

Existem alguns *softwares* que podem ser utilizados para os processos de Mineração de Dados, como exemplos são: Weka, RapidMiner, Phyton, R, SPSS, Alteryx, SAS Enterprise Miner, Orange, Data Science Studio, entre outros.

Alguns desses *softwares* são gratuitos, como o Weka e o Orange; outros devem ser adquiridos por meio de licenças de uso. Independente do software escolhido, o importante a observar é para qual finalidade ele será utilizado, pois alguns são destinados para análises preditivas, outros apenas para Mineração de Texto, e ainda tem aqueles com foco em estatística. Nesse estudo será utilizado o *RapidMiner* versão 9.8, com licença educacional. A interface desse software está ilustrada na Figura 5.

Figura 5. Interface do Software RapidMiner.



Fonte: Adaptado pelo autor (2020).

O *software RapidMiner* é desenvolvido pela RapidMiner Inc., com base em tecnologia Java e uma interface amigável, que permite configurar um processo de mineração com poucos passos. Além das tarefas relacionadas à mineração de dados, é possível também realizar tarefas com base em estatística.

A escolha do *RapidMiner* foi motivada por conta da interface do software, pois para executar um processo de mineração basta arrastar e soltar os componentes. Além disso, existe um tutorial integrado que vai sugerindo os próximos passos.

5 METODOLOGIA

O percurso metodológico de uma pesquisa serve para orientar o pesquisador sobre a necessidade de observar passos importantes tanto antes quanto durante a execução do estudo. Dessa forma, o pesquisador não perderá o foco do que é necessário realizar para alcançar os objetivos da pesquisa. Minayo (2010), corroborando Gil (2010) e Lakatos e Marconi (2017), salienta que a metodologia deve incluir o método, as técnicas e a criatividade do pesquisador.

Como delimitação da pesquisa, foram utilizados para compor a base de dados os resultados das avaliações em larga escala do Sistema de Avaliação da Educação Básica – SAEB, das edições de 2015 e de 2017, referentes ao Ensino Médio, que estão disponíveis no portal do INEP.

O percurso metodológico desse estudo foi dividido em três etapas, envolvendo a coleta, tratamento e agrupamento de dados, a etapa envolvendo os procedimentos estatísticos e, por fim, o uso de técnicas de mineração de dados para verificar os fatores preditivos, que podem contribuir para o melhor desempenho em avaliações em larga escala por alunos do ensino médio.

5.1 Coleta e preparação dos Dados

A Coleta de dados para esse estudo passou por duas etapas. Na primeira etapa foi realizada a pesquisa bibliográfica exposta no item 2.3 desse estudo, com objetivo de verificar as publicações existentes sobre o tema investigado. Com os resultados dessa etapa foi criado o Quadro 2, descrito no item 1.4.

Para a realização de uma análise inicial sobre o tema, foram selecionados alguns estudos que apresentavam em seus resultados informações que convergiam para o objetivo dessa pesquisa. A partir dos resultados das publicações analisadas, foram selecionados os fatores que, segundo os autores dessas publicações, puderam contribuir para desempenho educacional. Esses fatores preditores de desempenho estão relacionados no Quadro 3, também descrito no item 2.2 desse estudo.

Como segunda etapa, foram coletados os resultados das Avaliações SAEB das edições de 2015 e 2017, disponíveis no portal do INEP (INEP, 2020c). No pacote de dados disponibilizado pelo INEP, além dos desempenhos dos alunos, também estão inclusos os formulários que foram direcionados aos Alunos, aos Professores, Diretores e Representantes de Escola. O INEP coleta por meio desses

formulários informações sociodemográficas dos alunos e das atividades inerentes do corpo docente e da gestão escolar. Nos formulários direcionados às escolas são coletadas informações referentes à estrutura escolar - estrutura física, segurança, e recursos tecnológicos. Os resultados desses formulários também são disponibilizados no pacote de dados de cada edição do SAEB, já separados por tipo de participante, ou seja, uma base de dados para professores, outra para diretores e uma outra para as escolas. Os resultados dos formulários direcionados aos alunos já estão inseridos na base de dados referente aos desempenhos.

Além dos dados sobre os desempenhos dos alunos e dos formulários, o INEP também disponibiliza os dicionários de dados e as escalas de desempenho das proficiências avaliadas, documentos esses essenciais para a fase de preparação dos dados.

Os dados coletados foram devidamente tratados para que não apresentassem inconsistências ou outro tipo de problema na fase de análise, como exemplo a retirada de registros das bases de dados que continham campos em branco ou nulos ou que foram erroneamente preenchidos. Essa fase de tratamento de dados envolveu as tarefas de depuração, codificação, categorização e agrupamento desses dados, com intuito de adequá-los ao objetivo e às proposições do estudo.

Após a fase de tratamento dos dados, como resultados, foram criadas quatro bases de dados individuais contendo as informações referentes aos alunos do Ensino Médio, dos Professores, dos Diretores e das Escolas que ofertam o Ensino Médio. Com as bases prontas, o próximo passo foi separar as questões de cada tipo de formulário que convergiam com o quadro de fatores preditores de desempenho encontrados na literatura, (Quadro 3 descrito no item 2.2 desse estudo), procedimento esse similar ao que foi realizado por Andrade e Laros (2007), Laros et al. (2010), Franco e Menezes Filho (2012), Santos (2015) e Garcia et al. (2016).

Para uma melhor visualização, os resultados do agrupamento das questões referente aos formulários dos ALUNOS estão descritos no Quadro 14:

Quadro 14. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes aos Alunos.

ALUNOS		
1	Autores	
	Simões e Ferrão (2005); Bezerra e Kassouf (2006); Luz (2006) ; Gonçalves e França (2008); Brooke e Soares (2008); Cunha et al. (2009); Rodrigues, Rios-Neto e Pinto (2011); Laros, Marciano e Andrade (2012); Bezerra et al. (2016); Travitzki, Ferrão e Couto (2016)	
Fatores	Nível socioeconômico da família/aluno	Questões
		NÃO POSSÍVEL ASSOCIAR QUESTÕES DO FORMULÁRIO DOS ALUNOS
2	Autores	
	Bezerra e Kassouf (2006); Barbosa e Wilhelm (2009); Machebe, Ezegebe e Onuoha (2017), Castillo-Vergara (2018)	
Fatores	Renda familiar	Questões
		NÃO POSSÍVEL ASSOCIAR QUESTÕES DO FORMULÁRIO DOS ALUNOS
3	Autores	
	Bezerra e Kassouf (2006); Lobo (2017); Fantinato e Cia (2017)	
Fatores	Escolaridade da mãe	Questões
		19. Até que série sua mãe, ou a mulher responsável por você, estudou?
4	Autores	
	Bezerra e Kassouf (2006); Lobo (2017); Fantinato e Cia (2017)	
Fatores	Escolaridade dos pais	Questões
		19. Até que série sua mãe, ou a mulher responsável por você, estudou? 23. Até que série seu pai, ou o homem responsável por você, estudou?
5	Autores	
	Barbosa e Wilhelm (2009)	
Fatores	Alunos que trabalham	Questões
		38. Atualmente você trabalha fora de casa (recebendo ou não um salário)?
6	Autores	
	Simões e Ferrão (2005); Bezerra e Kassouf (2006); Marturano e Pizato (2015); Brito (2017); Jovarini, Leme e Correia-Zanini (2018)	
Fatores	Sempre fazer as lições escolares e motivação para aprender	Questões
		44. Você gosta de estudar Língua Portuguesa? 45. Você faz o dever de casa de Língua Portuguesa? 52. Você gosta de estudar Matemática? 53. Você faz o dever de casa de Matemática?
7	Autores	
	Correia-Zanini (2013); Marturano e Pizato (2015); Jovarini, Leme e Correia-Zanini (2018)	
Fatores	Fatores estressantes (clima sala de aula, notas baixas, apoio para realiza as tarefas, etc)	Questões
		NÃO POSSÍVEL ASSOCIAR QUESTÕES DO FORMULÁRIO DOS ALUNOS
8	Autores	
	Sebanc, Guimond e Lutgen (2016); Fernandes et al. (2018).	

Fatores	Habilidades sociais dos alunos (relacionamento com colegas, empatia, etc)	Questões	NÃO POSSÍVEL ASSOCIAR QUESTÕES DO FORMULÁRIO DOS ALUNOS
9	Autores		
	Bezerra e Kassouf (2006); Andrade e Laros (2007); Gonçalves e França (2008); Laros, Marciano e Andrade (2010); Brito (2017); Fernandes et al. (2018); Paula, Franco e Silva (2018); Lacruz et al. (2019)		
Fatores	Atraso escolar	Questões	41. Você já foi reprovado? 42. Você já abandonou a escola durante o período de aulas e ficou fora da escola o resto do ano?
10	Autores		
	Dwyer et al. (2007); Gonçalves e França (2008)		
Fatores	Uso do computador	Questões	13. Na sua casa tem computador? O que você consulta para fazer o dever de casa de Língua Portuguesa? 50. Sites da internet. O que você consulta para fazer o dever de casa de Matemática? 58. Sites da internet.
12	Autores		
	Gonçalves e França (2008); Zafani e Omote (2016); Brito (2017).		
Fatores	Cobrança e incentivo dos pais do aluno	Questões	27. Seus pais ou responsáveis incentivam você a estudar? 28. Seus pais ou responsáveis incentivam você a fazer o dever de casa e/ou os trabalhos da escola? 29. Seus pais ou responsáveis incentivam você a ler? 30. Seus pais ou responsáveis incentivam você a ir à escola e/ou não faltar às aulas?

Fonte: Desenvolvido pelo Autor (2020)

Os fatores preditores de desempenho relacionados aos Professores e as respectivas questões do Formulário SAEB estão descritos no Quadro 15:

Quadro 15. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes aos Professores

PROFESSORES			
1	Autores		
	Laros e Marciano (2010)		
Fatores	Estimular o trabalho colaborativo entre os professores	Questões	NESTA ESCOLA, COM QUE FREQUÊNCIA VOCÊ FEZ O SEGUINTE: 55. Participou de reuniões com colegas que trabalham com a mesma série(ano) para a(o) qual leciona. 57. Envolveu-se em atividades conjuntas com diferentes professores (por exemplo, projetos interdisciplinares). 69. A equipe de professores leva em consideração minhas ideias.
2	Autores		
	Soares et al. (2010); Fonseca e Namen (2016); Lauermann (2017); Schildkamp et al. (2020); Zakariya, Bjorkestol e Nilsen (2020).		
Fatores	Expectativas do professor	Questões	95. (PARA TODOS OS PROFESSORES) Quantos dos alunos desta turma você acha que concluirão o ensino médio? 96. (PARA TODOS OS PROFESSORES) Quantos dos alunos desta turma você acha que entrarão na universidade?

3		Autores	
Bezerra e Kassouf (2006); Dourado, de Oliveira e Santos (2007); Ortigão (2008); Hanushek e Woessmann (2017); Lauermann (2017); Mesquita (2018); Symeonidis (2019); Voinea (2019); Nascimento e Gomes (2020); Schildkamp et al. (2020); Zakariya, Bjorkestol e Nilsen (2020).			
Fatores	Escolaridade/formação dos professores	Questões	4. Qual é o mais alto nível de escolaridade que você concluiu (até a graduação)? 8. Indique o curso de pós-graduação de mais alta titulação que você possui. 19. Considerando-se todas as escolas em que você trabalha atualmente como professor (a), qual é a sua carga horária semanal? (Considere a carga horária contratual: horas-aula mais horas para atividades, se houver. Não considere aulas particulares).
4		Autores	
Bezerra e Kassouf (2006); Fini (2009); Fonseca e Namen (2016); Simielli (2017).			
Fatores	Nível de salários professores	Questões	10. Como professor, qual é, aproximadamente, o seu salário bruto? (com adicionais, se houver)
5		Autores	
Pereira e Mori (2012); Amâncio-Vieira et al. (2015); Fonseca e Namen (2016); Pettersson et al. (2016); Vidal e Vieira (2017); Jovarini, Leme e Correia-Zanini (2018); Schaik et al. (2018).			
Fatores	Atividades pedagógicas do corpo docente	Questões	20 Considerando-se todas as escolas em que você trabalha atualmente como professor(a), quantas horas semanais são dedicadas às atividades extraclasse (formação e estudo, planejamento, produção de recursos didáticos, etc.)? (GOSTARÍAMOS DE SABER QUAIS OS RECURSOS QUE VOCÊ UTILIZA PARA FINS PEDAGÓGICOS, NESTA TURMA: 50. Internet, 45 Livros, 44 Jornais e revistas). FORMULÁRIO DOS ALUNOS (51. O(A) professor(a) corrige o dever de casa de Língua Portuguesa? 59. O(A) professor(a) corrige o dever de casa de Matemática?)

Fonte: Desenvolvido pelo Autor (2020)

Os fatores preditores de desempenho relacionados aos Diretores de Escola e as respectivas questões do Formulário SAEB estão descritos no Quadro 16:

Quadro 16. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes aos Diretores

DIRETORES			
1		Autores	
Gonçalves e França (2008); Barbosa e Wilhelm (2009)			
Fatores	Participação dos pais na administração das escolas	Questões	29. O Conselho Escolar é um colegiado geralmente constituído por representantes da escola e da comunidade que tem como objetivo acompanhar as atividades escolares. Neste Anos, quantas vezes se reuniu o Conselho escolar? 30. Além de você, quem participa do Conselho escolar?
2		Autores	
Filho, Nuñez e Ribeiro (2009); Souza e Ribeiro (2017); Vieira e Bussolotti (2018); Barros et al. (2019); Pereira e Gasque (2019); Marangoni et al. (2019); Dallazem e Coelho (2019); Nascimento e Chiusoli (2019); Ribeiro, Síveres e Brito (2019); Zambello et al. (2019); Nichele e Mello (2020); Morais et al. (2020); Pereira (2020); Souza (2020).			

Fatores	Experiência do diretor	Questões	13. Qual é a sua carga horária de trabalho semanal, como Diretor, nesta escola? 15. Por quanto tempo você trabalhou como Professor(a) antes de se tornar Diretor(a) ? 16. Há quantos anos você exerce funções de Direção? 17. Há quantos anos você é Diretor(a) desta escola ? 18. Há quantos anos você trabalha na área de educação?
3	Autores		
Filho, Nuñez e Ribeiro (2009)			
Fatores	Estabilidade do corpo docente	Questões	28. Qual é o percentual de Professores com vínculo estável nesta escola? 75. Alta rotatividade do corpo docente.
4	Autores		
Filho, Nuñez e Ribeiro (2009); Lacruz et al. (2019)			
Fatores	Frequente de faltas por parte dos professores	Questões	73. Alto índice de faltas por parte dos professores.
5	Autores		
Filho, Nuñez e Ribeiro (2009); Laros e Marciano (2010)			
Fatores	Ocorrência de roubos e episódios de violência nas escolas	Questões	SOBRE OS FATOS DE VIOLÊNCIA LISTADOS ABAIXO, DIGA SE ELES ACONTECERAM OU NÃO ESTE ANO, NESTA ESCOLA: 90. Agressão verbal ou física de alunos a professores ou funcionários da escola. 91. Agressão verbal ou física de alunos a outros alunos da escola. 92. Você foi vítima de atentado à vida. 93. Você foi ameaçado por algum aluno. 94. Você foi vítima de furto (sem uso de violência). 95. Você foi vítima de roubo (com uso de violência). 96. Alunos frequentaram a escola sob efeito de bebida alcoólica. 97. Alunos frequentaram a escola sob efeito de drogas ilícitas. 98. Alunos frequentaram a escola portando arma branca (facas, canivetes etc.). 99. Alunos frequentaram a escola portando arma de fogo.
6	Autores		
Filho, Nuñez e Ribeiro (2009); Drabach (2013); Cária e Lambert-De-Andrade (2016); Back et al. (2016); Garcia et al. (2016); Oliveira e Waldhelm (2016); Philippsen (2018); Vieira e Bussolotti (2018); Aldrup et al. (2018); Masci, De Witte, Agasisti (2018); Freitas (2019); Riddle e Apple (2019)			
Fatores	Processos de Gestão Escolar	Questões	O FUNCIONAMENTO DA ESCOLA FOI DIFICULTADO POR ALGUM DOS SEGUINTE PROBLEMAS? 67. Insuficiência de recursos financeiros. 68. Inexistência de professores para algumas disciplinas ou séries. 69. Carência de pessoal administrativo. 70. Carência de pessoal de apoio pedagógico (supervisor, coordenador, orientador educacional). 71. Falta de recursos pedagógicos. 72. Interrupção das atividades escolares. 73. Alto índice de faltas por parte dos professores. 74. Alto índice de faltas por parte dos alunos. 75. Alta rotatividade do corpo docente. 76. Indisciplina por parte dos alunos.
7	Autores		
Laros e Marciano (2010); Mohr e Naujorks (2017); Góis e Rocha (2019)			

Fatores	Aumentar o controle sobre a incidência de faltas na sala de aula	Questões
	Sanar as causas da repetência do ano letivo	
<p>41. Nesta Escola, há alguma ação para redução das taxas de abandono? 42. Nesta escola, há alguma ação para redução das taxas de reprovação? 43. Nesta escola, há alguma ação para o reforço escolar à aprendizagem dos alunos (Monitoria, Aula de reforço, Recuperação, etc.)? 44. Nesta escola, indique com que frequência você discute com os professores medidas com o objetivo de melhorar o ensino e a aprendizagem dos alunos. INDIQUE COM QUAL FREQUÊNCIA SÃO DESENVOLVIDAS AS SEGUINTE ATIVIDADES PARA MINIMIZAR AS FALTAS DOS ALUNOS NESTE ANO E NESTA ESCOLA: (45. Os professores conversam com os alunos para tentar solucionar o problema. 46. Os pais/responsáveis são avisados por comunicação da escola. 47. Os pais/responsáveis são chamados à escola para conversar sobre o assunto em reunião de pais. 48. Os pais/responsáveis são chamados à escola para conversar sobre o assunto individualmente. 49. A escola envia alguém à casa do aluno. 74. Alto índice de faltas por parte dos alunos.</p>		

Fonte: Desenvolvido pelo Autor (2020)

E, por fim, os fatores preditores de desempenho relacionados aos Representantes de Escola e as respectivas questões do Formulário SAEB estão descritos no Quadro 17:

Quadro 17. Relação de Fatores de Desempenho e questões do Formulário SAEB referentes os Representantes de Escola

ESCOLA			
1	Autores		
	Gonçalves e França (2008); Laros, Marciano e Andrade (2012); Amâncio-Vieira e Pereira (2016); Gustafsson, Nilsen e Hansen (2018); Murphy (2019);		
Fatores	Nível socioeconômico da escola	Questões	Questão “NIVEL_SOCIO_ECONOMICO” já disponibilizada junto com os resultados da escola. Essa questão não está no formulário.
2	Autores		
	Soares e Alves (2003); Felício e Fernandes (2005); Bezerra e Kassouf (2006); Gonçalves e França (2008); Gatti (2009); Schneider e Gouveia (2011); Amâncio-Vieira e Pereira (2016); Fonseca e Namen (2016); Calixto, Segundo e Gusmão (2017); Brito (2017); Razak et al. (2018); Damon et al. (2019); Ford (2019); Lasater, Bengtson e Albiladi (2020); Vasconcelos et al. (2020)		
Fatores	A infraestrutura escolar (salas de aula, laboratórios; material didático; materiais audiovisuais, etc.)	Questões	13. Salas de Aula. 16. Banheiros. 20. São iluminadas de forma adequada. 21. São arejadas de forma adequada. 37. Computadores para uso dos alunos. 38. Acesso à internet para uso dos alunos. 39. Computadores para uso dos professores. 40. Acesso à internet para uso dos professores. 42. Fitas de vídeo ou DVD (educativas). 47. Projetor de slides/datashow. 53. Internet com conexão Banda Larga. 57. Biblioteca. 58. Sala de leitura. 59. Quadra de esportes. 60. Laboratório de informática. 61. Laboratório de Ciências. 68. Os livros podem ser manuseados e emprestados. OS SEGUINTE USUÁRIOS DA BIBLIOTECA (OU SALA DE LEITURA) LEVAM LIVROS PARA CASA: 72. Os alunos 73. Os professores

Fonte: Desenvolvido pelo Autor (2020)

Uma cópia dos formulários aplicados aos Alunos do Ensino Médio, Professores, Diretores e Escola estão nos Anexos dessa pesquisa, incluindo-se também uma cópia das escalas de proficiências utilizadas para comparar o nível de desempenho dos alunos.

Após essas etapas realizadas e com as bases de dados prontas, o próximo passo foi o desenvolvimento dos artigos. Para essa tarefa foram utilizadas, individualmente e conforme o objetivo específico, as bases de dados prontas. Para uma melhor visualização, o Quadro 18 ilustra os artigos desenvolvidos e para qual ou quais objetivos específicos foi direcionado.

Quadro 18. Relação de artigos que compõem a Tese.

Nº	Obj. Esp.	Título	Autores	Publicação	Situação
1	a	ANÁLISE DO DESEMPENHO ESCOLAR NA AVALIAÇÃO SAEB 2015.	Ivonaldo Vicente Silva, Márcia Terra da Silva, Saturnina Martins	XXXVIII Encontro Nacional de Engenharia de Produção (ENEGEP) - 2018	Aprovado e Apresentado
2	b, c	PREDICTIVE SUCCESS FACTORS IN SCHOOL PERFORMANCE: An analysis of the large-scale assessment in Brazil	Ivonaldo Vicente Silva, Márcia Terra da Silva, Saturnina Martins	ICT, Society and Human Beings 2019 International Conference in Porto, Portugal	Aprovado e Apresentado
3	a	Ensino Médio e as Avaliações em Larga Escala do SAEB: Uma análise do desempenho apresentado no período de 2005 a 2017.	Ivonaldo Vicente da Silva, Dra. Márcia Terra da Silva	Revista Educação Por Escrito	Submetido
4	b, c	Teaching practices and school performance: discovering predictors of success using Data Mining.	Ivonaldo Vicente da Silva, Dra. Márcia Terra da Silva e Dra. Nilza Duarte da Silva Lima.	Computers & Educatiion	Submetido
5	b, c	Predictive factors of school performance in SAEB assessments: influence of school management.	Ivonaldo Vicente da Silva, Dra. Márcia Terra da Silva e Dra. Nilza Duarte da Silva Lima.	Research, Society and Development - ISSN 2525-3409	Aprovado e Publicado
6	b, c	Use of Data Mining to identify the technological resources that contributed to school performance in large-scale evaluations of Brazilian high school.	Ivonaldo Vicente da Silva e Dra. Márcia Terra da Silva	IHIET-AI 2021 - 4rd International Conference on Human Interaction and Emerging Technologies - Strasbourg, France	Aprovado e será Apresentado em abril/2021

Fonte: Elaborado pelo Autor (2020).

Os artigos foram desenvolvidos seguindo uma determinada sequência, primeiramente, Artigo 1, no qual foram observados os desempenhos dos alunos nas avaliações do SAEB da edição de 2015. Para isso, foram utilizadas apenas técnicas da Estatística Descritiva. Em sequência e já utilizando as técnicas de Mineração de Dados baseadas em Árvores de Decisão, o Artigo 2 foi desenvolvido com objetivo de descobrir quais fatores estavam associados ao bom desempenho dos alunos nas avaliações do SAEB. Até esse ponto foram utilizados os resultados da edição de 2015 por que os resultados da edição de 2017 ainda não tinham sido disponibilizados pelo governo.

Dando sequência e já com os dados da edição de 2017 disponíveis, o Artigo 3 foi desenvolvido com intuito de verificar a evolução dos desempenhos dos alunos do Ensino Médio nas edições do SAEB, compreendendo o período de 2005 a 2017. Para esse artigo também foram utilizadas apenas as técnicas da Estatística Descritiva.

O Artigo 4 foi desenvolvido com intuito de observar quais fatores presentes nas atividades dos professores poderiam estar associados com o melhor desempenho dos alunos do Ensino Médio. Para isso foram utilizados os resultados da edição de 2017 do SAEB e as técnicas de Mineração de Dados baseadas em Árvores de Decisão e *Random Forest*.

Para o Artigo 5 também foram utilizados os resultados da edição de 2017 do SAEB com intuito de descobrir quais fatores presentes nas atividades dos Gestores Escolar (Diretores de Escola) poderiam estar associados com o melhor desempenho dos alunos nas avaliações. Como técnicas de Mineração de Dados foram utilizadas as Árvores de Decisão, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes*.

Finalmente, o Artigo 6 foi desenvolvido com o objetivo de descobrir se o nível de acesso a recursos de informática e biblioteca ofertados pelas escolas poderiam estar associados com os melhores desempenhos dos alunos nas avaliações do SAEB. Para isso foram utilizados os resultados da edição de 2017 do SAEB e as técnicas de Mineração de Dados baseadas em Árvores de Decisão, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes*.

A sequência de desenvolvimento dos artigos procurou observar, individualmente, como cada “ator” do ambiente educacional pode ter suas ações ou atividades associadas com o desempenho escolar dos alunos do Ensino Médio. De

forma a ilustrar essa sequência, a Figura 6 apresenta a sequência dos artigos e como cada um deles está associado com o desenvolvimento desse estudo.

Figura 6 Sequência dos Artigos que compõem a Tese.

T E S E	Artigo 1 - Descreve o Desempenho dos alunos nas avaliações de 2015
	Artigo 2 - Descreve os fatores dos alunos que estão associados ao Desempenho nas avaliações de 2015.
	Artigo 3 - Descreve a evolução do Desempenho dos alunos nas avaliações (período 2005 a 2017)
	Artigo 4 - Descreve os fatores dos professores que estão associados ao Desempenho nas avaliações de 2017
	Artigo 5 - Descreve os fatores dos gestores escolar que estão associados ao Desempenho nas avaliações de 2017
	Artigo 6 - Descreve se os fatores relacionados com o nível da oferta de recursos tecnológicos e de biblioteca estão associados ao Desempenho nas avaliações de 2017

Fonte: Elaborado pelo Autor (2020).

Os artigos estão inclusos, em sua íntegra, no Capítulo Resultados em Artigos para que seja possível observar a metodologia utilizada e os resultados encontrados.

6 RESULTADOS EM ARTIGOS

Neste capítulo serão descritos os artigos que foram desenvolvidos para compor a tese. Os artigos obedeceram uma sequência que utilizou como lógica observar, no ambiente educacional, os fatores dos alunos, dos professores, dos gestores e das escolas, que estão associados com o desempenho dos alunos do Ensino Médio.

Inicialmente, o primeiro artigo demonstrou por meio dos resultados como estavam classificados os desempenhos dos alunos na edição do SAEB de 2015. Como metodologia foram utilizadas as técnicas de estatística descritiva. Os dados foram coletados do website do INEP e após as tarefas de tratamento, esses dados foram agrupados por proficiências, gêneros e por níveis das escalas de desempenhos. Os resultados encontrados indicaram que a maior concentração de alunos apresentava desempenhos iguais ou inferiores ao Nível 3 da escala de desempenho.

Essa foi a primeira visão sobre como os alunos se desempenhavam nas avaliações em larga escala do SAEB. Então, como próximo passo, foi desenvolvido o segundo artigo, no qual se procurou identificar os fatores preditivos dos alunos com desempenho igual ou superior a 350 pontos, compatíveis com o nível 6 da escala de desempenho. Como metodologia foram utilizadas as técnicas de estatística descritiva para agrupar os dados e da Mineração de dados para a descoberta dos fatores preditivos. Por meio dos resultados das Árvores de Decisão, os resultados demonstraram que o nível de escolaridade dos pais ou responsáveis pelos alunos pode ser considerado um fator importante e fundamental para o desempenho escolar.

Com os resultados dos dois primeiros artigos, o próximo passo foi analisar como os resultados apresentados nas avaliações do SAEB estavam evoluindo, e, para isso, foi desenvolvido o terceiro, que teve por finalidade analisar os desempenhos dos egressos do Ensino Médio no período de 2005 a 2017. A análise realizada nesse artigo foi dividida em duas etapas; a primeira envolveu os desempenhos médios apresentados pelos estados e regiões para o período de 2005 a 2017 e na segunda foram analisados os desempenhos dos alunos em duas edições do SAEB, uma de 2005 e outra de 2017.

Por meio das técnicas da estatística descritiva, os resultados encontrados indicaram que os níveis de desempenhos apresentados, tanto dos alunos quanto

dos estados e regiões, em sua maioria não ultrapassaram os 325 pontos, ficando no máximo no nível 4 da escala de desempenho.

Apenas como comparativo pode-se observar que os resultados do primeiro artigo convergem para os resultados encontrados no terceiro, indicando assim um tipo de padrão nos desempenhos dos alunos nas avaliações de larga escala do SAEB.

Finalizados esses três artigos, já estavam resolvidas as questões envolvendo os desempenhos dos alunos e os fatores que podem estar associados com um melhor desempenho nas avaliações do SAEB. Assim, os próximos passos foi descobrir quais fatores referentes aos professores e diretores estavam associados com os melhores desempenhos dos alunos. Para isso foram desenvolvidos o quarto e quinto artigo.

No quarto artigo tratou-se de identificar os fatores dos professores que estavam associados com os melhores desempenhos dos alunos e, para isso, utilizou-se como metodologia as técnicas de Mineração de Dados Educacionais nos resultados da edição de 2017 do SAEB. Os resultados encontrados por meio de Árvores de Decisão e da *Random Forest* mostraram que, além do nível socioeconômico da escola, a experiência do professor e a melhora nas habilidades e competências pedagógicas estão intimamente relacionadas ao desempenho escolar.

No quinto artigo o objetivo foi observar quais fatores referente às atividades dos Diretores de Escola estavam associados com os desempenhos dos alunos do Ensino Médio, nas avaliações em larga escala do SAEB no ano de 2017. A metodologia utilizada contemplou as técnicas de Mineração de Dados Educacionais com base em Árvores de Decisão, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes* e os resultados demonstraram que o Tempo de Experiência do Diretor de Escola, nível de pós-graduação, ações para controle de reprovações e Nível Socioeconômico da escola estavam associados com os melhores desempenhos apresentados pelos alunos do Ensino Médio, nas avaliações do SAEB em 2017.

Faltava então observar os dados do ambiente escolar, e para isso foi desenvolvido o sexto artigo. O objetivo desse artigo foi observar se a qualidade de acesso aos recursos tecnológicos e de biblioteca, por alunos e professores no ambiente escolar, estavam associados com os melhores desempenhos apresentados pelos alunos do Ensino Médio, na edição de 2017. A metodologia utilizada contemplou as técnicas de Mineração de Dados Educacionais com base em

Árvores de Decisão, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes*, e os resultados encontrados pela *Random Forest*, que apresentou o melhor desempenho dentre os utilizados, indicaram que o nível de acesso a computadores por alunos e professores, o acesso à internet pelos alunos e a oferta de internet banda larga pela escola estão associados com os melhores desempenhos apresentados pelos alunos nas avaliações do SAEB, em 2017.

Apenas para lembrar que os dados utilizados em cada artigo foram as bases de dados descritas na Metodologia desse estudo. Então os dados foram tratados e separados conforme a dimensão analisada, ou seja, alunos, professores, diretores e escolas. As tarefas que envolveram a fase de tratamento de dados compreenderam na retirada dos registros em banco de dados, que estavam em branco, nulos ou que estavam preenchidos erroneamente. Com as bases de dados prontas, a fase de desenvolvimento dos artigos utilizou os dados conforme a necessidade.

Encerrando essa fase de descrição, os artigos são apresentados a seguir conforme ordem de 1 a 6. Cada periódico ou congresso utiliza uma formatação própria para que os artigos sejam submetidos, então apenas para manter um padrão estético, a formatação apresentada para dois artigos obedecerá a formatação original da tese.

6.1 ARTIGO 1 – Análise do Desempenho Escolar na Avaliação SAEB 2015.

Ivonaldo Vicente da Silva

Dra. Márcia Terra da Silva

RESUMO

A necessidade de profissionais qualificados para o mercado de trabalho, principalmente pelo advento de novas tecnologias e processos produtivos, faz com que sejam observadas primeiramente a qualidade da formação dos alunos que finalizam seus estudos no ensino médio, pois são esses que chegarão rapidamente no mercado de trabalho. Com isso o objetivo desse estudo foi analisar o desempenho dos alunos egressos do ensino médio na avaliação do sistema SAEB do ano de 2015. Os desempenhos foram analisados separadamente por proficiência e gênero. Os resultados encontrados indicam uma forte necessidade nas adaptações das políticas educacionais, que possam corrigir as deficiências apresentadas, e garantir um melhor acesso ao mercado de trabalho pelos egressos do ensino médio e a proporcionar condições suficientes para uma ideal formação profissional.

Palavras-chave: Desempenho escolar; Avaliação SAEB; SAEB;

ASBTRACT

The need for professionals qualified for the job market, mainly due to the advent of new technologies and productive processes, means that the quality of the education of the students who finish their studies in high school is observed first, since they are the ones who will arrive quickly in the job market. Thus, the objective of this study was to analyze the performance of high school students in the evaluation of the SAEB system of the year 2015. The performances were analyzed separately by proficiency and gender. The results indicate a strong need for adapting education policies to address the shortcomings presented and to ensure better access to the labor market by high school graduates and to provide adequate conditions for an ideal vocational training.

Keywords: School performance; SAEB Evaluation; SAEB;

1. Introdução

O uso da tecnologia em diversas operações da produção já não é novidade no mercado. Atualmente observa-se que o uso de sistema de informação, processos automatizados, robótica, internet das coisas e dos serviços vem evoluindo com o tempo e com as necessidades de mercado. Na Europa utiliza-se o conceito Indústria 4.0 para englobar atividades envolvendo automação, sistemas de informação e principalmente conectividade (AIRES; MOREIRA; FREIRE, 2017).

Os mercados estão cada vez mais necessitados de processos eficientes, que produzam resultados rápidos, seguros e lucrativos. Exemplo disso são os bancos que atualizam continuamente seus serviços baseados em sistemas de informação, buscando atrair novos clientes e passar ao mercado uma imagem de modernidade, praticidade, agilidade e segurança.

A cada ano novos profissionais entram no mercado de trabalho, com isso, existe a necessidade de se preparar um ambiente propício à absorção desse novo recurso. Além dos profissionais formados em cursos de graduação e pós-graduação, existe uma camada de indivíduos que vai buscar uma oportunidade no mercado de trabalho, mas que ainda não possui uma formação superior, são os alunos egressos do ensino médio.

O jovem egresso do ensino médio, na maioria das vezes, não possui experiência profissional, com isso fica condicionado a trabalhos que exigem um nível menor de preparo ou qualificação (MESQUITA; LELIS, 2015). Esses jovens podem desenvolver qualificações necessárias para algumas áreas do mercado de trabalho por meio de cursos técnicos em nível médio, cursos livres e até mesmo cursos vinculados ao sistema de aprendizagem industrial (SENAI).

No Brasil o adolescente pode cursar o ensino médio em escolas públicas ou na rede privada, neste caso deve arcar com os custos das mensalidades, que não são tão acessíveis à grande maioria dos jovens e de suas respectivas famílias. Observando ainda a questão do acesso ao ensino, inúmeras são as necessidades de infraestrutura nos diversos municípios brasileiros (GARCIA et al., 2014, MESQUITA; LELIS, 2015). Além disso, em alguns casos existem até mesmo a falta de professores qualificados para atuarem em determinados cursos na rede pública/privada (GATTI, 2014).

De certo que para um país evoluir em todas as suas dimensões sejam elas políticas, econômicas ou sociais, deve priorizar a educação, desde os anos iniciais da denominada Educação Básica. No Brasil a Educação Básica é organizada segundo a Lei de Diretrizes e Bases (Lei nº 9394/96) em Ensino Fundamental I e II e Ensino Médio. O Ensino Fundamental I e II abrangem os cursos do 1º ao 9º ano e o Ensino Médio envolve as três últimas séries da Educação Básica. O Jovem que deseja ingressar em um curso de nível superior normalmente deve ter concluído todos os ciclos da educação básica.

Para ingressar em um curso superior de uma universidade pública é necessário que se faça um processo seletivo denominado vestibular. Esse vestibular pode ser o grande avaliador do nível de conhecimento dos alunos egressos do ensino médio. A concorrência por uma vaga é acirrada, exemplo disso é relação vaga/candidato do curso de medicina da Universidade de São Paulo/Ribeirão Preto chegando ao nível de 75 candidatos para uma vaga (LOURENÇO,2017). Dessa forma o aluno que deseja ingressar em um curso ofertado por uma universidade pública já tem ciência que necessita de preparo acadêmico, ou seja uma base sólida de conhecimentos necessários para ser aprovado em tal vestibular. As universidades privadas também se utilizam de processo seletivo para que o indivíduo ingresse em um curso superior, embora na maioria dos casos com menor rigor em relação aos vestibulares das universidades públicas. Observa-se com o exposto que o egresso do ensino médio necessita ter um nível de conhecimento considerado essencial para o mercado de trabalho e também para o ingresso em curso superior (MESQUITA; LELIS, 2015).

O Governo Federal por meio do seu Ministério da Educação e Cultura (MEC), iniciou um processo de avaliação em larga escala para identificar quais os índices de desenvolvimento da educação com objetivo de avaliar a qualidade do ensino no país. Desde 1995, o MEC vem adaptando as avaliações conforme as necessidades nacionais. O Instituto responsável por essas avaliações em larga escala é o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) que após analisados os dados de cada avaliação, disponibiliza os resultados no seu portal (BRASIL, 2018a).

As avaliações em larga escala que o INEP disponibiliza são as do sistema SAEB que compreende os alunos do 5º ano do Ensino Fundamental 1, 9º ano do Ensino fundamental 2 e Ensino Médio (BRASIL, 2018d). Outra avaliação em larga escala é o ENEM utilizada pelos alunos do Ensino Médio como um dos critérios para o ingresso em cursos de nível superior principalmente em universidades públicas, pois a nota obtida nessa avaliação pode ser utilizada como parâmetro para o ingresso (BRASIL, 2018c). O INEP também avalia os cursos superiores por meio da avaliação ENADE (BRASIL, 2018b).

As avaliações em larga escala são utilizadas em vários países com o mesmo objetivo que é avaliar a qualidade do ensino. Exemplo disso é a avaliação PISA que é coordenada pela Organização para Cooperação e Desenvolvimento Econômico

(OCDE) e envolveu em sua última edição, no ano de 2015, 70 países (BRASIL, 2018e).

Independente do país, do sistema de ensino e da modalidade de oferta, se público ou privado, a qualidade de ensino é um dos fatores primordiais para a formação do indivíduo. Se o egresso do ensino médio precisa de oportunidades para entrar no mercado de trabalho um dos fatores que pode ser grande diferencial para o sucesso, é o seu nível de conhecimento. Uma das formas de se avaliar o nível de conhecimento dos alunos do ensino fundamental e médio, são os resultados das proficiências em avaliações de larga escala nacionais, nesse caso o do sistema SAEB (AMÂNCIO-VIEIRA et al., 2015; FONSECA; NAMEN, 2016; AMÉRICO; LACRUZ, 2017).

Observando o exposto o objetivo desse estudo foi em analisar o desempenho dos alunos egressos do ensino médio na avaliação do sistema SAEB do ano de 2015. O que se quis observar foi como esses alunos se desempenharam nas proficiências de Língua Portuguesa e Matemática, inclusive observando o fator gênero, ou seja, quais as diferenças apresentadas nessas proficiências por meninos e meninas. A Avaliação SAEB é nacional, com isso o resultado apresentado foi em nível nacional.

1.1 PNE (Plano Nacional de Educação)

No Brasil, o desenvolvimento da educação é tema de debate pelo governo, inclusive em 2014, o governo federal alterou por meio da Lei nº 13005/14 a maneira como o Plano Nacional de Educação (PNE) era conduzido, passando desde então, a ser uma exigência constitucional. O PNE determina as diretrizes, metas e estratégias relacionadas à educação para um determinado período, por exemplo, o plano que estará em vigência durante o desenvolvimento desse estudo, compreende o período de 2014 a 2024 (BRASIL, 2010; BRASIL, 2014a).

O Plano Nacional de Educação é composto por 20 metas que tem por finalidade melhorar a educação no país. Cada meta procura abranger uma dimensão necessária para a estruturação de uma política educacional estruturada e de fácil acesso aos cidadãos brasileiros. Uma das metas que converge com o objetivo desse estudo é a meta nº 7, que traz em seu texto a seguinte colocação: “fomentar a qualidade da educação básica em todas as etapas e modalidades, com melhoria do fluxo escolar e da aprendizagem, de modo a atingir as seguintes médias nacionais

para o IDEB: 6,0 nos anos iniciais do ensino fundamental; 5,5 nos anos finais do ensino fundamental; 5,2 no ensino médio” (BRASIL, 2014b).

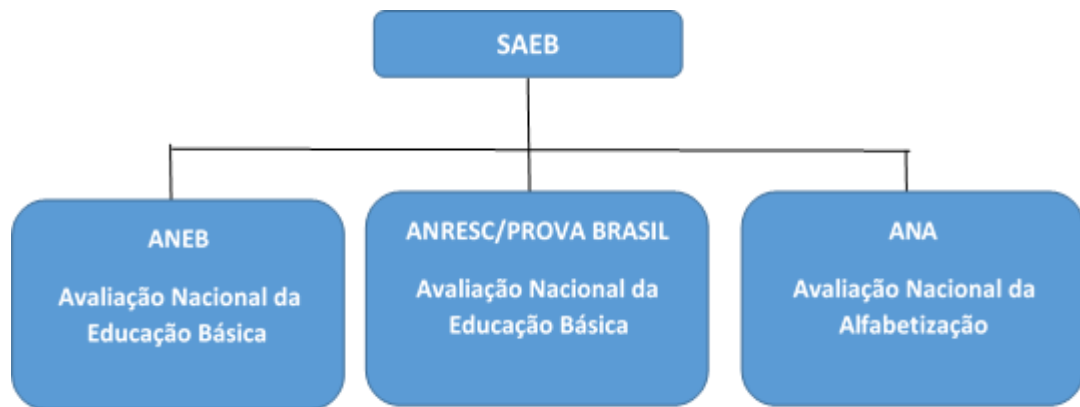
A qualidade de ensino, segundo o Plano Nacional de Educação, deve ser assegurada pelo governo de modo geral, dessa forma as avaliações em larga escala passaram a contribuir como um instrumento avaliador, fomentando informação e conhecimento necessário para que as adaptações nas políticas educacionais sejam realizadas durante a vigência do plano, objetivando sempre o alcance das metas propostas no PNE.

1.2. Sistema SAEB

O governo brasileiro, por meio do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), iniciou em 1990 um processo de avaliação de aprendizagem, com intuito de observar a qualidade do sistema de ensino. Neste ano, criou o Sistema de Avaliação da Educação Básica (SAEB), que tinha como objetivo primordial avaliar como a educação básica brasileira estava contribuindo para a formação dos indivíduos.

Em 2005, o governo decidiu reestruturar o sistema de avaliação de aprendizagem e criou dois sistemas de avaliação, um denominado Avaliação Nacional de Educação Básica (ANEB) que manteve as características iniciais do SAEB, e outra avaliação denominada Avaliação Nacional do Rendimento Escolar (ANRESC), conhecida como PROVA BRASIL, que tem como objetivo avaliar a qualidade de ensino das escolas públicas. Em 2013, outra avaliação foi incorporada ao sistema SAEB, Avaliação Nacional da Alfabetização (ANA), e tem com o objetivo medir o desempenho na alfabetização e letramento em língua portuguesa e matemática (INEP, 2018). A Figura 1 ilustra o sistema SAEB.

Figura 1 – Estrutura do Sistema de Avaliação da Educação Básica (SAEB)



Fonte: INEP (2018). Adaptado pelo autor

O sistema evoluiu desde 1990, principalmente no que se refere ao público-alvo para a realização do teste, bem como na estrutura da avaliação (base do conhecimento). A Avaliação Nacional do Rendimento Escolar - ANRESC, conhecida como Prova Brasil é uma avaliação que envolve os alunos dos 5º e 9º anos do Ensino Fundamental da rede pública. Essa avaliação, efetuada bianual, é censitária, dessa forma as escolas públicas com no mínimo 20 alunos matriculados nesses cursos devem participar.

O objetivo da ANRESC/Prova Brasil é mensurar o desempenho dos alunos nas escolas das redes públicas para avaliar os níveis de aprendizagem em relação a Língua Portuguesa (Leitura) e Matemática. Os resultados são apresentados por unidade escolar e servem para adaptação e ou atualização do sistema pedagógico utilizado por cada escola, incluindo ações de políticas públicas para a melhoria da qualidade da educação como um todo.

Já a Avaliação Nacional da Educação Básica - ANEB utiliza os mesmos instrumentos e formas da Prova Brasil, no entanto possui algumas características diferenciadoras, a primeira é que não é censitária, inclui na avaliação as escolas do setor privado e por fim, os alunos da 3ª série do Ensino Médio também participam.

Os resultados apresentados, por unidade escolar e sua localização, contribui para o entendimento de como cada região está se desenvolvendo em relação aos índices de aprendizagem dos alunos público-alvo das avaliações ANRESC/Prova Brasil e ANEB.

Outra Avaliação que compõe o sistema SAEB é a Avaliação Nacional de Alfabetização (ANA), que tem por objetivo medir os níveis de alfabetização e letramento em Língua Portuguesa e Matemática dos alunos do 3º ano do Ensino

Fundamental das escolas das redes públicas. Também censitária, com isso as escolas que possuem no mínimo 10 alunos matriculados nesse curso devem participar da avaliação. O Quadro 1 ilustra as características de cada avaliação.

Quadro 1 – Características das Avaliação do SAEB.

AVALIAÇÃO	Público Participante
ANRESC	Censitária <ul style="list-style-type: none"> ➤ 5º e 9º ano do Ensino Fundamental de escolas públicas, urbanas ou rurais, com 20 ou mais alunos matriculados no ano.
ANEB	Amostral: <ul style="list-style-type: none"> ➤ 5º e 9º ano do Ensino Fundamental de escolas públicas, urbanas ou rurais, com 10 a 19 alunos matriculados; ➤ 3ª série do Ensino Médio de escolas públicas, urbanas ou rurais, com mais de 10 alunos; ➤ 5º e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio de escolas privadas, urbanas ou rurais, com mais de 10 alunos.
ANA	Amostral <ul style="list-style-type: none"> ➤ Escolas públicas com pelo menos dez estudantes matriculados no 3º ano do Ensino Fundamental,

Fonte: INEP(2018). Adaptado pelo Autor (2018)

O INEP integrou os resultados da Prova Brasil e do Censo Escolar para viabilizar o Indicador de Desenvolvimento da Educação Básica (IDEB). Esse indicador tem como um dos princípios básicos que o aluno aprenda e passe de ano enquanto que o desempenho é medido por meio da Prova Brasil e a aprovação pelo Censo Escolar. Esse sistema composto pela Prova Brasil, Censo Escolar e IDEB propõe demonstrar um nível de qualidade do sistema de ensino brasileiro.

Os resultados das avaliações são divulgados no próprio site do INEP e podem ser utilizadas para análises e pesquisas acadêmicas, entre outros estudos. No entanto os resultados apontam as médias alcançadas pelos alunos em nível nacional, não observando diretamente os fatores ou dificuldades de cada região ou unidade escolar, o que pode comprometer de certa forma uma análise de desempenho escolar.

1.3 IDEB (Índice de Desenvolvimento da Educação Básica)

O Índice de Desenvolvimento da Educação Básica foi criado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) em 2007 e

desde sua criação tem por objetivo medir a qualidade do aprendizado pelos alunos brasileiros. Esse índice é reconhecido como um indicador que possibilita o acompanhamento da evolução da qualidade de ensino no país.

O IDEB é calculado a partir de dois itens oriundos do sistema educacional brasileiro, o primeiro é a taxa de rendimento escolar (taxa de aprovação) e o outro são as médias de desempenho nas avaliações SAEB, coordenadas pelo INEP. Os índices de aprovação são coletados no Censo Escolar (realizado anualmente), e o desempenho das avaliações SAEB, são coletados no portal INEP. Cada escola da rede ensino tem uma meta diferenciada, no entanto o objetivo é que todas alcancem a pontuação 6 até 2022. A Figura 2 apresenta a evolução do IDEB e as metas até 2021.

Figura 2 – Evolução e Metas do Índice de Desenvolvimento da educação Básica (IDEB)

Anos Iniciais do Ensino Fundamental												
IDEB Observado							Metas					
Total	2005	2007	2009	2011	2013	2015	2007	2009	2011	2013	2015	2021
		3.8	4.2	4.6	5.0	5.2	5.5	3.9	4.2	4.6	4.9	5.2
Anos Finais do Ensino Fundamental												
IDEB Observado							Metas					
Total	2005	2007	2009	2011	2013	2015	2007	2009	2011	2013	2015	2021
		3.5	3.8	4.0	4.1	4.2	4.5	3.5	3.7	3.9	4.4	4.7
Ensino Médio												
IDEB Observado							Metas					
Total	2005	2007	2009	2011	2013	2015	2007	2009	2011	2013	2015	2021
		3.4	3.5	3.6	3.7	3.7	3.7	3.4	3.5	3.7	3.9	4.3

Fonte: INEP (2018f). Adaptado pelo Autor (2018)

Pode-se observar na Figura 2, que a meta para os Anos Iniciais do Ensino Fundamental já foi alcançada em 2015 o que não aconteceu com as metas para os Anos finais do Ensino Fundamental e para o Ensino Médio, já indicando com isso, necessidades de adaptações nas políticas educacionais.

2. Percorso Metodológico

O percurso metodológico de uma pesquisa serve para orientar o pesquisador sobre a necessidade de observar passos importantes tanto antes quanto durante a execução do estudo, para que não perca o foco do que é necessário realizar para

alcançar os objetivos da pesquisa. Minayo (2010) salienta que a metodologia deve incluir o método, as técnicas e a criatividade do pesquisador. Para Gil (2010) a metodologia científica pode ser entendida como um conjunto de procedimentos tanto intelectuais quanto técnicos que o pesquisador utiliza com objetivo de alcançar o conhecimento.

Para compor esse estudo foi realizado primeiramente uma pesquisa bibliográfica em bases de dados Scielo, Portal Capes e Google Acadêmico, com objetivo de se observar as publicações que abordavam o referido tema. Utilizando as palavras chaves “desempenho escolar SAEB 2015”, “Sistema de Avaliação da Educação Básica”, “SAEB 2015” chegou-se ao resultado de 43 publicações realizadas no período compreendido entre 2015-2018. Após essa fase, foi realizada a leitura dos resumos dessas publicações com objetivo de se encontrar fatores importantes que pudessem contribuir para essa pesquisa, e também de verificar a existência de estudos semelhantes.

O resultado da pesquisa bibliográfica foi de suma importância para esse estudo, pois além de proporcionar fundamentos necessários para análise dos resultados também identificou a inexistência de um estudo com objetivo semelhante.

Após a pesquisa bibliográfica, foi realizada a coleta de dados na base de dados do INEP. Os dados coletados compreenderam os resultados da avaliação SAEB do ano de 2015, dos egressos do ensino médio. Em seguida foram realizados o processo de tratamento e normalização dos dados que compreende a exclusão de dados em branco, nulos, sem padrão exato e aqueles que por falta de alguma identificação também não puderam ser analisados. Como resultado dessa fase chegou-se ao total de 114225 avaliações registradas no banco de dados, destas 39325 foram realizadas por alunos do sexo feminino e 31373 pelos do sexo masculino.

O INEP disponibiliza junto com os resultados, um dicionário de dados que auxilia no tratamento das informações. Também é disponibilizado uma escala de proficiência tanto para Língua Portuguesa quanto para Matemática. Essa escala é composta por valores entre 0 e 500 pontos e dividida em 10 níveis. Cada nível abrange habilidades pertinentes a avaliação que o INEP divulga como sendo o mínimo necessário para que o aluno domine em termos de conhecimento. O nível 1 inicia em 225 pontos e o nível 10, pontuação igual ou superior a 450.

Segundo INEP, alunos com pontuação inferior a 225 pontos merecem uma atenção especial pois não conseguiram absorver o mínimo necessário para o seu nível de formação, sendo um indicativo que merece atenção no agrupamento e na análise de dados (BRASIL, 2018d).

Após realizados os devidos agrupamentos, acompanhando as escalas de proficiência, foram realizadas as pesquisas em banco de dados, utilizando como filtro o desempenho na proficiência, combinando o gênero e escala. Exemplo de um filtro realizado, foi contar quantos alunos obtiveram nota inferior a 225 pontos em Língua Portuguesa, separados por sexo (feminino/masculino).

Com a tabela completa, com base na estatística descritiva e nas publicações localizadas, os dados foram analisados e os resultados dessas análises apresentados no item Apresentação dos Resultados.

3. Apresentação dos Resultados

Essa fase do estudo é destinada a analisar os resultados encontrados nas pesquisas realizadas, tanto na fase bibliográfica, quanto na coleta e tratamento de dados. O primeiro passo para a análise é apresentar os resultados em formato de tabela. Na Tabela 1, os dados são apresentados devidamente agrupados por escala. Cada nível da escala abrange a quantidade de alunos da 3ª série do Ensino Médio, que obtiveram desempenho equivalente ao nível, tanto para proficiência Língua Portuguesa (LP) quanto para Matemática (MAT).

Tabela1. Quantidade de alunos em cada Proficiência do SAEB 2015.

Dados		Feminino				Masculino			
Níveis	Escalas	LP	%LP	MAT	%MAT	LP	%LP	MAT	%MAT
Nivel0	< 225	6555	16,67	6863	17,45	7476	23,83	4829	15,39
Nivel1	225 < 250	5394	13,72	8987	22,85	4470	14,25	5679	18,1
Nivel2	250 < 275	6122	15,57	8155	20,74	4736	15,1	5627	17,94
Nivel3	275 < 300	6376	16,21	5437	13,83	4419	14,09	4128	13,16
Nivel4	300 < 325	5778	14,69	3444	8,76	3902	12,44	3184	10,15
Nivel5	325 < 350	4747	12,07	2499	6,35	3278	10,45	2594	8,27
Nivel6	350 < 375	2937	7,47	1907	4,85	2078	6,62	2139	6,82
Nivel7	375 < 400	1208	3,07	1159	2,95	846	2,7	1577	5,03
Nivel8	400 < 425	203	0,52	631	1,6	165	0,53	999	3,18
Nivel9	425 < 450	5	0,01	213	0,54	3	0,01	505	1,61
Nivel10	>=450	0	0	30	0,08	0	0	112	0,36
Totais...:		39325		39325		31373		31373	

Fonte: Elaborado pelo Autor (2018)

Como primeira análise, foi percebido um percentual maior de meninas (20,22%) na Avaliação SAEB 2015, totalizando 39325 enquanto que a participação dos meninos foi de 31373.

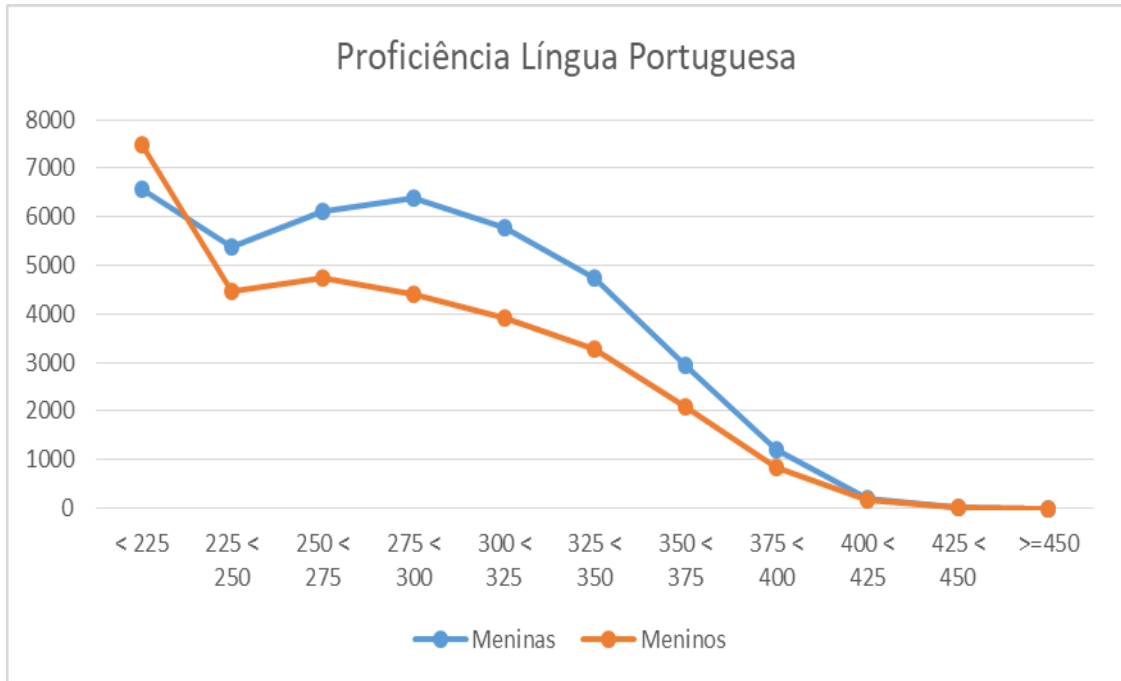
Em relação ao desempenho, uma observação importante está no fato de que 16,67% das meninas obtiveram proficiência menor que 225 em Língua Portuguesa e 17,45% em Matemática. Já para os meninos 23,83% ficaram abaixo dos 225 pontos em Língua Portuguesa e 15,39% em Matemática. Juntos totalizam 19,84% (14031 alunos) que não obtiveram pontuação nível 1 em Língua Portuguesa. E para Matemática um percentual de 16,53% (11692 alunos).

Na outra ponta da escala, no nível 10, apenas 30 alunas obtiveram desempenho maior que 450 pontos em Língua Portuguesa, já para os meninos o resultado foi um pouco melhor 112 alunos obtiveram desempenho compatível com esse nível, totalizando com isso um percentual 73,21% maior. Em Língua Portuguesa nem os meninos nem as meninas alcançaram desempenho compatível com o nível 10.

Em relação a quantidade de alunos por nível, a moda em Língua Portuguesa está no nível 0 tanto para as meninas quanto para os meninos. Em Matemática a moda está no nível 1 para ambos. Se observado a indicação do INEP para os alunos com pontuação menor que 225 (nível 0), pode-se deduzir que existe a necessidade de adaptações em políticas educacionais que possam primeiramente identificar as deficiências no sistema de ensino e posteriormente propor ações corretivas que garantam o aprendizado. As deficiências de aprendizado em Língua Portuguesa e Matemática já foram observadas por Laros e Marciano (2008; 2010) nos estudos envolvendo desempenhos dos alunos da 3ª série do ensino médio na avaliação SAEB de 2001.

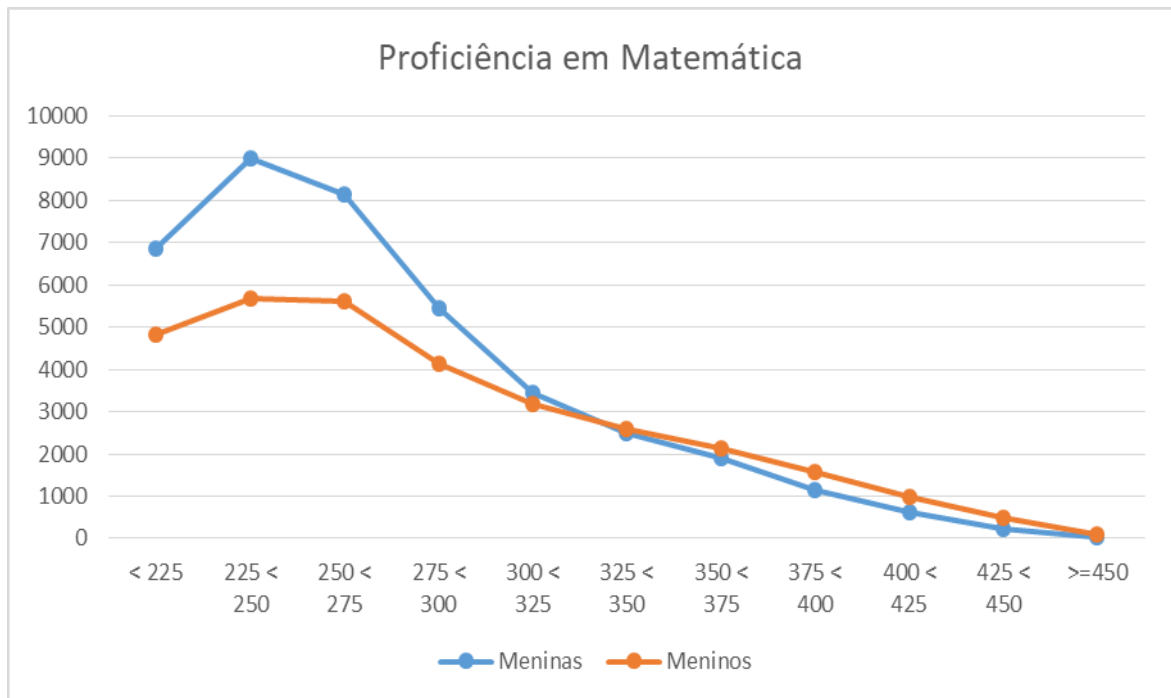
Outra análise que merece atenção está no fato de que a mediana para a proficiência em Língua Portuguesa está no nível 3 para as meninas e no nível para os meninos. Em Matemática a mediana está no nível 2 para ambos. Se observado o 3º quartil (75%), a proficiência em Língua Portuguesa para ambos está no nível 4, já para Matemática as meninas estão no nível 4 enquanto que os meninos no nível 5. Para ilustrar as Figuras 3 e 4 apresentam os gráficos relativos a quantidade de alunos em cada nível, tanto para Língua Portuguesa quanto para Matemática. A Figura 5 apresenta todas as proficiências juntas.

Figura 3 – Proficiência em Língua Portuguesa



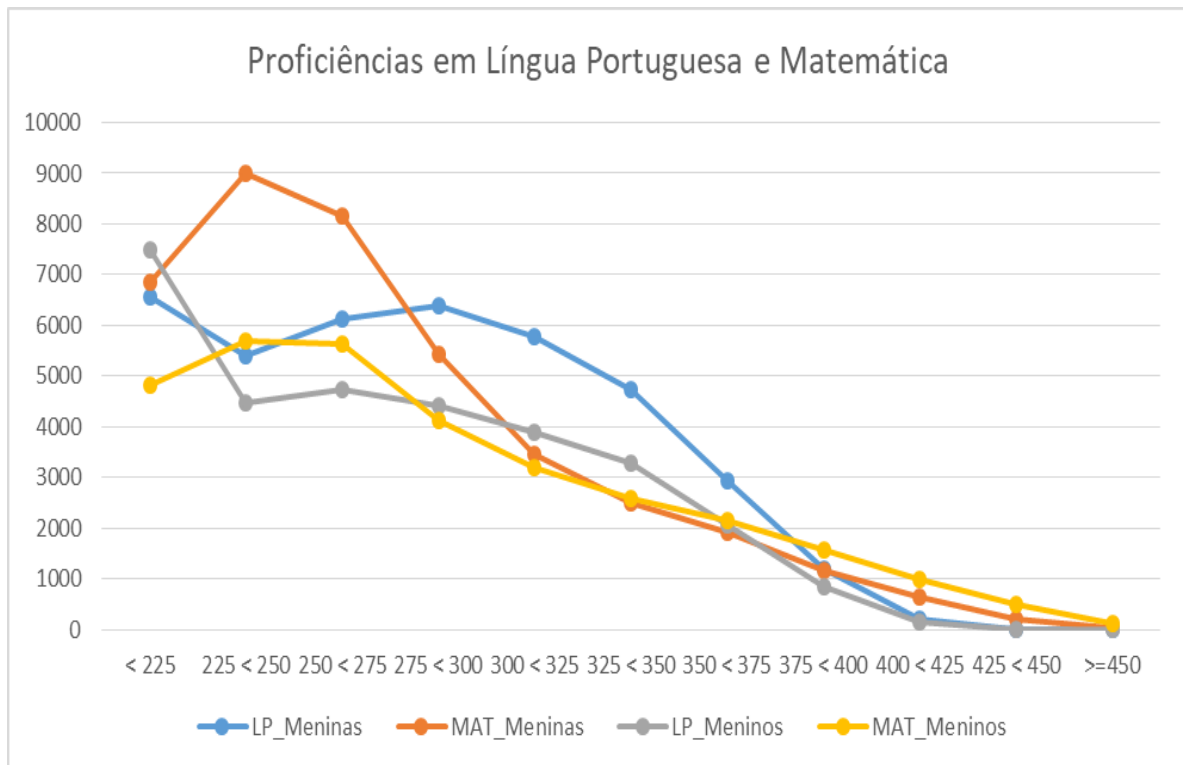
Fonte: Desenvolvido pelo Autor (2018), a partir dos resultados da Tabela 1.

Figura 4 – Proficiência em Matemática



Fonte: Desenvolvido pelo Autor (2018), a partir dos resultados da Tabela 1.

Figura 5 – Proficiências em Língua Portuguesa e Matemática



Fonte: Desenvolvido pelo Autor (2018), a partir dos resultados da Tabela 1

De acordo com a Figura 5 observa-se que existe uma quantidade maior de meninas no nível 1 (≥ 225 e < 250) indicando uma dificuldade maior de aprendizado, já sinalizado por Simões e Ferrão (2005), Andrade, Franco e Carvalho (2016), dessa forma pode-se observar a necessidade de adaptações nas políticas educacionais, nos planos curriculares e até mesmo nas práticas pedagógicas, para que em curto prazo sejam corrigidas essas deficiências de aprendizado.

Se for utilizado, apenas como comparativo, o nível 7 como um valor mínimo exigido para cada proficiência, os percentuais apresentados foram, para as meninas em Língua Portuguesa 3,07% e meninos 2,7%. Já em Matemática o percentual para as meninas foi 2,95% e para os meninos 5,03%. O índice maior apresentado no nível 7 para os meninos pode indicar um aprendizado maior em matemática em relação às meninas.

Após observados os valores apresentados na Tabela 1 é de suma importância que seja desenvolvido um plano de ação para que sejam corrigidas as deficiências tanto em Língua Portuguesa quanto em Matemática, visando com isso

dar subsídios suficiente para que esses alunos, aos finalizarem os estudos no ensino médio, possam continuar seus estudos em uma universidade pública e ainda pleitear melhores condições no mercado de trabalho.

Considerações Finais

Esse estudo teve por objetivo observar o desempenho dos alunos egressos do ensino médio e que realizaram a avaliação SAEB/2015. Os dados foram coletados, tratados e agrupados conforme escala de proficiência indicada pelo INEP. As análises dos resultados foram realizadas com base nos conceitos da estatística descritiva e comparados com outras publicações. Como limitação observada, está no fato de que não foram analisadas outras variáveis que pudessem de alguma forma gerar um viés no aprendizado dos alunos, sendo observado apenas o desempenho final na avaliação SAEB.

Após análises realizadas o que se observou foi que existe um percentual considerável de alunos que estão ainda no nível 0, em relação a moda os valores observados indicam nível para as meninas e meninos nas duas proficiências. Para mediana apenas as meninas na proficiência em Língua Portuguesa estão no nível 3, para as demais proficiências o nível observado é o 2. Outro fator importante que foi observado é que 75% dos alunos não chegaram ao nível 6 de aprendizado.

Os resultados apresentados indicam uma forte necessidade de adaptação nas políticas educacionais visando melhorias na qualidade de ensino, para que, em curto prazo sejam corrigidas todas as deficiências de aprendizado observadas. Como sugestão para um estudo futuro, está em analisar o desempenho dos alunos egressos do 9º ano do ensino fundamental, para com isso, observar se já existem deficiências de aprendizado neste período da formação.

Referências

AIRES, R. W. DO A.; MOREIRA, F. K.; FREIRE, P. DE S. INDÚSTRIA 4.0: COMPETÊNCIAS REQUERIDAS AOS PROFISSIONAIS DA QUARTA REVOLUÇÃO INDUSTRIAL. International Congress of Knowledge and Innovation - Ciki, v. 1, n. 1, 7 set. 2017.

AMÉRICO, B. L.; LACRUZ, A. J. Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla. Revista de Administração Pública, v. 51, n. 5, p. 854–878, out. 2017.

ANDRADE, M.; FRANCO, C.; CARVALHO, J. P. Gênero e Desempenho em Matemática ao final do Ensino Médio: Quais as relações? Anais, p. 1–16, 2016.

BRASIL. Ministério da Educação. Conferência Nacional de Educação. Construindo o sistema nacional articulado: o plano nacional de educação, diretrizes e estratégias de ação. Documento Final. Brasília: MEC/CONAE, 2010. 163p. Disponível em http://conae.mec.gov.br/images/stories/pdf/pdf/documetos/documento_final_sl.pdf

_____. LEI 13005, de 25 de junho de 2014. Estabelece o Plano Nacional de Educação – PNE e dá outras providências. 2014a. Disponível em http://www.planalto.gov.br/CCIVIL_03/_Ato2011-2014/2014/Lei/L13005.htm

_____. PNE. Dispõe sobre o Caderno Conhecendo as 20 Metas do Plano Nacional de Educação, texto da Lei do PNE (nº 13.005/2014); Brasília, 2014b. Disponível em <http://pne.mec.gov.br/planos-de-educacao>

_____. INEP. Dispõe sobre o Inep. 2018a. Disponível em: <http://portal.inep.gov.br/web/guest/sobre-o-inep>.

_____. INEP. Dispõe sobre o Enade. 2018b. Disponível em: <http://portal.inep.gov.br/web/guest/enade>.

_____. INEP. Dispõe sobre o Enem. 2018c. Disponível em: <http://portal.inep.gov.br/web/guest/enem>.

_____. INEP. Dispõe sobre o Saeb. 2018d. Disponível em: <http://portal.inep.gov.br/web/guest/educacao-basica/saeb>.

_____. INEP. PISA. 2018e. Disponível em: <http://portal.inep.gov.br/web/guest/pisa>.

_____. IDEB – Resultados e Metas. 2018f. Disponível em <http://ideb.inep.gov.br/resultado/resultado/resultadoBrasil.seam?cid=2562031>

SIMÕES, M.; FERRÃO, M. E. Competência percebida e desempenho escolar em Matemática. Estudos em Avaliação Educacional, v. 16, n. 32, p. 25–42, 2005.

FONSECA, S. O. DA; NAMEN, A. A. MINERAÇÃO EM BASES DE DADOS DO INEP: UMA ANÁLISE EX-PLORATÓRIA PARA NORTEAR MELHORIAS NO SISTEMA EDUCACIONAL BRASILEIRO. Educação em Revista, v. 32, n. 1, p. 133–157, mar. 2016.

GARCIA, P. S. et al. A infraestrutura das escolas de ensino fundamental da Região do Grande ABC paulista. Revista Ibero-Americana de Estudos em Educação, v. 9, n. 3, p. 614–631, 2014.

GATTI, B. A. Formação inicial de professores para a educação básica: pesquisas e políticas educacionais. Estudos em Avaliação Educacional, v. 25, n. 57, p. 24, 30 abr. 2014.

GIL, A. C. Métodos e técnicas de pesquisa social. 6.ed. São Paulo: Atlas, 2010.

LAROS, J. A.; MARCIANO, J. L. ÍNDICES EDUCACIONAIS ASSOCIADOS À PROFICIÊNCIA EM LÍNGUA PORTUGUESA: UM ESTUDO MULTINÍVEL. p. 19, 2008.

LOURENÇO, Ana. Fuvest divulga relação candidato/vaga do vestibular 2017. Guia do Estudante. Publicado em 10.nov. 2016. Atualizado em 16.mai.2017. Disponível em <https://guiadoestudante.abril.com.br/universidades/fuvest-divulga-relacao-candidatovaga-do-vestibular-2017/>.

MESQUITA, S. S. DE A.; LELIS, I. A. O. M. Cenários do Ensino Médio no Brasil. Ensaio: Avaliação e Políticas Públicas em Educação, v. 23, n. 89, p. 821–842, dez. 2015.

MINAYO, Maria Cecília de Souza (org.). Pesquisa social: teoria, método e criatividade. 29. ed. Petrópolis, RJ: Vozes, 2010.

6.2 ARTIGO 2. Predictive Success Factors in School Performance: An analysis of the large-scale assessment in Brazil

Me. Ivonaldo Vicente da Silva
Dra. Márcia Terra da Silva
Me. Saturnina Alves da Silva Martins

ABSTRACT

With the advent of new technologies and the strengthening of the concept of Industry 4.0, it becomes important to observe the issue of the need for more qualified and up-to-date professionals concerning the labor market. Thus, there is a need to investigate the level of school performance, especially among high school graduates, who will be the ones who will most quickly enter the labor market. In order to compose the research data, the results of the SAEB large-scale evaluation of the year 2015 in the Portuguese Language and Mathematics proficiency were used. The data were collected, standardized and the results presented in groupings by levels, obeying a scale of 0 to 10. RapidMiner software was used for data mining and decision tree construction. The central objective was to identify the best predictive factors for students' performance equal to or greater than 350 points (level 6). The results showed that there is a significant percentage of students with scores lower than 225 points in the two proficiencies analyzed, considered by the government as a concern, since these students did not learn the necessary minimum. The decision trees demonstrated that the level of schooling of the parents or guardians of students can be considered important and fundamental factors for school performance

KEYWORDS: School Performance; SAEB; Data Mining;

1. INTRODUCTION.

In Brazil, the young person who wishes to complete a higher education course must pass through all the basic education cycles, composed of Elementary and Middle School. The Elementary School consists of a cycle of nine grades and the High School for three. In some states, the fourth grade of high school is already being tested.

Based on the need for better monitoring of the quality of education, the government created an indicator called the Basic Education Performance Indicator (IDEB) and established it through the National Education Plan - PNE (MEC, 2014). One of the goals of the PNE is that the IDEB index for national education should be 6.0 by the year 2022. Each Brazilian state has its proportional target, but that converges to the national target set for 2022, however in 2015 no state has reached its goal for high school

The IDEB index is composed of data from the School Census and the performance in the large-scale evaluation of the Basic Education Evaluation System (SAEB). This system evaluates students graduating from Elementary and High School (INEP, 2018a).

Each year the labor market receives several individuals, high school graduated, looking for occupation and professional development. A large proportion of these individuals do not yet have professional qualifications, but they need to start their careers.

It is known that the Brazilian educational system suffers from several deficiencies, as its structure, teaching training, safety, access to technological means among other factors. However, several authors corroborate the fact that preparing the young, and future professionals for the job market requires a model of education based on quality and directed towards the changes of this market (Castro and Carvalho, 2013; Alves and Silva, 2013 ; Ferraz and Abreu, 2015; Hypolito and Hypolito, 2015)

Some studies point out that family participation can be a crucial factor in students' school performance, because when there is better monitoring and interest, especially parental participation, in the school routine of children and young people, it is observed that school performance (Nogueira et al. 2015, Gomes, 2016; Zafani and Omote, 2016; Mato et al, 2017; Silveira and Brittes, 2017).

With the advent of the so-called Fourth Industrial Revolution, the labor market starts to demand more skills and abilities from the new professional. Competencies and Skills based on communication, technological knowledge and abstraction power to understand the environment based on logical and mathematic reasoning, become basic prerequisites for the individual ables to maintain and design his professional career (Sátyro et al. 2018)

The high school graduate, usually still without much qualification, needs support and guidance in order to address the labor market, since on the one hand there is a lack of experience and on the other a labor market that seeks increasingly qualified professionals (Comazzetto et al. 2016). It is important to note that the individual with the highest educational level and professional qualification is more likely to evolve professionally. (Sabóia et al. 2016; Messias and Santos, 2018).

Based on a scenario that places on the opposite sides the young person with no professional experience and the highly selective labor market, it is necessary to

investigate the level of school performance of high school graduates and identify the factors that contribute or influence their performance. The population universe was composed of the results of the proficiencies in Portuguese Language and Mathematics of the students participating in the SAEB evaluation of the year 2015.

The results of this work will be able to contribute with greater knowledge regarding the difficulties of learning in each proficiency analyzed and the factors that contribute to better performance in large scale evaluation, so that the government and society can adapt the public policies, in order to strengthen the work in these predictive factors and in others that are directly related to the school performance.

2. BODY OF PAPER

2.1 Methodological Path

In order for the researcher to be able to conduct his or her work in a targeted manner and to mitigate possible failures, both in data collection and in interpretations of results, it is necessary to define the entire methodological course, including method, techniques and, in particular, researcher creativity, always observing the objectives one wishes to attain.

Firstly, for the development of this study, bibliographic research was carried out in databases such as Scielo, Google Scholar and Web of Science in order to identify the predictive factors of school performance.

Next, SAEB2015 data were collected on the website of the National Institute of Anísio Teixeira - INEP, including complementary documents such as data dictionaries, proficiency scales, among others (INEP, 2018b). The results range from 0 to 500 and are organized in a range of 25 points, with the lowest, from 0 to 225, being considered as level zero. 70698 students participated in this edition, 39326 girls and 31373 boys.

Table 1 presents the results for the proficiencies in Portuguese Language and Mathematics organized by level and gender of the students. The groupings were performed according to the performance of each proficiency, that is, in each level was described the total number of students who had a final grade within the analyzed range. Ex: Level 1 $> = 225$ and < 250 points.

After this phase, in order to try to identify important factors in the prediction of good results in the SAEB evaluation, some attributes were separated, according to Frame 1, for the construction of Decision Trees with support of the data mining

technique. The software used for data mining was RapidMiner. The results will be analyzed in the topic Analysis of Results.

Table 1. Number of students in each SAEB 2015 Proficiency

Scale	Girls				Boys			
	PORT	% PORT	MAT	% MAT	PORT	% PORT	MAT	% MAT
Level 0 < 225	6555	16,669	6863	17,452	7476	23,829	4829	15,392
Level 1 225 < 250	5394	13,716	8987	22,853	4470	14,248	5679	18,102
Level 2 250 < 275	6122	15,568	8155	20,737	4736	15,096	5627	17,936
Level 3 275 < 300	6376	16,214	5437	13,826	4419	14,085	4128	13,158
Level 4 300 < 325	5778	14,693	3444	8,758	3902	12,437	3184	10,149
Level 5 325 < 350	4747	12,071	2499	6,355	3278	10,448	2594	8,268
Level 6 350 < 375	2937	7,469	1907	4,849	2078	6,624	2139	6,818
Level 7 375 < 400	1208	3,072	1159	2,947	846	2,697	1577	5,027
Level 8 400 < 425	208	0,529	631	1,605	168	0,535	999	3,184
Level 9 425 < 450			213	0,542			505	1,610
Level 10 >=450			30	0,076			112	0,357
Total...:	39325	100,00	39325	100,00	31373	100,00	31373	100,00

Note: In Portuguese Language (PORT) Level 8 includes all students with a mark above 400

Source: Developed by the Author (2018)

2.2 Analysis of Results

It is observed a high percentage of students in level 0, (see Table 1), considered by the government as a worrying factor because at this level, the student did not learn the minimum necessary for the proficiency. In Portuguese Language, the total number of girls at level zero was 16.67% and boys 23.83%, whereas, in math, 17.45% of the girls and 15.39% of the boys have this low evaluation.

If the scale used for analysis is Level ≥ 6 of the scale for proficiency, in Portuguese language the total number of girls was 11.07% and boys 9.86%. In Mathematics, the total number of girls was 10.02% and boys 17.0%. Both the large number of students at level zero and the low number of students at the level ≥ 6 are worrisome. The differences in performance in Mathematics and Portuguese Language, among boys and girls, have already been observed in the studies of (Laros and Marciano, 2008; Laros et al. 2010), in studies involving high school students in the SAEB 2001 edition.

Then the predominant factor was verified so that the students obtained better performances in this evaluation. In this way it was necessary to use the data mining technique.

Data mining can be used to discover patterns in databases, and from these standards generate sufficient and necessary knowledge in support of decision-making (Silva et al. 2016; Provost and Fawcett, 2016).

The variables/attributes were chosen and prepared based on the forms applied to the students and the data/responses for each variable/attribute were collected from the results available by INEP (INEP, 2018b). The variables/attributes used for this study are described in Frame 1.

Frame 1. Variables/Attributes used for the Decision Tree

Description	ShortName
What is your gender?	SEX
Have your mother, or the woman responsible for you, studied until what grade?	MOTHER_STUDY
Have your father, or the man responsible for you, studied until what grade?	FATHER_STUDY
Do your parents or guardians encourage you to study?	PARENTS_ENCOURAGE_STUDY
Do your parents or guardians encourage you to do homework or school tasks?	PARENTS_ENCOURAGE_HOMEWORK
Do your parents or guardians encourage you to read?	PARENTS_ENCOURAGE_READ
Do your parents or guardians encourage you to go to school or not miss classes?	PARENTS_ENCOURAGE_GOTOSCHOOL
Do your parents or guardians talk to you about what happens at school?	PARENTS_TALK_ABOUTSCHOOL
Do you like to study Portuguese Language?	LIKE_PORTUGUESE
Do you do your homework of Portuguese Language?	DO_HOMEWORK_PORTUGUESE
Does the teacher correct homework in Portuguese Language?	TEACHER_CORRECT_PORTUGUESE
Do you like to study Mathematics?	LIKE_MATHEMATICS
Do you do Mathematics homework?	DO_HOMEWORK_MATHEMATICS
Does the teacher correct Mathematics homework?	TEACHER_CORRECT_MATHEMATICS
Do you use your school's library or reading room?	USE_LIBRARY
Level of Proficiency (Level <6; Level >= 6)	LEVEL_PROFICIENCY

Source: Developed by the Author (2019)

Data normalization process was performed so that there was no inconsistency and variable/attribute with no response. In this way, all the data that were duly completed were used.

RapidMiner software was used and the technique used was classification. Data were separated into two groups one for training and one for testing. From these groups the decision trees were created, one with data for proficiency in Portuguese Language and another for Mathematics. The results were grouped using level 6 of the proficiency scale as a parameter, so the label parameter of analysis was the level on the proficiency scale (Level < 6 or Level >= 6), with the objective of identifying the predominant factors for students to perform better. The results are described in

Frames 2 and 3. The accuracy percentage for Portuguese Language was 89.04% and for Mathematics 86.42%.

Frame 2 Decision Tree for Proficiency Portuguese Language

```

MOTHER_STUDY = <5th grade Elementary School: <Level6 {<Level6=6412, >= Level6=128}
MOTHER_STUDY = >=5th grade Elementary School e <9º Ano: <Level6 {<Level6=6385, >= Level6=293}
MOTHER_STUDY = >=9th grade Middle School e <High School: <Level6 {<Level6=5169, >= Level6=387}
MOTHER_STUDY = >=High School e <College degree: <Level6 {<Level6=14346, >= Level6=1958}
MOTHER_STUDY = >=College degree
| FATHER_STUDY = <5th grade Elementary School: <Level6 {<Level6=352, >= Level6=29}
| FATHER_STUDY = >=5th grade Elementary School e <9º Ano
| | PARENTS_ENCOURAGE_STUDY = No: <Level6 {<Level6=4, >= Level6=3}
| | PARENTS_ENCOURAGE_STUDY = Yes
| | | PARENTS_ENCOURAGE_GOTOSCHOOL = No: <Level6 {<Level6=9, >= Level6=0}
| | | PARENTS_ENCOURAGE_GOTOSCHOOL = Yes
| | | | PARENTS_ENCOURAGE_HOMEWORK = No
| | | | | PARENTS_TALK_ABOUTSCHOOL = No
| | | | | TEACHER_CORRECT_PORTUGUESE = Sometimes: <Level6 {<Level6=6, >= Level6=0}
| | | | | TEACHER_CORRECT_PORTUGUESE = Never or almost never: <Level6 {<Level6=1, >= Level6=1}
| | | | | TEACHER_CORRECT_PORTUGUESE = Teacher does not give homework to do at home: >= Level6
| | | | | {<Level6=0, >= Level6=2}
| | | | | TEACHER_CORRECT_PORTUGUESE = Always or almost always: <Level6 {<Level6=8, >= Level6=3}
| | | | | PARENTS_TALK_ABOUTSCHOOL = Yes: <Level6 {<Level6=12, >= Level6=1}
| | | | | PARENTS_ENCOURAGE_HOMEWORK = Yes: <Level6 {<Level6=572, >= Level6=61}
| FATHER_STUDY = >=9th grade Middle School e <High School: <Level6 {<Level6=759, >= Level6=153}
| FATHER_STUDY = >=High School e <College degree: <Level6 {<Level6=2924, >= Level6=781}
| FATHER_STUDY = >=College degree: <Level6 {<Level6=3667, >= Level6=1636}
| FATHER_STUDY = Never attended a school: <Level6 {<Level6=43, >= Level6=4}
| FATHER_STUDY = I don't Know: <Level6 {<Level6=736, >= Level6=98}
MOTHER_STUDY = Never attended a school: <Level6 {<Level6=1191, >= Level6=14}
MOTHER_STUDY = I don't Know: <Level6 {<Level6=3206, >= Level6=87}

```

Source: Developed by the Author (2019)

Only for didactic purposes, we will use the term SUCCESS for the level of proficiency equal to or greater than 6 (Level > = 6), both for Portuguese Language and Mathematics.

It can be observed that, in the Portuguese Language case, the most critical factor presented in the decision tree is the MOTHER_STUDY. The positive variation of the mothers' schooling provides an increase of percentage of the students' success. The first four rows of the decision tree indicate that this variable alone can be used to predict the percentage of SUCCESS of the group of students.

For the condition MOTHER_STUDY="Finalized the College degree" the percentage of SUCCESS reaches 30%, the greater percentage among all, when the condition FATHER_STUDY="Finalized the College degree" is also true.

To complete, if the condition MOTHER_STUDY="Never Studied" the percentage of SUCCESS reaches 1.6%, and if MOTHER_STUDY="Not Finalized 5th grade

Elementary School" to 4.38%, thus reinforcing the importance of the schooling level of mother or woman responsible for the student.

Some variables/attributes were not relevant to the point of appearing in the decision tree, some of which could initially have been considered as factors for school performance, such as if the student enjoys Portuguese Language (LIKE_PORTUGUESE) and uses the school library (USE_LIBRARY). From the results presented, it can be assumed that these factors are not essential for good performance in Portuguese Language proficiency.

Based on the results presented, it is clear that the schooling of the mother or woman responsible for the student is fundamental for the school performance, a fact already observed by some authors (Luiz, 2006; Gerard and Booth, 2015; Benner et al. 2016; Lobo, 2017; Fantinato and Cia, 2017)

Then the decision tree for mathematical proficiency will be analyzed.

Frame 3 Decision Tree for Mathematical Proficiency

```

LIKE_MATHEMATICS = No: <Level6 {<Level6=21310, >=Level6=1465}
LIKE_MATHEMATICS = Yes
| FATHER_STUDY = <5th grade Elementary School: <Level6 {<Level6=3860, >=Level6=214}
| FATHER_STUDY = >=5th grade Elementary School e <9º Ano: <Level6 {<Level6=3492, >=Level6=383}
| FATHER_STUDY = >=9th grade Middle School e <High School
| | MOTHER_STUDY = <5th grade Elementary School: <Level6 {<Level6=320, >=Level6=8}
| | MOTHER_STUDY = >=5th grade Elementary School e <9º Ano: <Level6 {<Level6=503, >=Level6=52}
| | MOTHER_STUDY = >=9th grade Middle School e <High School: <Level6 {<Level6=674, >=Level6=88}
| | MOTHER_STUDY = >=High School <College degree: <Level6 {<Level6=928, >=Level6=189}
| | MOTHER_STUDY = >=College degree
| | | TEACHER_CORRECT_MATHEMATICS = Sometimes: <Level6 {<Level6=42, >=Level6=18}
| | | TEACHER_CORRECT_MATHEMATICS = Never or almost never: >=Level6 {<Level6=6, >=Level6=10}
| | | TEACHER_CORRECT_MATHEMATICS = Teacher does not give homework to do at home: <Level6
| | | {<Level6=14, >=Level6=7}
| | | TEACHER_CORRECT_MATHEMATICS = Always or almost always: <Level6 {<Level6=297, >=Level6=107}
| | MOTHER_STUDY = Never attended a school: <Level6 {<Level6=25, >=Level6=2}
| | MOTHER_STUDY = I don't Know: <Level6 {<Level6=47, >=Level6=2}
| FATHER_STUDY = >=High School <College degree
| | MOTHER_STUDY = <5th grade Elementary School: <Level6 {<Level6=272, >=Level6=26}
| | MOTHER_STUDY = >=5th grade Elementary School e <9º Ano: <Level6 {<Level6=506, >=Level6=73}
| | MOTHER_STUDY = >=9th grade Middle School e <High School: <Level6 {<Level6=635, >=Level6=148}
| | MOTHER_STUDY = >=High School <College degree: <Level6 {<Level6=3160, >=Level6=896}
| | MOTHER_STUDY = >=College degree
| | | TEACHER_CORRECT_MATHEMATICS = Sometimes: <Level6 {<Level6=172, >=Level6=120}
| | | TEACHER_CORRECT_MATHEMATICS = Never or almost never: <Level6 {<Level6=34, >=Level6=23}
| | | TEACHER_CORRECT_MATHEMATICS = Teacher does not give homework to do at home: >=Level6
| | | {<Level6=29, >=Level6=54}
| | | TEACHER_CORRECT_MATHEMATICS = Always or almost always: <Level6 {<Level6=1120, >=Level6=570}
| | MOTHER_STUDY = Never attended a school: <Level6 {<Level6=22, >=Level6=2}
| | MOTHER_STUDY = I don't Know
| | | PaisIncentivamEstudar = No: >=Level6 {<Level6=1, >=Level6=2}
| | | PaisIncentivamEstudar = Yes: <Level6 {<Level6=124, >=Level6=10}
| FATHER_STUDY = >=College degree
| | MOTHER_STUDY = <5th grade Elementary School: <Level6 {<Level6=47, >=Level6=6}
| | MOTHER_STUDY = >=5th grade Elementary School e <9º Ano: <Level6 {<Level6=96, >=Level6=32}
| | MOTHER_STUDY = >=9th grade Middle School e <High School: <Level6 {<Level6=173, >=Level6=78}
| | MOTHER_STUDY = >=High School <College degree: <Level6 {<Level6=918, >=Level6=602}
| | MOTHER_STUDY = >=College degree: >=Level6 {<Level6=1499, >=Level6=1585}
| | MOTHER_STUDY = Never attended a school: <Level6 {<Level6=8, >=Level6=2}

```


3. CONCLUSIONS

The objective of this study was to analyze the students' performance in the SAEB 2015 large scale evaluation and to identify predictive factors of success in relation to students' performance in Portuguese Language and Mathematics proficiencies.

Initially the proficiency results showed that there is a high percentile of students, (reaching 23.8% in Portuguese Language, for boys), who presented performance compatible with level zero in both proficiencies. Likewise, the small number of students who reached a performed above level 6 is also worrisome.

In this way, it becomes necessary to have a more judicious view that can identify the deficiencies in the educational environment that may impair the performance of the students, and from that, propose changes in the educational plans and even in public policies that can offer better conditions for the learning at all levels of education (Oliveira, 2010; Oliveira and Waldhelm, 2016).

After the phases described in the methodological course, the results presented in the decision trees indicate that the level of education of parents or guardians of students is a primary factor and influences school performance. When looking at the decision trees, the results indicate that the family's role in accompanying the student's school routine may be preponderant so that the student has condition to develop and project his professional career.

The labor market, more and more selective, requires skilled and updated professionals, especially in relation to new technologies and innovations, so it is necessary that the young graduates of High School have a minimum of condition, in terms of knowledge, to work in these companies. What has been observed is that the level of proficiency presented mainly in Mathematics becomes worrisome because it may be that this egress is not minimally prepared to face professional challenges that require a higher level of logical reasoning and mathematical calculations. In an environment in constant technological change, factors related to mathematical and logical reasoning may be fundamental for the maintenance of the individual in the labor market. Inclusive the level of performance presented may be an influential factor in the choice of profession (Araujo et al. 2018).

The results found may contribute to future research in the area of education, principally in the sense of improving public policies aimed at education and the participation of the family in the student's school routine. As a limitation, there is the

question of the population universe of this edition of the SAEB, with the suggestion that this study can be replicated in other large-scale evaluations, including other educational levels, such as the 5th year of Elementary Education or international like PISA

ACKNOWLEDGMENT

This work was carried out with the support of the Coordination of Improvement of Higher Level Personnel - Brazil (CAPES) - Financing Code 001

REFERENCES

- Alves, T. and Silva, R.M. da 2013. Estratificação das oportunidades educacionais no Brasil: contextos e desafios para a oferta de ensino em condições de qualidade para todos. *Educação & Sociedade* 34(124), pp. 851–879.
- Araújo, A.J.N. et al. 2018. Ensino profissionalizante, desempenho escolar e inserção produtiva : uma análise com dados do ENEM. <http://ppe.ipea.gov.br> . Available at: <http://repositorio.ipea.gov.br/handle/11058/8353>
- Castro, J.A. de and Carvalho, C.H.A. de 2013. Necessidades e possibilidades para o financiamento da educação brasileira no plano nacional de educação. *Educação & Sociedade* 34(124), pp. 829–849. doi: 10.1590/S0101-73302013000300010.
- Comazzetto, L.R. et al. 2016. A Geração Y no Mercado de Trabalho: um Estudo Comparativo entre Gerações. *Psicologia: Ciência e Profissão* 36(1), pp. 145–157.
- Fantinato, A.C. and Cia, F. 2017. Envolvimento parental, competência social e o desempenho acadêmico de escolares. *Psicologia Argumento* 29(67). Available at: <https://periodicos.pucpr.br/index.php/psicologiaargumento/article/view/20401>
- Ferraz, R.R.N. and Abreu, J.S. 2015. QUANDO A CULPA DO MAU DESEMPENHO ESCOLAR NÃO É EXCLUSIVAMENTE DO ALUNO. *UNILUS Ensino e Pesquisa* 12(28), pp. 96–100.
- GIL, A.C. 2010. Métodos e técnicas de pesquisa social. 6th ed. São Paulo: Atlas.
- Gomes, J.A.M. 2016. OS RECURSOS DO AMBIENTE FAMILIAR E A INFLUÊNCIA NO RENDIMENTO ACADÊMICO. *Revista Lugares de Educação* 6(12), pp. 103–124.
- Hypolito, Á.L.M. and Hypolito, Á.L.M. 2015. Trabalho docente e o novo Plano Nacional de Educação: valorização, formação e condições de trabalho. *Cadernos CEDES* 35(97), pp. 517–534. doi: 10.1590/CC0101-32622015150376.
- INEP 2018a. IDEB - SAEB. Available at: <http://portal.inep.gov.br/web/guest/educacao-basica/saeb>
- INEP 2018b. INEP - Microdados. Available at: <http://portal.inep.gov.br/web/guest/microdados>.
- Lakatos, E.M. and Marconi, M. de A. 2017. Fundamentos de Metodologia Científica. Edição: 8a. Atlas.
- Laros, J.A. and Marciano, J.L. 2008. ÍNDICES EDUCACIONAIS ASSOCIADOS À PROFICIÊNCIA EM LÍNGUA PORTUGUESA: UM ESTUDO MULTINÍVEL. *Avaliação Psicológica*. p. 19. Available at: http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S1677-04712008000300010&lng=en&nrm=iso&tlng=pt

Laros, J.A. and Marciano, J.L.P. Moura, J.A 2010. FATORES QUE AFETAM O DESEMPENHO NA PROVA DE MATEMÁTICA DO SAEB: UM ESTUDO MULTINÍVEL. Avaliação Psicológica, p. 14. Available at: <http://www.redalyc.org/resumen.oa?id=335027283004>

Lobo, G.D. 2017. Avaliação do Desempenho Escolar dos Estudantes da Região Nordeste que Realizaram o ENEM: Uma Análise com Modelos Hierárquicos. Revista Espacios 38, p. 13.

LUZ, L.S. 2006. OS DETERMINANTES DO DESEMPENHO ESCOLAR: A ESTRATIFICAÇÃO EDUCACIONAL E O EFEITO VALOR ADICIONADO. XV Encontro Nacional de Estudos Populacionais, ABEP Caxambu-MG

Matos, D.A.S. et al. 2017. Impactos das práticas familiares sobre a proficiência em Língua Portuguesa e Matemática no Ensino Fundamental. Available at: <http://www.repositorio.ufop.br/handle/123456789/9797>.

MEC 2014. PNE - Plano Nacional de Educação - Plano Nacional de Educação - Lei nº 13.005/2014. Available at: <http://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>

Messias, J.F. and Santos, O.S. dos 2018. O DÉFICIT EDUCACIONAL E A EXCLUSÃO TECNOLÓGICA NO MERCADO DE TRABALHO BRASILEIRO. Brasil Para Todos - Revista Internacional 6(1). Available at: https://ojs.eniac.com.br/index.php/Anais_Sem_Int_Etn_Racial/article/view/536

MINAYO, M.C. de S. (or). 2010. Pesquisa social: teoria, método e criatividade. 29th ed. Petrópolis/RJ: Vozes.

Nogueira, R.K.S. et al. 2015. FAMÍLIA E ESCOLA: RELAÇÃO QUE PROMOVE O BOM RENDIMENTO ESCOLAR DOS ALUNOS DE DUAS TURMAS DO ENSINO FUNDAMENTAL EM UMA ESCOLA DA REGIÃO ARAPIRAQUENSE. Anais do Congresso de Inovação Pedagógica em Arapiraca 1(1). Available at: <http://www.seer.ufal.br/index.php/cipar/article/view/1916>

Oliveira, A.C.P. de and Waldhelm, A.P.S. 2016. Liderança do diretor, clima escolar e desempenho dos alunos: qual a relação? Ensaio: Avaliação e Políticas Públicas em Educação 24(93), pp. 824–844.

Oliveira, J.A.P.F. 2016. O QUE É UM BOM PROFESSOR, NA VISÃO DO ALUNO: um estudo de caso em uma Instituição de Ensino Superior em Oliveira-MG. REVISTA ACADÊMICA FEOL 1(2), pp. 47-67–67.

Oliveira, L.M. de 2010. Políticas educacionais na formação da professora dos anos iniciais do ensino fundamental em cursos de licenciatura. Ensaio: Avaliação e Políticas Públicas em Educação 18(67), pp. 235–252.

Provost, F. and Fawcett, T. 2016. Data Science para Negócios. Rio de Janeiro: Alta Books.

Sabóia, A.L. et al. 2016. Adolescentes e jovens no Brasil: escolarização e inserção no mercado de trabalho. Anais, pp. 1–17.

Sátyro, W.C. et al. eds. 2018. Indústria 4.0: Conceitos e Fundamentos. Edição: 1a. Blucher.

Silva, L.A. da et al. 2016. Introdução à Mineração de Dados. Rio de Janeiro: Elsevier. Available at: <https://www.loja.elsevier.com.br/introducao-a-mineracao-de-dados-9788535284461.html>.

Silveira, R.B. and Brittes, L.R. 2017. A PARTICIPAÇÃO DA FAMÍLIA NA ESCOLA: DESDOBRAMENTOS SOBRE A EVASÃO ESCOLAR E A EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA NA LÓGICA NEOLIBERAL. Educação Básica Revista 3(1), pp. 29–46.

Sousa, M. de. 2017. Eficácia escolar e o desempenho de meninos e meninas. Aleph , p. 55

6.3 ARTIGO 3. Ensino Médio e as Avaliações em Larga Escala do SAEB: Uma análise do desempenho apresentado no período 2005 a 2017.

HIGH SCHOOL EDUCATION AND SAEB LARGE SCALE EVALUATIONS: AN ANALYSIS OF THE PERFORMANCE PRESENTED IN THE PERIOD 2005 TO 2017.

Ivonaldo Vicente da Silva

Márcia Terra da Silva

RESUMO

A proposta desse estudo foi analisar os desempenhos dos egressos do Ensino Médio, nas avaliações em larga escala do SAEB. A análise foi realizada em duas etapas, a primeira envolveu os desempenhos médios estaduais e regionais no período de 2005 a 2017, e na segunda, foram analisados os desempenhos dos egressos nas edições de 2005 e 2017. Por meio do método de pesquisa mista (quanti-quali), envolvendo estatística descritiva e análise da literatura, os resultados encontrados, indicaram que os níveis de desempenhos apresentados, tanto dos alunos quanto dos estados e regiões, em sua maioria não ultrapassaram os 325 pontos, ficando no máximo no nível 4 da escala de desempenho. Os resultados desse estudo podem contribuir para o desenvolvimento de ações no ambiente escolar, que possam envolver governo, profissionais da educação, famílias e comunidade, com intuito de criar alternativas para proporcionar melhores condições de aprendizagem aos alunos do Ensino Médio.

Palavras-chave: Desempenho Escolar; Avaliação em larga escala; Ensino Médio;

ASBTRACT

The purpose of this study was to analyze the performance of high school graduates, in the large-scale evaluations of SAEB. The analysis was carried out in two stages, the first involved the average state and regional performances in the period from 2005 to 2017, and in the second, the performances of the graduates in the 2005 and 2017 editions were analyzed. Using the mixed research method (quanti-quali), involving descriptive statistics and literature analysis, the results found, indicated that the performance levels presented, both by students and by states and regions, mostly did not exceed 325 points, staying at most at level 4 performance scale. The results of this study can contribute to the development of actions in the school environment, which may involve government, education professionals, families and the community, in order to create alternatives to provide better learning conditions for high school students.

Keywords: School Performance; Large-scale evaluation; High school;

1. Introdução

Desde o início do século XIX o homem procura melhorar e aperfeiçoar os sistemas de produção. As Revoluções Industriais contribuíram, cada uma ao seu modo e foco, para a evolução dos sistemas de produção. Em um primeiro período, de 1760 a 1860, surgiu na Europa a máquina de tear mecânica, a indústria têxtil (tecidos de algodão) e a máquina vapor, esse período foi denominado como primeira Revolução Industrial.

Denominada como segunda Revolução Industrial, o período de 1860 a 1900 teve um papel significativo para os sistemas de produção pois neste momento foram introduzidos como matérias-primas produtivas o aço, a energia elétrica e os combustíveis baseados no petróleo, além disso, os produtos químicos e a locomotiva a vapor foram as grandes sensações da época. A terceira Revolução Industrial, englobando o período do século XX e XXI, foi marcada pelo surgimento do computador e das diversas tecnologias afins. Outro item de grande impacto nesse período foi o surgimento do celular, indispensável para o mundo moderno.

A partir dos anos 1980, o uso da tecnologia em diversas operações da produção se torna essencial. O uso de sistema de informação, processos automatizados, robótica, internet das coisas e dos serviços vem evoluindo com o tempo e com as necessidades de mercado. Na Europa, a partir dos anos 2000, o conceito de Indústria 4.0 passa a ser utilizado para englobar atividades produtivas envolvendo automação, sistemas de informação e principalmente conectividade (AIRES, MOREIRA e FREIRE, 2017; SACOMANO et al., 2018).

O cenário econômico proveniente desse período cria uma lacuna de produção que necessita ser rapidamente preenchida, no entanto, para se obter uma maior produção são necessários melhores processos produtivos, e por conta disso surge a demanda por profissionais mais qualificados e preparados para as novas tecnologias e inovações (BALTAR, 2015; COMAZZETTO et al., 2016), e qualificação normalmente está atrelada à formação educacional.

Sobre a formação educacional, no Brasil o ciclo básico da educação é composto pelos cursos do Ensino Fundamental em nove anos, e pelo Ensino Médio que normalmente é ofertado em um ciclo de três anos (BRASIL, 1996, p. 9) . Esses cursos são ofertados pelo governo e pela iniciativa privada, sendo que a maioria dos alunos estão matriculados em escolas públicas, podendo essas serem administradas pelo governo federal, estadual ou municipal.

No entanto, o jovem ao terminar o Ensino Médio normalmente não possui experiência profissional, e em um mercado com poucas ofertas para iniciantes, a condição da inexperiência proporciona uma alta disponibilidade de indivíduos com pouca ou quase nenhuma qualificação, mas que ainda necessita trabalhar (MESQUITA e LELIS, 2015). Uma alternativa para que esses jovens possam obter qualificação necessária para alguns postos de trabalho, são os cursos técnicos ou cursos vinculados ao Sistema de Aprendizagem Industrial (SENAI).

Além dos cursos técnicos, os jovens também podem ao finalizar o Ensino Médio, ingressar em um curso de graduação. Esses cursos são ofertados tanto em instituições públicas quanto particulares. É importante observar que o jovem egresso do Ensino Médio, necessita de conhecimentos essenciais para o ingresso em um curso de graduação, da mesma forma que precisa ter conhecimentos mínimos para exercer atividades em um posto de trabalho (MESQUITA e LELIS, 2015; BALTAR, 2015; BORGES, 2018; MESSIAS e SANTOS, 2018).

E a necessidade de qualificação para o exercício laboral reforça, como pré-requisito, que esse indivíduo tenha tido condições de desfrutar de um modelo de educação com qualidade, que fosse atualizado e direcionado para as exigências do mercado de trabalho, condição essa que nem sempre é verdadeira (BUENO, 2019; CHIRINÉA e BARREIRO, 2009). O sistema educacional brasileiro possui demandas e características distintas, sendo que as dificuldades apresentadas são as mais diversificadas possíveis, como exemplo, dificuldades em infraestrutura (GARCIA et al., 2014; VASCONCELOS et al., 2020), professores qualificados e motivados (PATTI et al., 2017), ambientes adequados (AMORIM et al., 2018), sistemas de ensino atualizados, (GATTI, 2014), além da capacitação dos gestores educacionais (RESCIA e GENTILINI, 2016).

Ainda sobre a formação dos indivíduos, a qualidade do acesso ao conhecimento no ambiente escolar se torna importante, por que os sistemas de produção e serviços estão evoluindo e se adaptando a sistemas inteligentes de produtividade, e por conta disso, as escolas precisam também de adaptação nos currículos escolares, de forma a contemplar os conteúdos primordiais com vistas as necessidades do mercado de trabalho (BUENO, 2019).

Pensando em qualidade de ensino, o Governo Federal por meio do seu Ministério da Educação e Cultura (MEC), desde 1995, vem constantemente adaptando os seus sistemas de Avaliações de Larga Escala com intuito de prover

informações suficientes para elaboração de planos que proporcionem melhorias em todo o sistema de ensino brasileiro.

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), que é o responsável pelo desenvolvimento e aplicação das Avaliações em Larga Escala, disponibiliza no seu website os resultados de cada edição (INEP, 2020a). As avaliações que compreendem o Sistema de Avaliação da Educação Básica (SAEB) são direcionadas para os alunos do 5º ano do Ensino Fundamental 1, do 9º ano do Ensino fundamental 2 e do Ensino Médio (INEP, 2020a).

Além das avaliações do SAEB, direcionadas para o Ensino Fundamental e Médio, existem outras, também em larga escala, que são direcionadas para alunos egressos do Ensino Médio ou de cursos de graduação, exemplos disso são o Exame Nacional do Ensino Médio (ENEM) que é direcionado aos egressos do Ensino Médio e o ENADE que avalia os cursos de graduação em território nacional (INEP, 2020b). Internacionalmente há a avaliação do PISA, direcionado a alunos de faixa etária dos 15 anos, é coordenado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) e envolveu em sua última edição, no ano de 2018, 80 países (OCDE, 2020). Importante salientar que independente do modelo de avaliação em larga escala, todas têm um ponto em comum, que é fornecer informações para o acompanhamento do desempenho educacional dos alunos, e essa informação se torna importante por que contribui para a análise do nível de qualidade da educação do país.

A qualidade da educação pode contribuir para a formação do indivíduo, e uma das alternativas para verificar o nível de conhecimento dos alunos, são os resultados das Avaliações em Larga Escala, que no Brasil consistem nas avaliações do sistema SAEB (AMÂNCIO-VIEIRA et al., 2015; AMÉRICO e LACRUZ, 2017; FONSECA e NAMEN, 2016; LOCKMANN e MACHADO, 2018).

Se por um lado o governo brasileiro procura acompanhar a evolução da qualidade de ensino por meio das Avaliações em Larga Escala, por outro, o mercado de trabalho pressiona o jovem egresso do Ensino Médio para que esse seja mais qualificado e preparado. E é nesse ponto que as análises dos resultados nas avaliações em larga escala, em conjunto com os curriculums escolares e a necessidades impostas pelo mercado de trabalho, podem fornecer informações importantes para o processo de tomada de decisão, inclusive no ambiente da educação.

Dessa forma, esse estudo se propõe a analisar os resultados das avaliações do SAEB aplicadas aos egressos do Ensino Médio, pois se torna importante observar o nível de desempenho desses jovens à medida que seguirão para exercer atividades no mercado de trabalho e ainda continuarão com o ciclo de formação profissional por meio de cursos de graduação.

A avaliação do SAEB destinada aos alunos egressos do Ensino Médio é composta por duas proficiências, sendo Língua Portuguesa e Matemática. Os resultados obtidos pelos alunos nessas avaliações, são apresentados em uma escala que pode variar entre 0 e 500 pontos, com isso, presume-se que quanto maior a nota obtida pelos alunos melhor foi o desempenho (INEP, 2020a). Os resultados das edições do SAEB são divulgados publicamente no website do INEP e podem ser utilizadas para estudos e análises pelos governos estaduais, pelos profissionais da educação e ainda pela sociedade em geral.

Além da divulgação dos resultados individuais dos alunos, o governo calcula a partir desses resultados, as médias estaduais, municipais e individualmente para cada escola, para as duas proficiências avaliadas. Essas informações podem ser utilizadas, inicialmente, para que os governos estaduais, escolas e profissionais da educação possam analisar qual o nível de desempenho dos alunos da sua região, e ainda, se esse nível evoluiu em relação a edição anterior do SAEB. Além disso, os gestores escolares e professores, podem realizar adaptações nas metodologias e práticas de ensino, com intuito de proporcionar melhorias no ambiente educacional, e com isso favorecer as condições de aprendizagem dos alunos.

Diante de um cenário que fica evidente a necessidade da análise da qualidade da educação, seja por necessidade do governo, das escolas ou dos indivíduos, se torna importante então, observar como os desempenhos escolares dos alunos estão evoluindo. Diante disso o objetivo desse estudo está separado em duas etapas, a primeira é analisar a evolução dos desempenhos médios estaduais e regionais referentes as edições do SAEB dos anos de 2005 a 2017. Como segunda etapa é, analisar os desempenhos individuais dos alunos nas edições de 2005 e 2017, e observar o nível de evolução entre elas. Ressaltando que o foco do estudo está nos resultados dos egressos do Ensino Médio nas proficiências Língua Portuguesa e Matemática.

Esse estudo se torna importante à medida que pode fornecer informações essenciais sobre o nível de desempenho e evolução dos alunos nas avaliações de

larga escala, e com isso, dar subsídios para o desenvolvimento de melhorias em todo ambiente escolar.

Esse artigo está estruturado em 6 capítulos, sendo o primeiro a introdução contemplando o contexto da pesquisa e os objetivos, logo após e na sequência a Revisão da Literatura, Método, Resultados, Considerações Finais e Referências.

2 Revisão da Literatura

A revisão da literatura se torna primordial para o entendimento de como o sistema educacional brasileiro é estruturado e como a proposta do acompanhamento da qualidade da educação está sendo analisada.

2.1 PNE (Plano Nacional de Educação)

O Plano Nacional de Educação (PNE) tem como objetivo principal determinar as diretrizes, metas e estratégias relacionadas à educação. Normalmente o plano tem uma duração definida e o que está em vigência no Brasil compreende o período de 2014 a 2024 (CONAE, 2010; MEC, 2014). No Brasil o PNE passou a ser uma exigência constitucional por meio da Lei nº 13005/14.

Dentre as 20 metas propostas no PNE, uma serviu como base para a análise do objetivo desse estudo, que é a meta nº 7 que tem em seu texto, (MEC, 2014):

Fomentar a qualidade da educação básica em todas as etapas e modalidades, com melhoria do fluxo escolar e da aprendizagem, de modo a atingir as seguintes médias nacionais para o Ideb: 6,0 nos anos iniciais do ensino fundamental; 5,5 nos anos finais do ensino fundamental; 5,2 no ensino médio.

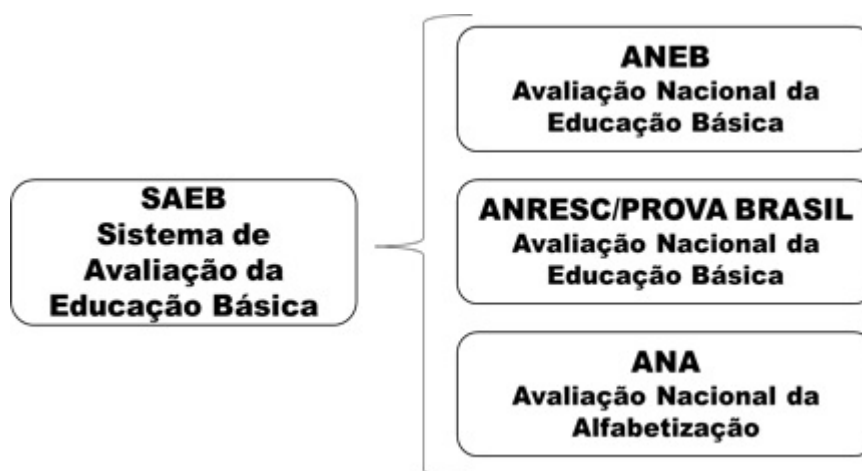
O PNE indica em seu texto que a qualidade da educação no Brasil deve ser assegurada pelo governo, e para acompanhar se os esforços e os investimentos públicos estão no caminho certo, foram implantadas as avaliações em larga escala que passaram a contribuir como um instrumento indicador do nível de desempenho dos alunos. De posse dessas informações as instituições responsáveis pela educação e a sociedade de um modo geral, podem pensar em adaptações nas políticas públicas para que o objetivo de qualidade, proposto no PNE, possa ser alcançado dentro do prazo estipulado.

2.2. Sistema SAEB

O Sistema de Avaliação da Educação Básica (SAEB) tem como objetivo avaliar se a educação básica brasileira está contribuindo para o desenvolvimento dos indivíduos, em termos de formação educacional.

Em 1990, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) iniciou os processos de avaliação da aprendizagem e em 2005 decidiu reestruturar o sistema de avaliação da aprendizagem, criando então a Avaliação Nacional de Educação Básica e a Avaliação Nacional do Rendimento Escolar, mais conhecida como PROVA BRASIL. Em 2013 foi incorporada ao sistema SAEB a Avaliação Nacional de Alfabetização que tem como foco observar o desempenho na alfabetização e letramento em língua portuguesa bem como em matemática (INEP, 2020a). A Figura 1 ilustra o sistema de avaliação proposta pelo SAEB.

Figura 1 – Estrutura do Sistema de Avaliação da Educação Básica (SAEB)



Fonte: Adaptado de INEP (INEP, 2020a).

O sistema SAEB tem evoluído desde a década de 1990, seja em relação ao público-alvo ou na estrutura da avaliação, incluindo a base de conhecimento para aplicação das avaliações. Para um melhor entendimento dos objetivos de cada modelo de avaliação, estão descritas no Quadro 1 as principais características de cada modelo de avaliação.

Quadro 1 – Características das Avaliação do SAEB para Ens. Fundamental e Médio.

Características do Sistema de Avaliação da Educação Brasileira.			
Público-Alvo	Abrangência	Formulação dos Itens	Disciplinas Avaliadas
5º e 9º ano do Ens. Fundamental	Escolas públicas – Censitário Escolas privadas – Amostral	Matrizes de Referência	Língua Portuguesa e Matemática
3ª e 4ª série do Ensino Médio	Escolas públicas – Censitário Escolas privadas – Amostral + Adesão	Matrizes de Referência	Língua Portuguesa e Matemática

Fonte: Adaptado com base nos dos INEP. (INEP, 2020a)

Com o objetivo de viabilizar o Indicador de Desenvolvimento da Educação Básica (IDEB), o INEP integrou os resultados da Prova Brasil e do Censo Escolar. O índice IDEB tem como um dos principais objetivos observar se o aluno “aprendeu” e “passou de ano”, pois utiliza os resultados das avaliações em larga escala como medida para o desempenho e os dados do Censo Escolar para observar as aprovações/reprovações dos alunos. A combinação dos resultados das avaliações em larga escala, Censo Escolar e IDEB foi desenvolvida como uma proposta de apresentar o nível de qualidade do sistema educacional brasileiro (INEP, 2020a).

Os resultados das avaliações são divulgados publicamente no próprio website do INEP e podem ser utilizadas para análises e pesquisas acadêmicas, entre outros estudos. Para compor esse estudo serão utilizados os resultados coletados no website do INEP e o foco será nos valores de desempenho obtidos pelos alunos nas proficiências analisadas, que são Língua Portuguesa e Matemática.

2.3 IDEB (Índice de Desenvolvimento da Educação Básica)

Criado em 2007 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) o Índice de Desenvolvimento da Educação Básica (IDEB) tem por objetivo observar e medir a qualidade da educação brasileira. É reconhecido como indicador que possibilita observar como a educação está evoluindo em termos de qualidade, principalmente com foco nas metas propostas no Plano Nacional de Educação (INEP, 2020c).

O IDEB é o resultado dos cálculos envolvendo dois itens oriundos do sistema educacional brasileiro, um deles é a taxa de rendimento escolar (taxa de aprovação) e o outro são as médias de desempenho nas avaliações SAEB, coordenadas pelo

INEP. Os índices de aprovação são coletados no Censo Escolar (realizado anualmente), e o desempenho das avaliações SAEB, são coletados no portal INEP.

Todos estados, municípios e escolas da rede de ensino têm a sua própria meta bienal, no entanto o objetivo é que todas alcancem a pontuação 6 até 2022. O Quadro 2 apresenta a evolução do IDEB e as metas até 2021.

Quadro 2 – Evolução e Metas do Índice de Desenvolvimento da educação Básica (IDEB)

Anos Iniciais do Ensino Fundamental																
IDEB Observado									Metas							
Total	2005	2007	2009	2011	2013	2015	2017	2019	2007	2009	2011	2013	2015	2017	2019	2021
	3.8	4.2	4.6	5.0	5.2	5.5	5.8	5.9	3.9	4.2	4.6	4.9	5.2	5.5	5.7	6.0
Anos Finais do Ensino Fundamental																
IDEB Observado									Metas							
Total	2005	2007	2009	2011	2013	2015	2017	2019	2007	2009	2011	2013	2015	2017	2019	2021
	3.5	3.8	4.0	4.1	4.2	4.5	4.7	4.9	3.5	3.7	3.9	4.4	4.7	5.0	5.2	5.5
Ensino Médio																
IDEB Observado									Metas							
Total	2005	2007	2009	2011	2013	2015	2017	2019	2007	2009	2011	2013	2015	2017	2019	2021
	3.4	3.5	3.6	3.7	3.7	3.7	3.8	4.2	3.4	3.5	3.7	3.9	4.3	4.7	5.0	5.2

Obs: Valores com fundo cinza escuro = Alcance da meta na edição

Fonte: Adaptado pelos Autores com base nos dados do INEP (INEP, 2020k).

O que se pode observar, no Quadro 2, é que só a meta para os Anos Iniciais do Ensino Fundamental tem sido alcançada desde 2007 o que não acontece com as metas para os Anos finais do Ensino Fundamental e para o Ensino Médio. Com base nessa informação se pode presumir que existe necessidade de maiores estudos, envolvendo os segmentos do governo e da sociedade, em prol do desenvolvimento de planos de ação ou ainda de políticas públicas que possam propor condições de melhorias em todo o sistema de ensino brasileiro.

2.4. Publicações sobre o tema.

Vários estudos publicados observaram os desempenhos nas Avaliações em Larga Escala como forma de entender ou acompanhar a evolução da qualidade do sistema de ensino brasileiro. Como exemplo, o de Correia (2017) que teve como objetivo analisar as avaliações em larga e a relações com o desempenho dos alunos em escolas públicas do estado de São Luís, no Maranhão, apresentou em suas

conclusões que o sistema SAEB e IDEB são referências válidas para medir o nível de qualidade da educação no país, e complementando, o autor indica que em se tratando de melhorias, devem ser observadas as questões de estrutura física das escolas, formação dos professores e inclusive, adaptações nos currículos escolares.

Já o estudo de Bruns e Nunes (2018) indicaram que os resultados das avaliações em larga escala do SAEB podem influenciar diretamente no Índice de Desenvolvimento da Educação Básica-IDEB de cada escola, demonstrando com isso o nível de qualidade da educação dessa instituição de ensino. No entanto, a classificação da escola com base apenas no índice IDEB pode não representar uma realidade referente ao nível de ensino e de desenvolvimento educacional dos alunos.

Observando os desempenhos dos alunos do ensino fundamental nas avaliações do SAEB no período de 2007 a 2017, Santos (2020) concluiu que os desempenhos desses alunos estão evoluindo a cada edição do SAEB, mas que é preciso propor atualizações nos currículos escolares e capacitar os professores e gestores em relação as matrizes de referências das proficiências analisadas nas avaliações, pois com isso, pode se atualizar o conteúdo programático das disciplinas inclusive propor novas metodologias de aprendizado. De acordo com os resultados encontrados e amparado pela análise do autor, pode se observar que medidas precisam ser implementadas para que os alunos possam se desenvolver melhor em termos de conhecimentos, e com isso alcançar melhores resultados nas avaliações do SAEB.

As avaliações do SAEB podem fornecer informações importantes para o desenvolvimento dos currículos escolares e das metodologias de ensino, além de contribuir para a capacitação do corpo docente, pois a partir do estudo e entendimento do processo avaliativo do SAEB, ações podem ser desenvolvidas para que os alunos se desenvolvam melhor e aprendam mais.

Outro ponto a ser observado é se o nível de investimentos pode influenciar o desempenho dos alunos nas avaliações do SAEB e conseqüentemente no índice IDEB. O estudo de Soares e Rosa (2020) procurou investigar essa questão tendo como base 11 escolas municipais de Santa Maria no Rio Grande do Sul, com características bem semelhantes pois apresentavam custo baixo por aluno, menores taxas de reprovação e corpo docente mais qualificado. Por meio de uma metodologia qualitativa os autores concluíram que das 11 escolas analisadas

apenas 4 superaram suas metas no IDEB, dessa forma os autores concluíram que o volume de investimentos pode não ter relação direta com o desempenho escolar, sugerindo que os melhores resultados alcançados podem ter sido por conta da prática de gestão escolar. Em um estudo anterior Oliveira Júnior et al. (2019) também concluíram que o volume de investimento não possuía relação direta com o desempenho escolar.

Com um olhar para as questões de políticas públicas, Gusmão e Amorim (2020) investigaram os marcos legais que regulamentam o Ensino Médio e resultados de pesquisas oriundas das avaliações em larga escala e puderam concluir que o SAEB pode contribuir para a melhoria da qualidade da educação escolar. Os resultados dos desempenhos dos alunos nas avaliações do SAEB podem contribuir para que o governo e a sociedade acompanhem como uma determinada região ou escola está se desenvolvendo em relação à qualidade de ensino.

Vários outros estudos foram analisados nessa fase da revisão da literatura, e os resultados contribuirão para uma melhor análise e fundamentação desse estudo.

3. Método

Para conduzir esse estudo foi escolhido o método misto de pesquisa composto de duas etapas, uma quantitativa e outra qualitativa. Para a etapa quantitativa foi utilizada as técnicas da Estatística Descritiva, para a construção dos agrupamentos de dados e cálculos das médias, modas e medianas. A etapa qualitativa envolveu as análises dos resultados encontrados, na etapa quantitativa, com base na literatura investigada.

Após a pesquisa bibliográfica realizada na fase da revisão da literatura, iniciou-se a fase de coleta de dados. Os dados coletados são referentes aos resultados das avaliações SAEB dos anos de 2005 a 2017, dos egressos do Ensino Médio e que estão disponíveis para consulta pública no website do INEP. Além disso, o governo também disponibiliza as médias estaduais e regionais para as proficiências Língua Portuguesa e Matemática que são calculadas com base no conjunto de resultados dos alunos para o estado ou região, com intuito de observar o quanto cada estado ou região evolui em termos de desempenho médio, em cada proficiência (INEP, 2020a). Com base nesses dados foram criadas as tabelas que

ilustram os desempenhos médios estaduais e regionais, para as proficiências Língua Portuguesa e Matemática.

As médias atribuídas aos estados, e que são disponibilizadas pelo governo, representam o desempenho médio dos alunos de todas escolas participantes do SAEB, sendo que essas escolas podem ser de áreas urbanas ou rurais, municipais, estaduais ou federais e ainda pertencerem à rede pública de ensino ou ainda da rede privada. Ressaltando que essas médias são disponibilizadas já separadas por proficiência, ou seja, uma média para Língua Portuguesa e outra para Matemática.

Após realizada essa fase, foram separados os dados dos desempenhos dos alunos relativos as edições de 2005 e 2017, para as duas proficiências, com a intenção de analisar se houve evolução, em termos de desempenho escolar, dos egressos do Ensino Médio nesse período. Após separados os resultados dessas edições, foi realizado o processo de tratamento desses dados, que permitiu a exclusão de dados em branco, nulos, sem padrão exato e aqueles que por falta de alguma identificação também não puderam ser analisados. Após essa fase chegou-se ao total de 22.204 avaliações em Língua Portuguesa e de 22.187 avaliações em Matemática na edição de 2005, e na edição de 2017, chegou-se ao total de 1.339.051 avaliações registradas no banco de dados, tanto para Língua Portuguesa quanto para Matemática. Importante salientar que após realizada fase de tratamento dos dados, a edição 2005, apresentou uma diferença de 17 avaliações a mais em Língua Portuguesa em relação ao total de Matemática, ocasionada pela retirada de dados em branco, nulos ou preenchidos erroneamente.

Em conjunto com os resultados das avaliações o INEP disponibiliza um dicionário de dados que auxilia no tratamento e análise das informações dos desempenhos dos alunos, incluindo as escalas de proficiências, tanto para Língua Portuguesa quanto para Matemática (INEP, 2020e). As escalas são compostas por valores entre 0 e 500 pontos e são divididas em até 10 níveis. Cada nível abrange habilidades mínimas necessárias que o aluno deve dominar em termos de conhecimento. Para a proficiência Língua Portuguesa a escala varia entre o Nível 1 e 8, já para Matemática a escala varia entre o Nível 1 e 10.

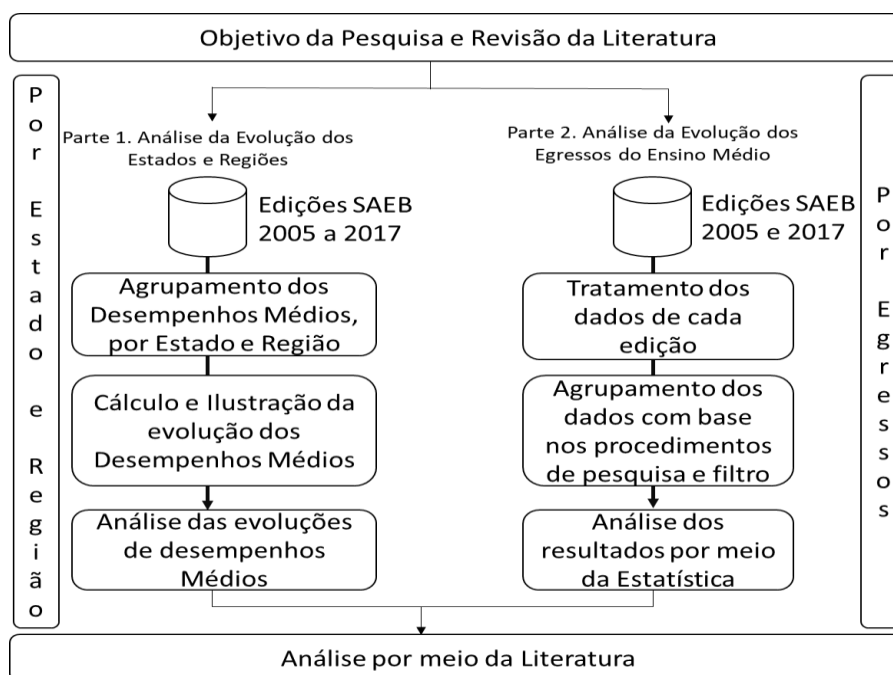
O Nível 1 da escala abrange o intervalo entre maior ou igual a 225 pontos e menor que 250. A escala vai evoluindo de nível a cada 25 pontos até chegar a sua totalidade que são 500 pontos. Para o INEP, os alunos que apresentarem pontuação inferior a 225 pontos merecem uma atenção especial pois não conseguiram absorver

o mínimo necessário para o seu nível de formação, sendo um indicativo que merece atenção no agrupamento e na análise de dados, neste caso são classificados como Nível Zero (INEP, 2020a).

Com os dados dos desempenhos dos alunos já tratados e separados por edição, o próximo passo foi agrupar os resultados por nível da escala de desempenho. Para isso foram realizados os procedimentos de pesquisa nos dados, utilizando como filtro o desempenho na proficiência e gênero do aluno. Como exemplo de filtro, foi contar quantos alunos não obtiveram nota superior a 225 pontos em Matemática e Língua Portuguesa, separados por gênero (feminino/masculino). O objetivo dessa escolha é tentar identificar se existe diferença de nível de aprendizado, nas proficiências analisadas, entre meninos e meninas, inclusive observar também, se houve evolução de desempenho dos egressos do Ensino Médio da edição de 2005 para a edição de 2017.

Em síntese, esse estudo possui duas vertentes para análise, uma envolvendo os desempenhos médio dos estados e regiões brasileiras no período de 2005 a 2017 e outra envolvendo os desempenhos dos egressos nas edições de 2005 e 2017, das avaliações do SAEB para o Ensino Médio. Para uma melhor ilustração do desenvolvimento dessa pesquisa, a Figura 2 ilustra os passos necessários para que fosse alcançado os objetivos propostos.

Figura 2. Desenho da Pesquisa.



Fonte: Desenvolvido pelos Autores (2020).

Após realizadas todas as tarefas metodológicas necessárias para o alcance dos objetivos dessa pesquisa, o próximo passo será a descrição e análise dos resultados encontrados.

4. Resultados

Após finalizada a fase metodológica que envolveu a coleta, tratamento e agrupamento dos dados, os resultados encontrados serão analisados em duas etapas, a primeira contemplando as evoluções dos desempenhos médios estaduais e regionais para as duas proficiências, e na segunda, os desempenhos dos egressos do Ensino Médio nas edições de 2005 e 2017, também para as duas proficiências. Para ilustrar os desempenhos médios estaduais e regionais foram criadas duas tabelas, sendo a Tabela 1 para os resultados da proficiência Língua Portuguesa e a Tabela 2, para Matemática.

Tabela 1. Médias das Proficiências em Língua Portuguesa e sua variação no período 2005-2017, para alunos egressos do Ensino Médio.

Estados	Língua Portuguesa							
	2005	2007	2009	2011	2013	2015	2017	Dif. 2017-2005
BRASIL	257,60	261,39	268,83	268,57	264,06	267,87	268,52	10,92
NORTE	241,99	245,29	260,45	255,33	248,46	259,35	252,28	10,29
Rondônia	257,67	257,77	271,11	268,76	263,05	268,09	268,33	10,66
Acre	252,58	259,59	264,29	255,59	261,52	263,12	264,45	11,87
Amazonas	231,27	238,21	250,57	257,87	246,16	264,76	254,46	23,19
Roraima	256,87	251,33	260,81	263,38	259,06	263,27	255,32	-1,55
Pará	243,17	245,41	257,25	250,50	244,90	255,72	245,78	2,61
Amapá	253,16	249,01	254,93	259,67	252,42	257,52	253,52	0,36
Tocantins	240,45	244,08	249,92	259,09	249,31	254,06	261,26	20,81
NORDESTE	246,22	248,49	258,36	255,23	249,34	256,20	258,59	12,37
Maranhão	231,66	241,24	247,33	248,76	241,78	248,49	254,07	22,41
Piauí	250,38	242,97	250,91	254,47	250,72	256,36	257,94	7,56
Ceará	257,79	256,79	263,06	262,49	255,94	257,03	266,28	8,49
Rio Grande do Norte	241,67	249,68	256,92	254,88	245,61	252,37	253,77	12,10
Paraíba	239,52	250,28	260,82	257,88	253,57	257,75	259,16	19,64
Pernambuco	250,17	246,49	255,27	254,35	262,09	270,39	269,36	19,19
Alagoas	243,28	243,72	251,16	247,65	241,55	250,81	256,60	13,32
Sergipe	260,18	242,09	261,85	266,76	252,84	258,42	266,41	6,23
Bahia	243,14	250,13	264,80	254,92	239,84	250,83	249,57	6,43
SUDESTE	262,70	268,94	273,30	277,35	274,07	273,86	274,83	12,13
Minas Gerais	268,71	276,71	274,78	276,33	268,82	269,29	275,86	7,15

Espírito Santo	267,72	261,67	278,39	265,75	265,90	278,00	283,72	16,00
Rio de Janeiro	257,03	260,51	268,33	275,05	277,62	276,70	274,48	17,45
São Paulo	261,57	268,77	273,68	279,09	275,83	274,66	274,05	12,48
SUL	272,37	274,51	282,40	276,77	270,56	274,46	276,42	4,05
Paraná	267,16	271,94	280,92	272,44	263,32	274,29	274,82	7,66
Santa Catarina	266,99	268,90	275,17	279,82	273,08	276,95	275,22	8,23
Rio Grande do Sul	282,47	281,49	289,57	280,42	278,67	272,79	279,24	-3,23
CENTRO-OESTE	261,56	262,63	269,25	272,79	268,13	273,19	273,68	12,12
Mato Grosso do Sul	272,56	274,43	285,47	284,62	274,51	280,28	273,60	1,04
Mato Grosso	255,73	253,80	260,38	264,00	256,13	265,05	263,20	7,47
Goiás	253,08	251,95	265,59	269,19	267,28	269,88	276,26	23,18
Distrito Federal	278,22	288,41	278,85	283,06	279,21	284,42	278,44	0,22

Fonte: Elaborada com base nos dados do INEP (INEP, 2020c).

Em relação às médias estaduais, tanto na Tabela 1 quanto na Tabela 2 os valores estão descritos conforme o ano da edição da avaliação em larga escala do SAEB. Foi criada uma coluna denominada “DIF2017-2005” que contempla a diferença entre o valor apresentado em 2017 e o valor apresentado em 2005 (2017-2005). Por meio dos resultados dessa coluna pode se observar qual foi o nível de evolução dos desempenhos médios entre a edição de 2005 e 2017, e se essa evolução foi positiva ou negativa.

Apenas para ressaltar, cada nível da escala, a partir do Nível 1, evolui em 25 pontos, dessa forma o Nível 1 da escala compreende de 225 a 250, Nível 2 de 250 a 275 e assim sucessivamente. Para a proficiência Língua Portuguesa, o Nível 8 da escala, compreende valores iguais ou superiores a 400 pontos, dessa forma não há na escala de desempenho, a indicação dos níveis 9 e 10. Além disso, para o governo, alunos com pontuação menor que 225 pontos, possuem classificação compatíveis com o Nível Zero, e por conta disso devem receber uma atenção especial, pois estão abaixo do mínimo exigido para a sua formação.

Para que se possa observar quais são as competências mínimas exigidas para cada nível da escala, o Quadro 3 ilustra os conteúdos referentes aos níveis de 1 a 4, para a proficiência Língua Portuguesa. Serão reproduzidos os conteúdos referentes aos níveis 1 a 4, apenas por que os maiores níveis de desempenho apresentados pelos estados estão nessa faixa, mas o documento oficial completo pode ser acessado diretamente no pacote de dados disponibilizado pelo INEP (INEP, 2020a).

Quadro 3. Relação de competências em Língua Portuguesa, para os níveis de 1 a 4 na escala de desempenho.

<p>Nível 1 Desempenho maior ou igual a 225 e menor que 250</p>	<p>Os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Identificar elementos da narrativa em história em quadrinhos. • Reconhecer a finalidade de recurso gráfico em artigos. • Reconhecer a relação de causa e consequência em lendas. • Inferir o sentido de palavra em letras de música e reportagens.
<p>Nível 2 Desempenho maior ou igual a 250 e menor que 275</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Reconhecer a ideia comum entre textos de gêneros diferentes e a ironia em tirinhas. • Reconhecer relações de sentido estabelecidas por conjunções ou locuções conjuntivas em letras de música e crônicas. • Reconhecer o uso de expressões características da linguagem (científica, profissional etc.) e a relação entre pronome e seu referente em artigos e reportagens. • Inferir o efeito de sentido da linguagem verbal e não verbal em notícias e charges.
<p>Nível 3 Desempenho maior ou igual a 275 e menor que 300</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Localizar informação explícita em artigos de opinião. • Identificar a finalidade de relatórios científicos. • Reconhecer relações de sentido marcadas por conjunções, a relação de causa e consequência e a relação entre o pronome e seu referente em fragmentos de romances. • Reconhecer o tema de uma crônica. • Reconhecer variantes linguísticas em artigos. • Reconhecer o sentido e o efeito de sentido produzido pelo uso de recursos morfossintáticos em contos, artigos e crônicas. • Reconhecer opiniões divergentes sobre o mesmo tema em diferentes textos. • Inferir informação, o sentido e o efeito de sentido produzido por expressão em reportagens e tirinhas.
<p>Nível 4 Desempenho maior ou igual a 300 e menor que 325</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Localizar informações explícitas em infográficos, reportagens, crônicas e artigos. • Identificar o argumento em contos. • Identificar a finalidade e a informação principal em notícias. • Reconhecer a relação entre os pronomes e seus referentes em contos. • Reconhecer elementos da narrativa em contos. • Reconhecer variantes linguísticas em contos, notícias e reportagens. • Reconhecer o efeito de sentido produzido pelo uso de recursos morfossintáticos em poemas. • Reconhecer ideia comum e opiniões divergentes sobre o mesmo tema na comparação entre diferentes textos. • Reconhecer ironia e efeito de humor em crônicas e entrevistas. • Reconhecer a relação de causa e consequência em piadas e fragmentos de romance. • Comparar poemas que abordem o mesmo tema. • Diferenciar fato de opinião em contos, artigos e reportagens. • Diferenciar tese de argumentos em artigos, entrevistas e crônicas. • Inferir informação, sentido de expressão e o efeito de sentido decorrente do uso de recursos morfossintáticos em crônicas. • Inferir o sentido decorrente do uso de recursos gráficos em poemas. • Inferir o efeito de sentido da linguagem verbal e não verbal e o efeito de humor em tirinhas.

Fonte: Adaptado de INEP (INEP, 2020c).

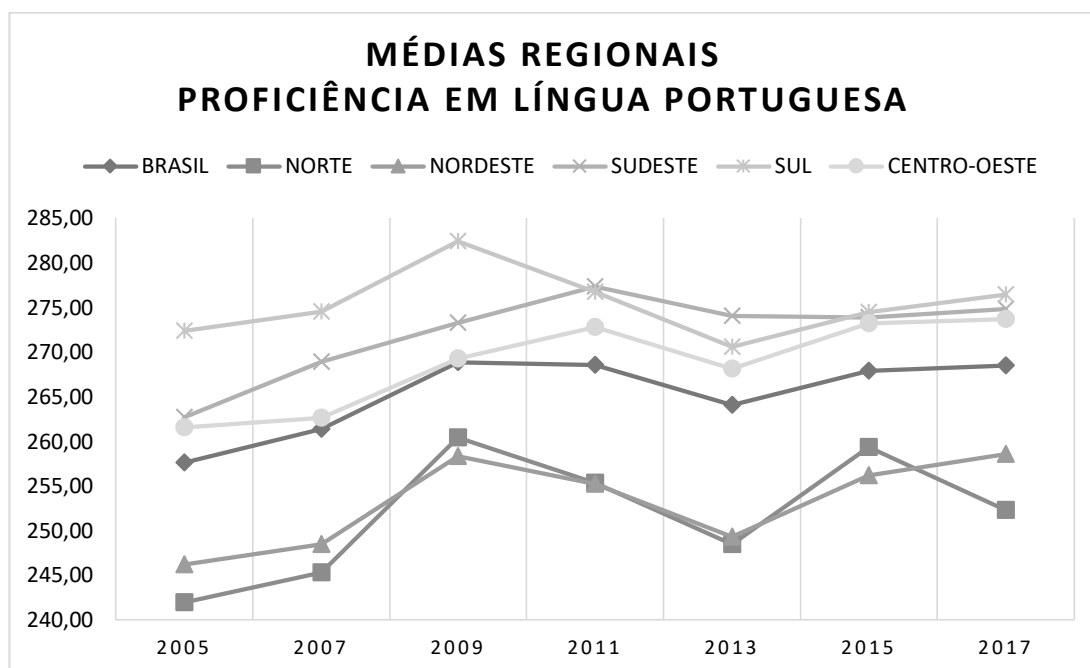
Iniciando as análises dos dados da Tabela 1, na edição de 2017, as médias das regiões para a proficiência Língua Portuguesa são compatíveis com o Nível 2 da escala (≥ 250 e < 275 pontos), com exceção da Região Sul, que apresentou média

de 276,42 pontos, compatível com o Nível 3 da escala de desempenho. E se observado por estado, os valores acima de 275 pontos, compatíveis com o Nível 3, foram alcançadas por Minas Gerais (275,86), Espírito Santo (283,72), Santa Catarina (275,22), Rio Grande do Sul (279,24), Goiás (276,26) e Distrito Federal (278,44). Abaixo de 250 pontos, ainda no Nível 1, estão os estados do Pará (245,78) e Bahia (249,57).

Independentemente do nível atual de cada estado e região, as médias alcançadas estão muito longe de um valor ideal, pois os estados com maior pontuação ainda estão no Nível 3. Dessa forma, cabe aos membros do corpo docente e gestores escolares, analisarem os conteúdos referentes a cada nível da escala e elaborar planos de ação que possam identificar as dificuldades de aprendizado dos alunos. Com base nesses planos de ação e em conjunto com as famílias dos alunos, é possível desenvolver estratégias de ensino que contribuam para a melhoria do aprendizado, pois são diversos os fatores que podem influenciar no nível de aprendizagem, desde a condição socioeconômica da família, a falta de motivação para os estudos, formação e expectativa do corpo docente e ainda experiência dos gestores escolares (BASSETTO, 2019; FONSECA e NAMEN, 2016; MARANGONI, ARAÚJO e SCHWARZ, 2019; SILVA e LEAL, 2020; TABERNEIRO, SERRANO e MÉRIDA, 2017)

Em relação a evolução no período de 2005-2017, observa-se que a média nacional evoluiu positivamente em 10,92 pontos. Já os estados de Roraima (-1,55) e Rio Grande do Sul (-3,23) apresentaram evoluções negativas. Em termos gerais a maior evolução positiva no período de 2005 a 2017 foi a do estado Amazonas (23,19) e a maior evolução negativa foi a do estado do Rio Grande do Sul (-3,23). A Figura 3 ilustra graficamente a evolução de cada região para a proficiência Língua Portuguesa no período de 2005 a 2017.

Figura 3. Evolução na Proficiência Língua Portuguesa no período 2005-2017



Fonte: Elaborado pelos Autores (2020).

Importante observar que somente investir em educação não é sinônimo de sucesso nas avaliações em larga escala, pois o nível de investimento deve estar acompanhado de mudanças de paradigmas, melhores práticas e metodologias de ensino, sistema de gerenciamento eficaz tanto dos recursos quanto dos ambientes escolares, e essas ações só podem ser efetivas com planejamento, organização e controle (SAVIAN e BEZERRA, 2013; TARDA e RODRIGUES, 2015; TERNOSKI et al., 2017).

A próxima análise envolve os desempenhos médios referentes a proficiência Matemática, para isso a Tabela 2 ilustra os dados para cada estado e região, inclusive o nível de evolução no período 2005 a 2017.

Tabela 2. Médias das Proficiências em Matemática e sua variação no período 2005-2017, para alunos egressos do Ensino Médio.

Estados	Matemática							Dif. 2017-2005
	2005	2007	2009	2011	2013	2015	2017	
BRASIL	271,29	272,89	274,72	274,83	270,15	267,60	270,63	-0,66
NORTE	250,25	253,24	261,53	255,36	250,90	257,03	252,04	1,79
Rondônia	271,13	271,16	275,46	276,37	271,15	265,84	272,03	0,90
Acre	257,10	264,85	261,53	257,76	259,25	257,96	263,28	6,18

Amazonas	241,99	247,43	251,40	257,32	247,86	259,66	250,93	8,94
Roraima	268,49	259,60	263,87	268,67	260,00	261,68	256,73	-11,76
Pará	248,70	252,64	256,76	248,73	247,08	254,49	246,29	-2,41
Amapá	260,28	252,21	254,61	254,41	251,74	252,75	249,11	-11,17
Tocantins	253,56	252,07	251,98	263,88	254,99	256,30	265,77	12,21
NORDESTE	258,16	258,15	261,30	257,85	254,81	256,92	259,95	1,79
Maranhão	239,90	246,19	247,00	247,06	244,04	249,11	251,68	11,78
Piauí	260,49	253,78	256,11	259,26	258,13	256,91	263,14	2,65
Ceará	265,95	265,31	267,31	265,82	260,95	261,07	267,37	1,42
Rio Grande do Norte	254,90	257,90	259,67	258,82	250,46	255,15	255,01	0,11
Paraíba	253,58	261,03	264,16	258,94	258,09	258,02	261,27	7,69
Pernambuco	255,42	257,75	258,81	258,10	266,92	266,99	271,39	15,97
Alagoas	258,99	251,76	254,81	248,73	248,17	252,76	257,93	-1,06
Sergipe	271,20	253,29	266,21	260,37	260,18	258,98	269,93	-1,27
Bahia	262,05	261,30	266,89	257,64	246,99	251,57	251,31	-10,74
SUDESTE	276,74	280,42	280,23	285,22	279,49	273,46	276,81	0,07
Minas Gerais	290,38	289,35	285,71	289,06	279,75	272,67	281,48	-8,90
Espírito Santo	282,39	277,45	286,96	281,97	279,41	281,43	291,64	9,25
Rio de Janeiro	268,91	271,87	277,01	286,84	283,47	274,34	277,09	8,18
São Paulo	272,92	279,43	278,10	283,41	278,36	273,09	274,12	1,20
SUL	292,52	291,89	294,47	290,14	282,16	274,69	281,55	-10,97
Paraná	284,77	290,15	290,77	283,77	273,93	273,77	279,49	-5,28
Santa Catarina	286,55	287,77	286,99	294,97	286,46	278,07	283,69	-2,86
Rio Grande do Sul	306,12	296,83	304,81	295,26	290,27	273,31	283,41	-22,71
CENTRO-OESTE	274,53	275,39	274,62	279,53	275,26	271,91	277,42	2,89
Mato Grosso do Sul	283,13	288,58	291,40	291,87	280,42	277,07	276,40	-6,73
Mato Grosso	263,56	264,57	263,69	269,58	262,69	263,83	265,93	2,37
Goiás	266,76	265,60	271,39	276,17	274,63	270,28	278,91	12,15
Distrito Federal	297,83	300,31	285,65	290,16	287,49	280,42	286,41	-11,42

Fonte: Elaborada com base nos dados do INEP (INEP, 2020c).

Para a proficiência Matemática, conforme Tabela 2, as médias em 2017 compatíveis com o Nível 1 (≥ 225 < 250) da escala foram as dos estados do Pará (246,29) e do Amapá (249,11). Compatíveis com o Nível 3 (≥ 275 < 300) foram as dos estados de Minas Gerais (281,48), do Espírito Santo (291,64), Rio de Janeiro (277,09), Paraná (279,49), Santa Catarina (283,69), Rio Grande do Sul (283,41), Mato Grosso do Sul (276,40), Goiás (278,91) e Distrito Federal (268,41). Os demais estados apresentaram médias compatíveis com o Nível 2 da escala (≥ 250 < 275).

De forma similar aos resultados para Língua Portuguesa, os valores de desempenhos médios apresentados pelos estados indicam que a maioria deles também estão compatíveis com o Nível 2 da escala de desempenho. Para se ter

uma noção dos conteúdos referentes aos níveis de 1 a 4, o Quadro 4 ilustra esses tópicos.

Quadro 4. Relação de competências em Matemática, para os níveis de 1 a 4 na escala de desempenho

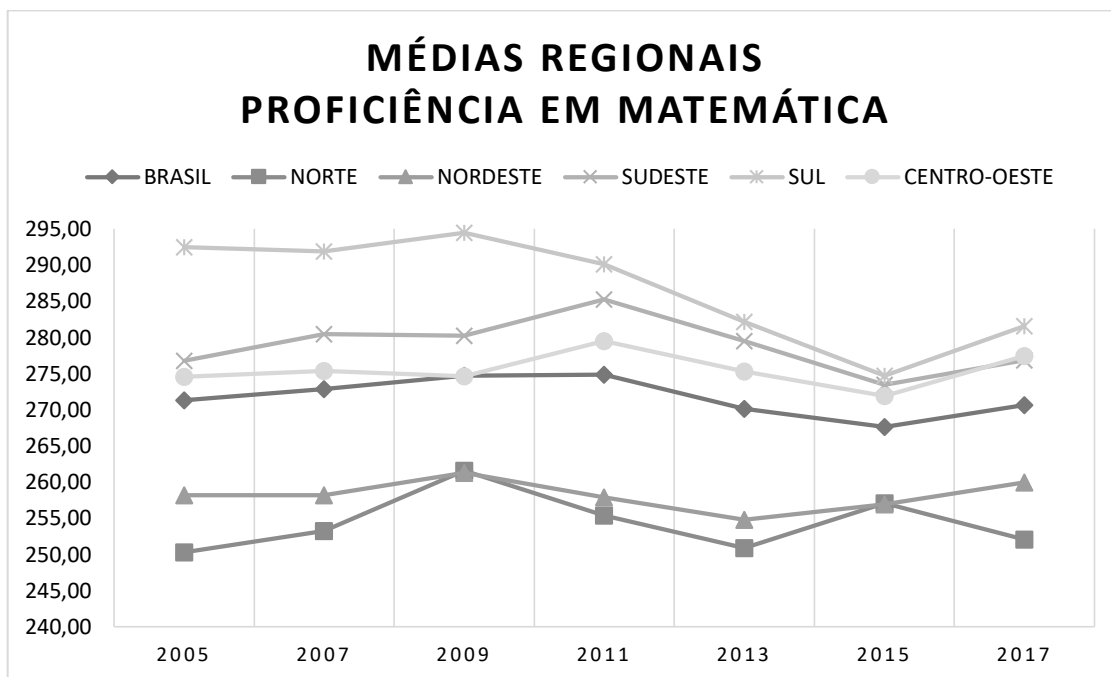
<p>Nível 1 Desempenho maior ou igual a 225 e menor que 250</p>	<p>Os estudantes provavelmente são capazes de: TRATAMENTO DE INFORMAÇÕES Associar uma tabela de até duas entradas a informações apresentadas textualmente ou em um gráfico de barras ou de linhas.</p>
<p>Nível 2 Desempenho maior ou igual a 250 e menor que 275</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: ESPAÇO E FORMA Reconhecer as coordenadas de pontos representados em um plano cartesiano localizados no primeiro quadrante. NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES Reconhecer os zeros de uma função dada graficamente. Determinar o valor de uma função afim, dada sua lei de formação. Determinar resultado utilizando o conceito de progressão aritmética. TRATAMENTO DE INFORMAÇÕES Associar um gráfico de setores a dados percentuais apresentados textualmente ou em uma tabela.</p>
<p>Nível 3 Desempenho maior ou igual a 275 e menor que 300</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES Reconhecer o valor máximo de uma função quadrática representada graficamente. Reconhecer, em um gráfico, o intervalo no qual a função assume valor máximo. Determinar, por meio de proporcionalidade, o gráfico de setores que representa uma situação com dados fornecidos textualmente. Determinar o quarto valor em uma relação de proporcionalidade direta a partir de três valores fornecidos em uma situação do cotidiano. Determinar um valor reajustado de uma quantia a partir de seu valor inicial e do percentual de reajuste. Resolver problemas utilizando operações fundamentais com números naturais.</p>
<p>Nível 4 Desempenho maior ou igual a 300 e menor que 325</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: GRANDEZAS E MEDIDAS Resolver problemas envolvendo área de uma região composta por retângulos a partir de medidas fornecidas em texto e figura. NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES Reconhecer o gráfico de função a partir de valores fornecidos em um texto. Determinar a lei de formação de uma função linear a partir de dados fornecidos em uma tabela. Determinar a solução de um sistema de duas equações lineares. Determinar um termo de progressão aritmética, dada sua forma geral. Determinar a probabilidade da ocorrência de um evento simples. Resolver problemas utilizando proporcionalidade direta ou inversa, cujos valores devem ser obtidos a partir de operações simples. Resolver problemas de contagem usando princípio multiplicativo.</p>

Fonte: Adaptado de INEP (INEP, 2020c).

Em relação ao nível de evolução, a média nacional para a proficiência Matemática evoluiu negativamente no período de 2005 a 2017, apresentando o valor de -0,66 pontos. Em relação às regiões, apenas a Região Sul também apresentou

evolução negativa em -10,97 pontos, as demais regiões apresentaram evoluções positivas. A maior evolução positiva foi a da Região Centro-Oeste com valor de 2,89 pontos. A Figura 4 ilustra as evoluções das regiões para a proficiência Matemática no período de 2005 a 2017.

Figura 4. Evolução na Proficiência Matemática no período 2005-2017



Fonte: Desenvolvido pelos Autores (2020).

Em relação às escalas de proficiências, para Matemática as escalas variam igualmente a de Língua Portuguesa, no entanto, para Língua Portuguesa a escala indica o Nível 8 como o último, compreendendo valores iguais ou superiores a 400 pontos. Já para Matemática, o último nível é o 10, com valores iguais ou superiores a 450 pontos.

Da mesma forma que existe a necessidade de adaptações no sistema educacional com objetivo de se melhorar o desempenho em Língua Portuguesa, para Matemática não é diferente. Se faz necessário, inclusive, um trabalho mais minucioso, pois 12 estados do total de 27 apresentaram evolução negativa no período 2005 a 2017, para essa proficiência.

Se observado a questão da qualidade de ensino, pode-se presumir que existe a necessidade de adaptações nas metodologias de ensino, em práticas relacionadas ao desenvolvimento da aprendizagem, um melhor acompanhamento por parte do

corpo docente, gestores escolares e membros das famílias, pois somente com um trabalho em conjunto, esses alunos podem se sentir amparados e motivados para o aprendizado. Não se pode isentar a participação do estado, pois para o desenvolvimento de um trabalho com qualidade, as escolas devem possuir recursos técnicos e de informática atualizados, os professores devem estar motivados, capacitados e bem remunerados, não esquecendo dos gestores escolares, que precisam ter condições para o exercício da função (BRITO, 2017; CALIXTO et al., 2017; PATTI ET AL., 2017; SILVA e LEAL, 2020; SIMIELLI, 2017; SOARES e ROSA, 2020).

O próximo passo foi analisar os desempenhos apresentados pelos alunos nas edições do SAEB de 2005 e de 2017. Após a fase de tratamento e agrupamento dos dados (já descritos na metodologia), a Tabela 3 foi desenvolvida. Nesta tabela, estão descritos a quantidade de aluno por nível de escala, para a proficiência Língua Portuguesa e Matemática, já separados por gênero e por edição.

Tabela 3. Quantidade de alunos em cada nível da escala, para proficiências Língua Portuguesa e Matemática, nas edições de 2005 e 2017.

Escalas		Ensino Médio 2005								Ensino Médio 2017							
		Feminino		Masculino		Feminino		Masculino		Feminino		Masculino		Feminino		Masculino	
		LP	%LP	LP	%LP	MT	%MT	MT	%MT	LP	%LP	LP	%LP	MT	%MT	MT	%MT
Nivel0	<225	2408	18,7	2081	22,2	2453	19,9	1318	13,4	150162	20,5	175535	28,9	174202	23,8	133558	22,0
Nivel1	>=225 e <250	1639	12,8	1118	11,9	1709	13,9	941	9,6	109909	15,0	93852	15,5	146261	20,0	95411	15,7
Nivel2	>=250 e <275	1887	14,7	1326	14,2	1814	14,7	1146	11,6	132177	18,1	100573	16,6	148927	20,4	104788	17,3
Nivel3	>=275 e <300	2008	15,6	1393	14,9	1566	12,7	1259	12,8	135254	18,5	96755	15,9	112474	15,4	94371	15,5
Nivel4	>=300 e <325	2057	16,0	1402	15,0	1401	11,4	1240	12,6	107385	14,7	74372	12,2	74889	10,2	78040	12,9
Nivel5	>=325 e <350	1617	12,6	1167	12,5	1150	9,3	1087	11,0	62897	8,6	42725	7,0	42870	5,9	54439	9,0
Nivel6	>=350 e <375	890	6,9	645	6,9	945	7,7	1033	10,5	26430	3,6	17962	3,0	19220	2,6	26881	4,4
Nivel7	>=375 e <400	299	2,3	195	2,1	726	5,9	866	8,8	6928	0,9	5006	0,8	8112	1,1	11690	1,9
Nivel8	>=400 e <425	41	0,3	31	0,3	399	3,2	587	6,0	645	0,1	484	0,1	3624	0,5	5600	0,9
Nivel9	>=425 e <450					157	1,3	322	3,3					1195	0,2	2454	0,4
Nivel10	>=450					16	0,1	52	0,5					13	0,0	32	0,0
Totais.....:		12846	100	9358	100	12336	100	9851	100	731787	100	607264	100	731787	100	607264	100

Fonte: Elaborado pelos Autores (2020).

Inicialmente pode se perceber o aumento na quantidade de alunos que realizaram as provas das avaliações do SAEB, em 2005 esse total foi de

aproximadamente 12.000 alunos já em 2017 o total de alunos chegou a quase 732.000, com isso, a variação apresentada aproximou-se de 6.100%. Essa variação pode ter sido resultado de ações do governo, que passou a controlar e exigir uma maior participação das escolas na realização das avaliações em larga escala.

Conforme dados apresentados na Tabela 3, pode-se observar que em Língua Portuguesa, na edição de 2005, 18,7% das meninas e 22,2% dos meninos não alcançaram pontuação maior que 225 pontos, ou seja compatível com o Nível Zero da escala. Já na edição de 2017, o percentual de alunos no mesmo nível da escala foi de 20,5% de meninas e de 28,9% de meninos. Percebe-se que em relação à edição de 2005, houve um aumento da quantidade de alunos no Nível Zero na edição de 2017, demonstrando que o sistema e as metodologias de ensino utilizadas não produziram efeitos positivos, pois ao invés da quantidade de alunos regredir houve na realidade um aumento de 1,8% de meninas e de 6,7% de meninos.

Para a proficiência Matemática o cenário não foi diferente, o percentual de alunos no Nível Zero foi de 19,9% de meninas e 13,4% na edição de 2005, já na edição de 2017 esses percentuais foram de 23,8% de meninas e de 22% de meninos. Observa-se que também houve um aumento da quantidade de alunos nesse nível, apresentando um acréscimo de 3,9% de meninas e de 8,6% de meninos.

Apenas ressaltando que, para o governo, os alunos com níveis de aprendizado compatíveis com o Nível Zero da escala, merecem uma atenção especial pois, não conseguiram absorver o mínimo necessário para o seu nível de formação (INEP, 2020a).

A análise por meio da Estatística Descritiva indicou que os valores da MODA, tanto na edição de 2005 quanto na 2017, e independente da proficiência, são compatíveis com o Nível Zero. E se forem observadas as MEDIANAS, na edição de 2005 em Língua Portuguesa os valores são compatíveis com o Nível 3 da escala tanto para os meninos quanto para as meninas, e para Matemática, esses valores são compatíveis com o Nível 3 para as meninas e Nível 4 para os meninos. Já na edição de 2017, os valores das MEDIANAS tanto para Língua Portuguesa quanto para Matemática, e independente do gênero do aluno, estão compatíveis com o Nível 2 da escala de desempenho.

Com isso, pode-se perceber que os valores das MODAS e das MEDIANAS demonstraram que, em 2017, houve uma redução no nível de conhecimento dos alunos se comparados com a edição de 2005.

Se os resultados dos desempenhos dos alunos foram analisados por meio dos QUARTIS, pode-se perceber que em 2005, os valores referentes ao terceiro QUARTIL, para Língua Portuguesa são compatíveis com o Nível 4 da escala, tanto para as meninas quanto para meninos, e em Matemática, os valores são compatíveis com o Nível 5 para as meninas e Nível 6 para os meninos.

Na edição de 2017, os valores referentes ao terceiro QUARTIL, em Língua Portuguesa, são compatíveis com o Nível 4 da escala, para as meninas e Nível 3 para os meninos. Para Matemática, nessa mesma edição, os valores para as meninas são compatíveis com o Nível 3 da escala, e para os meninos com o Nível 4. Em síntese, na edição de 2017 a maioria dos alunos não conseguiram desempenho superior a 325 pontos nas proficiências analisadas.

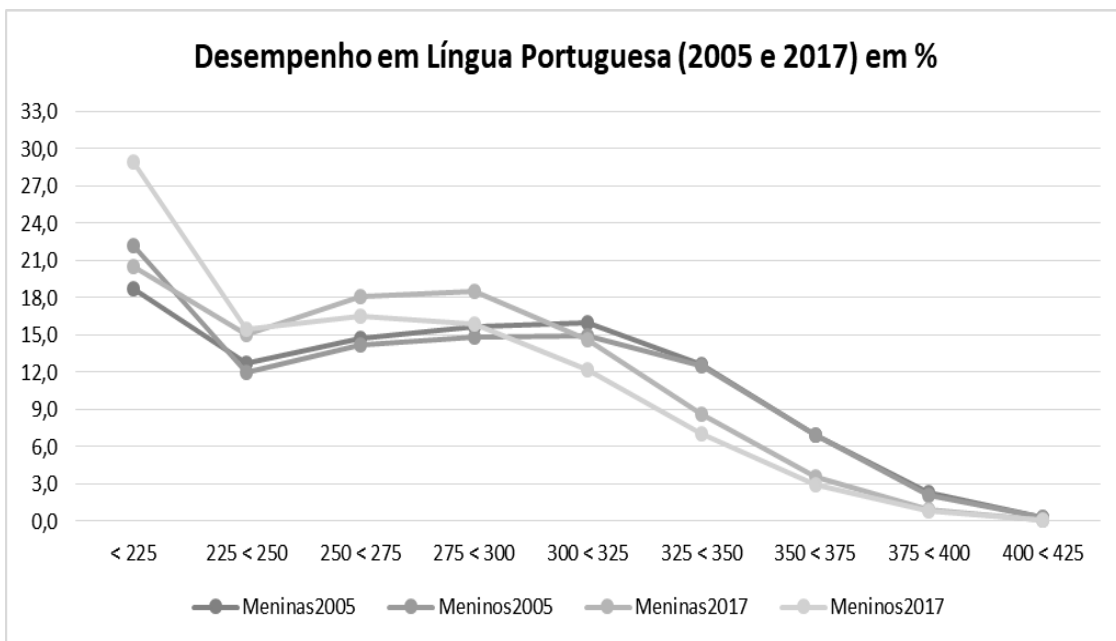
Se for utilizado, apenas como comparativo, o Nível ≥ 7 como um valor mínimo exigido para as proficiências, em Língua Portuguesa os percentuais apresentados na edição de 2005 seriam, para as meninas 2,6% e meninos 2,4%, e na edição de 2017, os percentuais seriam de 1% para as meninas e de 0,9% para os meninos. E em Matemática, na edição de 2005 os percentuais apresentados seriam de 10,5% para as meninas e de 18,5% para os meninos, e na edição de 2017, os percentuais seriam de 1,8% para as meninas e de 3,3% para os meninos.

Alguns estudos como os de Simões e Ferrão (2005), Marques (2010), Mesquita e Lelis (2015), Andrade et al. (2016), Tokarnia (2016), Américo e Lacruz (2017) e Marim e Bernardes (2017), já apontaram a necessidade de melhorias em toda estrutura do sistema educacional brasileiro, seja na qualidade da formação docente, nas práticas de gestão, na infraestrutura das escolas, para que se possa melhorar a qualidade ensino nas escolas brasileiras

Nota-se, em Matemática, uma quantidade maior de meninos com desempenho equivalente ao nível 10, tanto na edição de 2005 quanto na edição de 2017, isso pode indicar a necessidade de adaptações nas metodologias de ensino e outras práticas pedagógicas, que possam atrair e motivar as meninas para o estudo e empenho na área de exatas (ANDRADE et al., 2016; BRITO et al., 2016; GRAMANI e SCRICH, 2012; SANTOS e TOLENTINO-NETO, 2015). As Figuras 5 e

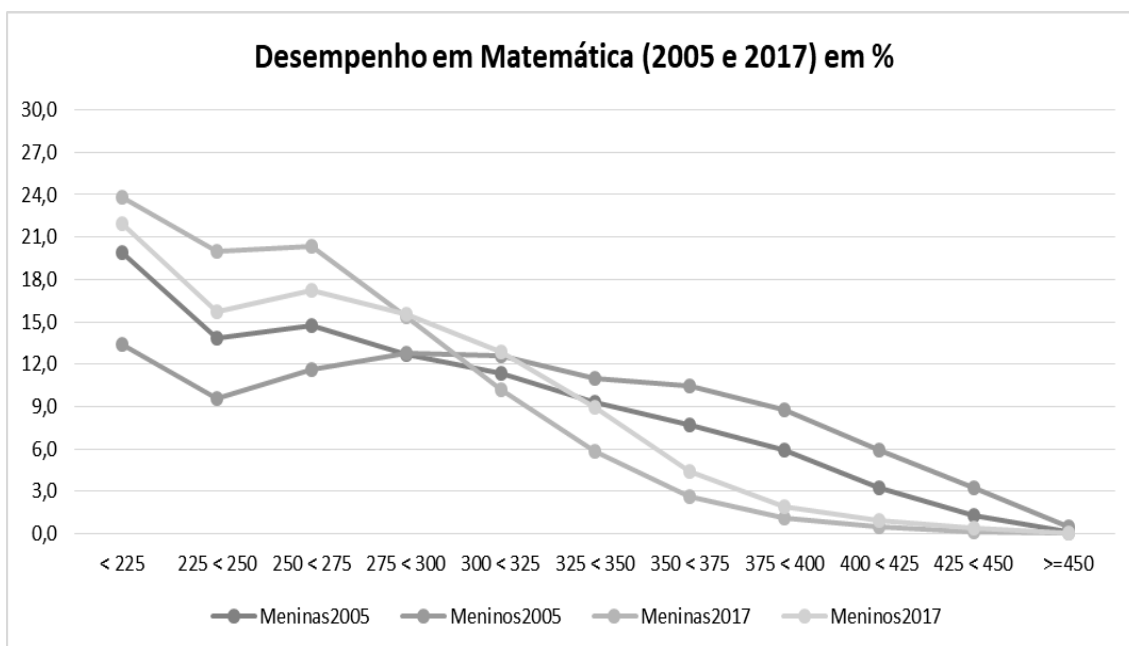
6 ilustram graficamente os desempenhos dos alunos nas edições de 2005 e 2017, tanto para Língua Portuguesa quanto para Matemática.

Figura 5. Desempenhos em Língua Portuguesa nas edições de 2005 e 2017.



Fonte: Desenvolvido pelos Autores (2020).

Figura 6. Desempenhos em Matemática nas edições de 2005 e 2017.



Fonte: Desenvolvido pelos Autores (2020).

Os resultados dos desempenhos dos alunos tanto em Língua Portuguesa quanto em Matemática mostram um cenário preocupante em relação à formação educacional do indivíduo, pois como já citado nesse estudo, o mercado de trabalho necessita cada vez mais de profissionais capacitados e com um bom nível de conhecimento.

Após análises realizadas nos dados expostos nesse estudo chega-se ao entendimento de que se torna necessário desenvolver ações que possam promover melhorias em todo ambiente educacional, pois a aprendizagem dos alunos pode ser influenciada por diversas situações. Não se pode apenas indicar que os motivos do fraco desenvolvimento educacional do aluno sejam ocasionados por uma condição ou outra. É importante que todas as dimensões da área educacional sejam observadas de igual forma, desde a condição socioeconômica da família, condições de infraestrutura das escolas, formação e capacitação do corpo docente e da gestão escolar (SILVA e SILVA, 2018; BASSETTO, 2019; NASCIMENTO e CHIUSOLI, 2019; ARAUJO et al., 2020; NASCIMENTO e GOMES, 2020).

Outro fator a ser observado é o volume de investimentos que o governo deve destinar à educação, pois mesmo que as metas de investimentos sejam cumpridas, se torna necessário uma melhor análise e controle no direcionamento desses valores, por que somente destinar o dinheiro não é certeza de que sejam alcançados melhores resultados nas avaliações em larga escala ou em níveis de aprendizado, conforme já observado por (GALVÃO, 2019; RODRIGUES e SOUSA, 2019; SOARES e ROSA, 2020).

O sistema educacional brasileiro carece de maior atenção tanto dos governos quanto da sociedade em geral, pois somente com uma educação de qualidade é que o indivíduo terá condição de exercer seu papel na sociedade, na família, no trabalho e principalmente como cidadão. Além disso, é a partir de uma educação com qualidade, que a nação poderá ter condições de se desenvolver economicamente, com melhorias nos sistemas de produção/serviços, inovação tecnológica, agronegócio, sustentabilidade e responsabilidade social.

5. Considerações Finais

O objetivo desse estudo foi analisar a evolução do desempenho dos alunos egressos do Ensino Médio, nas avaliações em larga escala do SAEB. Esse objetivo foi dividido em duas etapas, sendo que na primeira, foram analisadas as evoluções

dos desempenhos médios dos estados e das regiões brasileiras, no período de 2005 a 2017, e na segunda etapa, foram analisados os desempenhos dos egressos em nível nacional, tendo como base as edições de 2005 e de 2017. As proficiências contempladas nas análises foram Língua Portuguesa e Matemática.

Por meio das técnicas de Estatística Descritiva, os resultados indicaram que existe a necessidade da criação de planos de ação que possibilitem identificar defasagens de aprendizado, para que, com isso, os gestores tenham condições de desenvolver alternativas de melhorias para o ambiente educacional. Apenas ressaltando, que de acordo com os dados referentes aos desempenhos médios estaduais, no período 2005-2017, em Matemática, 12 estados apresentaram evolução negativa nessa proficiência.

Já em relação aos desempenhos dos egressos do Ensino Médio referentes as edições de 2005 e 2017, e se observados com base na MODA, percebe-se que a maior concentração de alunos está compatível com o Nível Zero da escala de desempenho, tanto em Língua Portuguesa quanto em Matemática, independente do gênero do aluno. E quando observados pela MEDIANA, pode-se perceber que na edição de 2005 os valores para Língua Portuguesa foram compatíveis com Nível 3 para meninos e meninas, e em Matemática, compatíveis com Nível 3 para as meninas e Nível 4 para os meninos. Já na edição de 2017, os valores para as MEDIANAS foram compatíveis com Nível 2, tanto para as meninas quanto para os meninos, nas duas proficiências analisadas, indicando com isso uma regressão de desempenho dos egressos da edição de 2017 em relação aos egressos da edição de 2005.

Com base nas análises realizadas e resultados apresentados, presume-se que existem necessidades de adaptações nos sistemas e metodologias de ensino, no entanto importante ressaltar que somente observar esses tópicos pode não ser suficiente para que se produza efeitos positivos no ambiente educacional. Se tornam necessários esforços mais amplos, envolvendo o governo, o ambiente escolar, as famílias e a comunidade onde as escolas atuam, para que em conjunto, debates e projetos sejam desenvolvidos em prol de melhorias para todo o sistema de ensino.

Deve-se ressaltar ainda, que os planos de ação que poderão ser desenvolvidos com intuito de propor melhorias no sistema educacional, deverão contemplar os aspectos sociais, socioeconômicos, culturais e específicos de cada região, pois não basta apenas formar o indivíduo para o mercado de trabalho, deve-

se além de tudo, formar o cidadão que exercerá seu papel na sociedade e na família.

Como sugestão para estudos futuros, seria interessante replicar a metodologia desse estudo contemplando os resultados de egressos dos 5º ou 9º anos do Ensino Fundamental, inclusive com os egressos do Ensino Médio na edição do SAEB de 2019, quando os dados estiverem disponíveis. Além disso, utilizar os resultados das edições do ENEM ou PISA para observar se houve evolução no nível de conhecimento dos alunos participantes dessas avaliações em larga escala.

Referências

- AIRES, R. W. DO A.; MOREIRA, F. K.; FREIRE, P. DE S. Indústria 4.0: Competências Requeridas aos Profissionais da Quarta Revolução Industrial. *International Congress of Knowledge and Innovation - Ciki*, v. 1, n. 1, 7 set. 2017.
- AMÂNCIO-VIEIRA, S. F. et al. A Relação entre Custo Direto e Desempenho Escolar: Uma Análise Multivariada nas Escolas de Ensino Fundamental de Londrina/PR. *Educação em Revista*, v. 31, n. 1, p. 169–194, mar. 2015.
- AMÉRICO, B. L.; LACRUZ, A. J. Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla. *Revista de Administração Pública*, v. 51, n. 5, p. 854–878, out. 2017.
- AMORIM, N. B. DE S. et al. A percepção ambiental dos estudantes do ensino médio sobre o cuidado com a sala de aula. *Revista Psicopedagogia*, v. 35, n. 107, p. 156–167, ago. 2018.
- ANDRADE, M.; FRANCO, C.; DE CARVALHO, J. P. Gênero e Desempenho em Matemática ao final do Ensino Médio: Quais as relações? *Anais*, p. 1–16, 2016.
- ARAUJO, J. M. DE et al. Fatores determinantes do desempenho educacional no Sudeste Brasileiro. *Gestão e Sociedade*, v. 14, n. 38, p. 3507–3525, 16 abr. 2020.
- BALTAR, P. Crescimento da economia e mercado de trabalho no Brasil. [s.l.] Texto para Discussão, Instituto de Pesquisa Econômica Aplicada (IPEA), 2015. Disponível em: <<https://www.econstor.eu/handle/10419/121610>>. Acesso em: 27 dez. 2018.
- BASSETTO, C. F. Family background and school performance: an approach with binary variables from SARESP results. *Revista Brasileira de Estudos de População*, v. 36, 2019.
- BORGES, J. C. A Qualificação Profissional do Trabalhador para o Mercado de Trabalho e Ambiente Organizacional. *faccat*, p. 17, 2018.
- BRASIL. Lei no 9.394, de 20 de dezembro de 1996, Estabelece as diretrizes e bases da educação nacional. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L9394.htm>.
- BRITO, C.; PAVANI, D.; LIMA JR, P. Meninas na Ciência: atraindo jovens mulheres para carreiras de Ciência e Tecnologia. *Revista Gênero*, v. 16, n. 1, 19 ago. 2016.
- BRITO, M. V. S. DE. Determinantes do sucesso escolar: uma análise multinível a partir dos dados do Pisa 2015. Dissertação de Mestrado—Brasília: Universidade de Brasília, 23 jun. 2017.
- BRUNS, J. P.; NUNES, C. D. C. Relações entre a Qualidade na Educação e as Avaliações em larga Escala no Brasil. *Professare*, v. 7, 12 ago. 2018.
- BUENO, A. L. Competências e empregabilidade: sentidos da formação nas políticas curriculares para o Ensino Médio. *Sociologias Plurais*, v. 5, n. 2, 30 dez. 2019.

- CALIXTO, K.; SEGUNDO, C.; GUSMÃO, R. P. DE. Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 28, n. 1, p. 1447, 27 out. 2017.
- CHIRINÉA, A. M.; BARREIRO, I. M. DE F. Qualidade da educação: eficiência, eficácia e produtividade escolar. *Revista on line de Política e Gestão Educacional*, v. 0, n. 7, 2009.
- COMAZZETTO, L. R. et al. A Geração Y no Mercado de Trabalho: um Estudo Comparativo entre Gerações. *Psicologia: Ciência e Profissão*, v. 36, n. 1, p. 145–157, 2016.
- CONAE. Conferência Nacional de Educação. Construindo o sistema nacional articulado: o plano nacional de educação, diretrizes e estratégias de ação. Documento Final. Brasília: MEC/CONAE, 2010. 163p. Disponível em: <http://conae.mec.gov.br/images/stories/pdf/documentos/documento_final_sl.pdf>.
- CORREIA, L. M. AVALIAÇÕES EM LARGA ESCALA: relações com o desempenho dos educandos nas avaliações nacionais. Dissertação de Mestrado—Maranhão - Brasil: Universidade Federal do Maranhão, 27 dez. 2017.
- FONSECA, S. O. DA; NAMEN, A. A. Mineração em Bases de Dados do INEP: Uma Análise Exploratória para Nortear Melhorias no Sistema Educacional Brasileiro. *Educação em Revista*, v. 32, n. 1, p. 133–157, mar. 2016.
- GALVÃO, F. V. Gastos educacionais e desempenho escolar: limites e possibilidades de uma abordagem. *Cadernos de Pesquisa*, v. 49, n. 171, p. 10–33, 30 abr. 2019.
- GARCIA, P. S. et al. A infraestrutura das escolas de ensino fundamental da Região do Grande ABC paulista. *Revista Ibero-Americana de Estudos em Educação*, v. 9, n. 3, p. 614–631, 2014.
- GATTI, B. A. Formação inicial de professores para a educação básica: pesquisas e políticas educacionais. *Estudos em Avaliação Educacional*, v. 25, n. 57, p. 24, 30 abr. 2014.
- GRAMANI, M. C. N.; SCRICH, C. R. Influência do desempenho educacional na escolha da profissão. *Cadernos de Pesquisa*, v. 42, n. 147, p. 868–883, 2012.
- GUSMÃO, F. A. F.; AMORIM, S. S. O percurso histórico do ensino médio no Brasil: uma reflexão sobre as políticas públicas de avaliação educacional. *Horizontes*, v. 38, n. 1, p. e020022–e020022, 30 abr. 2020.
- INEP. SAEB. Disponível em: <<http://portal.inep.gov.br/web/guest/educacao-basica/saeb>>.
- INEP. Enade. Disponível em: <<http://inep.gov.br/enade>>.
- INEP. INEP - IDEB. Disponível em: <<http://portal.inep.gov.br/web/guest/ideb>>.
- INEP. SAEB - Microdados. Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>.
- LOCKMANN, K.; MACHADO, R. Pátria educadora? Uma análise das propostas para o ensino público brasileiro. *Proposições*, v. 29, n. 1, p. 128–152, abr. 2018.
- MARANGONI, R. A.; ARAÚJO, F. S. DE G.; SCHWARZ, R. R. Diretor de escola: dificuldades e enfrentamentos. *Revista Triângulo*, v. 12, n. 2, p. 14–31, 8 maio 2019.
- MARIM, V.; BERNARDES, M. B. J. Formação inicial docente: em busca da qualidade da educação pública (Initial teacher training: in search of the quality of public education). *Crítica Educativa*, v. 3, n. 2, p. 237–252, 13 set. 2017.
- MARQUES, A. F. Aplicação da análise multivariada na infraestrutura e no desempenho das escolas públicas do Ensino Fundamental e Médio pertencentes ao Núcleo Regional de Educação de Paranaíba. *Acta Scientiarum. Technology*, v. 32, n. 1, 2010.
- MEC. PNE - Plano Nacional de Educação - Plano Nacional de Educação - Lei nº 13.005/2014. Disponível em: <<http://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>>.

- MESQUITA, S. S. DE A.; LELIS, I. A. O. M. Cenários do Ensino Médio no Brasil. Ensaio: Avaliação e Políticas Públicas em Educação, v. 23, n. 89, p. 821–842, 2015.
- MESSIAS, J. F.; SANTOS, O. S. DOS. O Déficit Educacional e a Exclusão Tecnológica no Mercado de Trabalho Brasileiro. Brasil Para Todos - Revista Internacional, v. 6, n. 1, 31 out. 2018.
- NASCIMENTO, M. C. DO; GOMES, G. R. R. Formação continuada docente para a utilização das TIC no processo de ensino e aprendizagem. Research, Society and Development, v. 9, n. 2, p. e33921998–e33921998, 1 jan. 2020.
- NASCIMENTO, V. P. DO; CHIUSOLI, C. L. O papel do gestor escolar: estudo de caso sobre os desafios da educação pública. Revista on line de Política e Gestão Educacional, v. 23, n. 2, p. 238–254, 6 mar. 2019.
- OCDE. PISA - PISA. Disponível em: <<http://www.oecd.org/pisa/>>. Acesso em: 29 set. 2018.
- OLIVEIRA JÚNIOR, M. C. DE et al. Recursos destinados à educação e resultados alcançados no Ideb de uma capital brasileira. Cadernos EBAPE.BR, v. 17, n. 3, p. 523–538, set. 2019.
- PATTI, Y. A. et al. Percepção de professores do ensino médio acerca da motivação docente. Revista Psicopedagogia, v. 34, n. 103, p. 53–64, 2017.
- RESCIA, A. P. O.; GENTILINI, J. A. Formação de gestores educacionais e escolares no contexto das tendências das reformas educacionais: consensos e dissensos. Revista Ibero-Americana de Estudos em Educação, v. 11, n. 25, p. 307–318, 2016.
- RODRIGUES, A. M. G.; SOUSA, E. P. DE. Eficiência dos Gastos Públicos em Educação Básica nos Municípios Brasileiros. Economic Analysis of Law Review, v. 10, n. 1, p. 193–219, 21 nov. 2019.
- SACOMANO, J. B. et al. (EDS.). Indústria 4.0: Conceitos e Fundamentos. Edição: 1a ed. [s.l.] Blucher, 2018.
- SANTOS, A. L. DOS. O Panorama das Notas em Matemática no SAEB dos Estudantes do 5º ano do Ensino Fundamental. Revista Educação Matemática em Foco, v. 8, n. 2, p. 4–27, 25 jun. 2020.
- SANTOS, J. B. P. DOS; TOLENTINO-NETO, L. C. B. DE. O que os dados do SAEB nos dizem sobre o desempenho dos estudantes em Matemática. Educação Matemática Pesquisa : Revista do Programa de Estudos Pós-Graduados em Educação Matemática, v. 17, p. 25, 2015.
- SAVIAN, M. P. G.; BEZERRA, F. M. Análise de eficiência dos gastos públicos com educação no ensino fundamental no estado do Paraná. Economia & Região, v. 1, n. 1, p. 26–47, 15 maio 2013.
- SILVA, R. M. DE S. DA; LEAL, D. A. A Família na vida escolar dos alunos: uma análise do relacionamento das instituições Escola – Família no município de Fazenda Rio Grande / The Family in the school life of students: an analysis of the relationship of the School - Family institutions in the municipality of Fazenda Rio Grande. Brazilian Journal of Development, v. 6, n. 8, p. 58476–58488, 18 ago. 2020.
- SILVA, T. R.; SILVA, J. B. L. A influência do nível socioeconômico na aprendizagem da leitura. Revista Educação, Psicologia e Interfaces, v. 2, n. 2, p. 7–28, 1 ago. 2018.
- SIMIELLI, L. E. R. Equidade e Oportunidades Educacionais: O Acesso a Professores no Brasil. Arquivos Analíticos de Políticas Educativas, v. 25, 2017.
- SIMÕES, M. DE F.; FERRÃO, M. E. Competência percebida e desempenho escolar em Matemática. Estudos em Avaliação Educacional, v. 16, n. 32, p. 25–42, 2005.

SOARES, C. S.; ROSA, F. S. DA. Gastos Públicos em Educação e Desempenho Escolar: Análise Qualitativa dos Resultados do IDEB em Santa Maria/RS. *Revista Eletrônica de Estratégia & Negócios*, v. 13, n. 0, p. 38–62, 8 maio 2020.

TABERNERO, C.; SERRANO, A.; MÉRIDA, R. Estudio comparativo de la autoestima en escolares de diferente nivel socioeconómico. *Psicología Educativa*, v. 23, n. 1, p. 9–17, 1 jun. 2017.

TARDA, K. B.; RODRIGUES, R. V. A Influência dos Gastos em Educação no Crescimento Econômico da Região Administrativa de Campinas. *Revista Estudo & Debate*, v. 22, n. 2, 23 dez. 2015.

TERNOSKI, D. M.; RIBEIRO, F.; CLEMENTE, A. A Influência da Aplicação de Recursos Públicos nas Áreas de Educação e Saúde com os Índices Sociais nos Municípios Paranaenses. *Revista Capital Científico - Eletrônica (RCCe) - ISSN 2177-4153*, v. 15, n. 1, p. 98–117, 12 mar. 2017.

THIMOTEO, F. E. P. A Avaliação da Educação Básica: uma análise qualitativa dos instrumentos contextuais do SAEB no período de 1995 a 2001. *Dissertação de Mestrado*- São Paulo: FGV, 2003.

TOKARNIA, M. Apenas 4,5% das escolas têm infraestrutura completa prevista em lei, diz estudo. Disponível em: <<http://agenciabrasil.ebc.com.br/educacao/noticia/2016-06/apenas-45-das-escolas-tem-infraestrutura-completa-prevista-em-lei-diz>>. Acesso em: 12 maio. 2018.

VASCONCELOS, J. C. et al. School infrastructure and public investment in Education in Brazil: the importance of educational performance. *Ensaio: Avaliação e Políticas Públicas em Educação*, n. AHEAD, 2020.

6.4 ARTIGO 4. Teaching practices and school performance: discovering predictors of good performance using Data Mining.

Ivonaldo Vicente da Silva

Márcia Terra da Silva

Nilsa Duarte da Silva Lima

ABSTRACT

This study aims to reveal which variables related to teachers' characteristics and professional activities could be associated with the best performing schools, relative to a large-scale assessment of Brazilians Secondary School students. To this end, the study applies standard Educational Data Mining techniques, such as Decision Trees and Random Forest, on the last national assessment databases. The results showed that, in addition to the school's socioeconomic level, the teacher's experience and improvements in teaching skills and competence are closely related to school performance. As a contribution to the literature, it has been concluded that the variables relating to teachers, taken individually, have greater weight as predictor factors of the students' performance than the socioeconomic school's status.

Keywords: School performance; Teaching experience; Data Mining.

1 INTRODUCTION

The purpose of this study is to observe the teachers' factors and activities associated with the educational performance of students enrolled in Secondary Schools, according to the National System to Assess Basic Education (SAEB in Portuguese).

In Brazil, the educational system consists of nine years of Primary School education and Secondary School, usually offered in three years (BRASIL, 1996). For students entering an undergraduate course, they must complete the entire basic educational cycle. Usually, children start Primary school at the age of six and complete the entire educational cycle in twelve years. In Brazil, according to the 2017 School Census, 48,608,093 students enrolled in the National Basic Education System, of which 8,887,061 were enrolled in private schools (INEP, 2019).

The quality of Brazilian education is far from ideal, and in public schools, the problems are becoming worse. With a public educational system that lacks investment, a high level of inequality to access resources and income, and even in some cases, a fragile family structure, young people need to struggle to stay in school and overcome obstacles to get a good education.

Unfortunately, there is still a considerable percentage of young people who do not finish the basic education cycle. According to the Brazilian Institute of Geography and Statistics - IBGE, in 2017, the illiteracy rate was 7.1% for men and 6.8% for women. In addition to this, the percentage of young people aged 25 or over who did not finish Primary School education was 33.8%, and among those who had started Secondary School Education, 4.4% did not finish it. (IBGE, 2018).

The schooling rate indicates the percentage of students in each age group enrolled in Basic Educational courses. According to IBGE, the school enrollment rate for young people between 15 and 17 years of age, in 2017, was approximately 87.2%. Of this percentage, only 68.4% of students attended Secondary School at the right age, indicating that 18.8% of students repeated years or dropped out of school. School dropouts were caused by various reasons and can be related to demotivation, lack of career expectations, the school environment, or even getting a job to supplement their family's income (IBGE, 2018).

Therefore, given the importance of basic education for young people, it is important to understand the conditions or facilitating factors behind a good school performance, knowing that this issue is more pressing in public education, where conditions are more precarious.

Moreover, maintaining young people at school requires investments in all education structure segments to ensure that students gain access to knowledge, regardless of their school-age. The specialized literature mentions the importance of investments in school infrastructure, teachers, school management, and administrative staff training, upgrading technological resources and laboratories, as well as other resources (Amâncio-Vieira et al., 2015; Lasater et al., 2020; Razak et al., 2018; van den Ham & Heinze, 2018; Ford, 2019.)

Among the essential investments, hiring, training and keeping teachers is frequently mentioned in the literature. After all, through the work of the teaching staff, students can feel more motivated to learn and to discover new knowledge (Soares et al., 2010; Pereira & Mori, 2012; Amâncio-Vieira et al., 2015; Fonseca & Namen, 2016; Vidal & Vieira, 2017). As such, understanding which activities and teacher behavior most affect the student results is key to planning how they are trained and how to improve their skills. Also, understanding how expectations and values become motivating factors for students can guide the care taken by school managers. As such, understanding which activities and teacher behavior most affect the student

results is key to planning how they are trained and how to improve their skills. Also, understanding how expectations and values become motivating factors for students can guide the care taken by school managers. (Soares et al., 2010).

As such, this study looks at the main factors related to teachers that are associated with the school performance of Secondary School students. To compose a database for this study, data from the teaching staff and schools was used, from the 'National System to Assess Basic Education' run by the Brazilian Government and made available on the Ministry for Education website.

The Assessment System is mandatory for all public schools and optional for private schools; therefore, most of the data comes from public schools since approximately 81% of students were enrolled in public schools in 2017 (INEP, 2019). One of the challenges for carrying out this study is the reliability of the data collected so that there are no distortions or errors during the analysis. Another challenge is dealing with a considerably large volume of data, and, because of this, the techniques to be used for analysis must be appropriate to this (Sivarajah et al., 2017).

One of the alternatives to study and analyze large volumes of data in order to obtain information on certain scenarios is Data Mining (Martínez-Abad & López, 2017). Through techniques, either supervised or unsupervised, Data Mining allows data to be analyzed, and it can even predict patterns and behaviors of the variables available. As such, its use has increased over time, and in the area of education, it has become a good alternative for analyzing performance factors and, so, contributes to the decision-making process (Provost & Fawcett, 2013; Martínez-Abad & López, 2017).

In developing this study, the chapter on the Review of Literature comes immediately after the Introduction. It addresses the publications that support the study. The chapter on the Methodology explains the methods used to develop this study. The results are then shown for discussion and reflection, and finally comes the chapter on Final Considerations.

2 REVIEW OF THE LITERATURE

Several aspects can be related to the quality of education, and they must be observed in their context. By observing students' needs, issues relating to access to essential resources such as sanitation, food, safety, status of the family, and the

family structure become paramount so that educational development is safer and more peaceful.

The government is socioeconomic expected to invest in and modernize the educational system, from the school infrastructure to training all those involved in the educational environment. When it comes to the school environment, in addition to an adequate infrastructure for teaching, the teachers, school managers, and others involved must have the skills and knowledge to perform their roles.

Studies already published support these aspects, repeating academic years at school, for example, becomes worrying as students are unable to finish their studies at the expected ages (Paula et al., 2018; Romero et al., 2018). Repeating academic years, besides affecting the career of the young person, also disturbs the school organization. This happens because by repeating academic years, the young person delays his entry into undergraduate courses, and also becomes another student in a primary education course classroom, occupying a vacancy for another student who is enrolling in the correct year for their age (INEP, 2020b).

The family, which is the main structure for the young person, should participate more effectively in his educational development. Parents' educational background can become a motivating factor for the student to feel more supported in their educational activities, even contributing to the social and professional development of these young people (Barbosa & Wilhelm, 2009; Brito, 2017; Jovarini et al., 2018).

Regarding teaching practices, it is important to note that the teacher's role becomes essential for students to learn since this professional works directly in the school environment with the student to develop and pass on knowledge (Pettersson et al., 2016; Schaik et al., 2018).

It is part of the teacher's activities to develop teaching tasks that students perform in the classroom and at home. These tasks aim at reinforcing the content developed in the classroom, and by correcting these exercises, the teacher is able to observe whether students present learning difficulties (Marturano & Pizato, 2015; Brito, 2017; Jovarini et al., 2018). It should also be noted that the volume of tasks and the degree of complexity must be at the same level as the conditions of how they are carried out by the students, avoiding exhaustion, stress, and demotivation.

Other aspects related to teachers are also important for school performance, such as holding a Master's degree or Doctorate, the level of relationship in the school

environment including their leaders, expectations regarding the students' educational performance, motivation, and satisfaction with their professional career, among others.

The level of expectation regarding students' educational performance becomes paramount for a favorable learning environment to be developed. The motivation, discipline, and commitment to carrying out all activities relating to the classroom's daily routine must be part of the teacher's profile. When the teacher is motivated, everyone notices it, mostly the students (Lauermann, 2017; Schildkamp et al., 2020; Zakariya et al., 2020).

Identifying all these factors found in the school environment, which directly relates to students' performance, is made easier with the support of technological resources. Using specific tools, information technology professionals can prepare reports that help managers in decision-making processes, and a good option for this is Educational Data Mining.

Educational Data Mining can be used to find patterns and correlations in situations involving a large volume of data and is even used to predict results. Through mining techniques and algorithms, the user can obtain reliable information to improve the decision-making process, such as improving the relationship with customers, reducing risks, predicting performance (Romero & Ventura, 2007; Witten et al., 2011; Provost & Fawcett, 2013).

For some authors, Data Mining is considered part of a process of knowledge discovery in databases (KDD), and it may use techniques and algorithms, supervised or not (Romero & Ventura, 2007; Provost & Fawcett, 2013).

In tasks where unsupervised data mining techniques are used, the algorithms do not use predetermined data type specifications; patterns in the data are discovered automatically. In supervised data mining, built models are used, and because of this, the types of data that will be used must be stated, highlighting the variables considered targets. The most frequently used techniques in supervised data mining are Decision Trees, Random Forest, Neural Networks and Clusters (Baker & Yacef, 2009; Martínez-Abad & López, 2017).

A Decision Tree encompasses a hierarchical model of nodes and leaves and with a graphic representation ready, prediction factors can be easily analyzed. These Trees present necessary and sufficient information for making important decisions in the analyzed environment. Examples of algorithms used to build Decision Trees are

ID3, J48, C4.5, CART and Random Forest (Provost & Fawcett, 2013; Silva et al., 2016; Martínez-Abad & López, 2017;).

The studies involving Educational Data Mining are broad and diverse; however, most of the publications use Decision Trees as an alternative to creating knowledge based on a large volume of data.

2.1 EDM Related Work

As an example of studies already published involving Educational Data Mining, Tuaha et al. (2019) used the Naive Bayes, Decision Trees, and Artificial Neural Network algorithms in order to predict students' behavior concerning punctuality in classes and parental participation in the learning process. As a result, the authors concluded that these factors are fundamental to students' academic success.

Studies by Martínez-Abad & López (2017) aimed to observe students' performance in the city of Baja California in Mexico. After analyzing a sample of 18,935 Secondary School students based on Decision Tree algorithms, they concluded that the results indicated that personal factors are more indicative of academic performance, followed by school and social factors. The authors compared results found in the available literature and concluded that Decision Trees could be used with confidence to predict academic performance, which was better, in some cases, than other algorithms.

In order to fill a gap about the use of Educational Data Mining in International Educational Assessment, Depren et al. (2017) carried out a study to predict the mathematics performance of students from the eighth grade who participated in the broad assessment -IEA's TIMSS scale, from Turkey. The Decision Tree, Bayesian Network, Logistic Regression, and Neural Network algorithms were used as the Mining technique. As a result, they found that the logistic regression showed better performance in relation to the other algorithms used, and that the student's confidence factor can contribute to the students' performance in the results presented for mathematical proficiency.

Also, using the results of students in mathematical proficiency, Bravo Sanzana et al. (2015) used the Random Forest and CART algorithms to identify the students' profiles in the large-scale assessments of 2011, in Chile. As a result, they

found that the parents' educational expectations, the type of school and the math skills index could influence the good performance in the assessments.

Using data from secondary school students in Portugal, Ünal (2020) applied Decision Tree, Random Forest, and Naive Bayes algorithms to predict these students' final grades in the Portuguese Language and Mathematics proficiencies. The author used data referring to students' grades in the analyzed proficiencies and social, demographic and school-related resources. The results indicated that of the three algorithms used, Random Forest demonstrated better performance to predict the good performance of students.

In order to observe the factors that affect performance in Mathematics in the large-scale PISA assessment of Turkish students in the 2015 edition, Güre et al. (2020) used the Multilayer Perceptron, Artificial Neural Networks, and Random Forest algorithms on a database of 4,422 students who performed the assessment. Besides, the analysis pointed out that the following attributes are the most important for the prediction: their level of success, the level of education of the mother, the students' motivation, their belief in epistemology, the level of interest of teachers, and the disciplinary environment of the classroom.

Other studies based on the use of Educational Data Mining contributed to the development of this investigation. However, these selected examples reveal techniques well applied, clarifying possibilities for further studies explore whenever necessary.

It is important to note that this study centered on analyzing factors present in the educational environment, especially the teaching staff's activities or behavior. Furthermore, in no way this analysis is expected to exhaust the subject of revealing the factors that impact students' performance in the Secondary School.

As such, the main objective of this study is to understand, through Educational Data Mining, the teaching staff's factors or activities that are associated with the teaching-learning environment, so that Secondary School students can obtain better results in large-scale assessments of the Basic Education Assessment System (SAEB).

3 METHOD

For this study, it was necessary to initially read publications relating to the topic, in order to identify relevant publications that could contribute to its development. This phase was critical because, as a result, a list of factors relating to the students' performance was obtained. These factors were then used to guide the data pre-processing, as explained below.

In the second step, data was collected from the Federal Government website, relating to large-scale assessments of the Basic Education Assessment System in Brazil. The data package includes the results of the students in the assessment, in addition to questionnaires completed by the students themselves, by teachers, by the Directors and School Representatives (INEP, 2020a).

The complete database relating to the 2017 assessment contains 1,966,507 records of Secondary School students 73,674 records of School Principals, 753,668 records of teachers and finally 73,674 records from Schools. Note that respondents cannot be identified, since the names of schools, teachers, and students are not recorded in the database.

3.1 Data pre-processing

This study focuses on the forms completed by teachers and by schools offering Secondary School courses in 2017. Initially, questionnaires from teachers working in Secondary Schools and the respective schools where they work were listed. As a result of this step, a database was created called "ProfSchool", which was used for the data mining phase.

The next step was to select the attributes for the data analysis. SAEB forms have 125 questions for teachers and 74 questions for schools, and from these, attributes were chosen based on the factors mentioned in the bibliography. The result of this step is described in Table 1.

Attributes	Form questions
Academic background and experience	
GENDER	1. What is your gender?
GRADE (graduate)	4. What is the highest level of education you have completed (and graduated)?
POSTGRADE (Postgraduate)	8. State the highest postgraduate degree course you have.
EXP_FUNCTION	13. How many years have you been a teacher?

(Time of experience in the profession)	
Need for Professional Improvement	
IMPROVE_NCP (Need for Improvement in National Curriculum Parameters)	26. Curriculum parameters or guidelines in your area of work.
IMPROVE_SPEC_CONTENT (Need for Improvement in Specific Content)	27. Specific content in my main teaching subject.
IMPROVE_EVALUAT_METH (Need for improvement in evaluation methodologies)	30. Student assessment methodologies.
IMPROVE ICT (Need for Improvement in Information and Communication Technology)	31. Information and Communication Technologies used in teaching.
Resources used for Teaching Purposes	
USE_BOOKS_EDU_PURPOSES (Uses books for Educational Purposes)	45. Printed literature in general.
USE_INTERNET_EDU_PURPOSES (Uses Internet for Educational Purposes)	50. The Internet.
Relationship with the Principal	
PRINC_TEAC_TEACHING_QUALITY (Principal and teachers - quality assurance of education)	59. The principal and teachers seek to ensure that the quality of teaching is a collective responsibility.
PRINC_ATTEN_LEARNING_STUDENT (Principal pays attention to student learning.)	61. The principal is especially attentive to aspects related to student learning.
PRINC_MOTIVATES_ME (Principal motivates me)	64. The principal encourages and motivates me to work.
PRINC_RESPECTS_ME (Principal respects me)	66. I feel the principal respects me.
HAVE_CONFIDENCE_IN_PRINCIPAL (I have confidence in principal)	67. I have confidence in the principal as a professional.
PARTICIPATE_DECISIONS (I participate in school decisions)	68. I take part in the decision making related to my work.
Reasons for possible learning problems in students	
LACK_INEF_SUPERVISION (Lack or inefficiency of supervision)	71. Lack of or inefficient supervision, coordination, and teaching guidance.
TEACHER DISSAT CAREER (Dissatisfaction with the teaching career)	75. Teacher dissatisfied and discouraged with teaching career.
Expectations about the future educational trajectory of the students	
EXPEC_FINISH_HIGH_SCHOOL (Level of expectation in relation to students completing high school.)	95. (For all Teachers). How many students in this class do you think will finish Secondary School?
EXPEC_ENTER_COLLEGE (Level of expectation in relation to students attending college)	96. (For all Teachers) How many students in this class do you think will go to university?
The Subject taught	
DISCIPLINE	105. In this class, which subjects do you teach?

How often does the teacher develop the following teaching practices in this class?	
PROPOSE_HOMEWORK (If the teacher proposes homework)	107. Propose homework.
CORRECT_HOMEWORK (If the teacher corrects the homework)	108. Correct homework with students.
DEV_GROUP_ACTIV (If the teacher develops group activities in the classroom)	109. Develop group activities in the classroom, so that students seek solutions to problems.
school data: socioeconomic level, average proficiencies and scale levels	
SOCIOECONOMIC_LEVEL (socioeconomic level)	Provided by the Federal Government.
AVG_LP (Average Portuguese language proficiency achieved by the school)	Calculated based on school grades averages on Portuguese.
AVG_MT (Average Mathematical proficiency achieved by the school)	Calculated based on school grades averages on Math.
SCALE_LEVEL_LP (School level in Portuguese language proficiency)	As provided by the government
SCALE_LEVEL_MT (School level in mathematical proficiency)	As provided by the government

Table 1. List of factors in questionnaires from Teachers and Schools.

The next step involved the treatment of the data to eliminate blank, null, or incorrectly completed fields, as such, all the records in the “ProfSchool” database were analyzed. With the result of this step, it was possible to obtain a database with all fields correctly completed, thus avoiding possible errors in the data mining phase. At the end of this stage, of the total number of 205,299 available records, only 69,980 could be used.

To facilitate reading the results in the mining phase, and even to identify the data better, the values from the responses in the “ProfSchool” database have been replaced by more understandable forms. As an example, for the question about the teacher's gender, there were two possible answers, "A" or "B". In the reports, they appear as "MASC" for answer "A" and "FEM" for answer "B". All changes were made based on the forms' alternatives.

Regarding student performance, the assessment scale ranges from 0 to 500 points, with Level 1 of the scale starting at 225 points and progressing every 25 points forming regular intervals. As such, Level 1 comprises performances between 225 and 250 points, Level 2 between 250 and 275 points and so on. The assessment

assumes that students who did not obtain 225 points have not yet been able to show the elementary and essential skills that are compatible with their stage of schooling (INEP, 2020a).

The information in the "ProfSchool" database regarding the performance of schools refers to the average that each school obtained from its students' results, so the higher the average value, the better student performed in assessments. This value was provided by the Federal Government and was already separated by proficiency. Together with the averages achieved, the Government also provides the socioeconomic level of each school, classified into seven levels, which are Very High, High, Medium High, Medium, Medium Low, Low and Very Low.

To facilitate the visualization and analysis of a school's performance in each analyzed proficiency, four groups were identified as follows: "LevelZero", containing schools with an average below 225 points, "Level1to3" for schools that have averages between 225 points and 300, "Level4to6" for schools with averages between 300 and 375 points, and finally "Level7to10" which includes schools with averages equal to or greater than 375 points.

Before exporting the data from the "ProfSchool" database to the specific software for Data Mining, a final check was done to identify possible form completion errors. For the data mining phase, a free educational version of the software RapidMiner Studio 9.7 was used, a choice based only on convenience and practicality.

3.2 Educational Data Mining

In the Data Mining phase, two models were used for analysis. In the first one, Decision Trees were built, and in the second one, a Random Forest was built; both can be used for both classification and regression tasks. The data set and target attributes used were the same for both models.

Decision Trees can be considered a support tool for decision making. It generates a graph in a tree format to illustrate the results using supervised learning algorithms. The process of creating a Decision Tree consists of finding rules on the variables, which return the greatest gain in information (Provost & Fawcett, 2013). The first node or rule is the one that produces the greatest gain of information and, therefore, reveals the main predictive attribute for the target condition. The Decision Tree allows you to learn about the data's behavior by observing the path from the

main or root node to the leaf or final node, and this knowledge is an important support to the decision (Gomes & Almeida, 2017; Martínez-Abad et al., 2020).

Random Forests are also considered a supervised learning algorithm, which creates a combination of Decision Trees (a forest), each tree using a combination of the attributes surveyed. One of the differences of this algorithm for conventional Decision Trees is that in Random Forests, a combination of decision trees trained with a bagging method is created, which consistently increases the overall result. Several small trees with inferior performance are created, but together they can increase performance to predict the attributes, using the average between the trees to indicate the best prediction attributes. Random Forests still have the advantage of identifying the relative importance of each attribute for the prediction (Ahmed & Hikmat Sadiq, 2018).

The Decision Tree algorithms use measurements of impurity degree, such as the Gini Index and Entropy. After calculating the Entropy, the criterion to construct the tree must be chosen. For that, it is necessary to observe which criterion can present a better gain of information regarding the target attribute. The criteria that can be used for this purpose are: `information_gain`, `gain_ratio`, `gini_index`, `accuracy`, and `least_square`.

In order to find a criterion that could better explain the results regarding Decision Trees and Random Forests, several tests were carried out with the available criteria options, and finally `information_gain` was chosen to both construct Decision Trees and for the Random Forest, for having the best information gain. The formulas used for the Entropy Calculation (1) and Information Gain (2) are:

$$\text{Entropy (Y)} = - \sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Where Y is the set of cases and $p(c | Y)$ the value of Y is the proportion of class c.

$$\text{Information Gain} = (Y, a) = \text{Entropy (Y)} - \sum_{v \in \text{Values}(a)} \frac{|Y_v|}{|Y|} \text{Entropy (Y}_v) \quad (2)$$

Where values (a) represent all possible values in a set of cases. Y_v is a subclass of class Y_v related to class (a).

With the data in RapidMiner, the first step was to create Decision Trees to identify which factors could be associated with student performance. Two trees were

created, one targeting “NivelEscalaLP” for Portuguese language proficiency and the other “NivelEscalaMT” for Mathematics. Several tests were carried out until the highest accuracy, the Kappa index, and the error percentage were found. The best setting was criterion = information_gain, maximal depth = 12, apply pruning, confidence = 0.1, Cross Validation number of folds = 10, for the two proficiencies analyzed.

Results were analyzed based on the path to reach the highest level on the school's performance scale, such as “Level7to10”. It is worth mentioning that if the school does not present occurrences in the “Level7to10” class, the previous class will be used for analysis, that is, “Level4to6” and if there are also no records in that class, the previous class should be used, in this case, the “Level1to3” class, and so on.

Then, the Random Forest algorithm was applied to identify how relevant each attribute is in relation to the target attribute. Several tests were also performed for the Random Forest, changing the number of aggregate trees (bagging) to 100, 70, 50 trees, and the settings used were criterion = information_gain, maximal depth = 12, apply pruning, confidence = 0.1, Cross Validation number of folds = 10.

In this step, the results were analyzed based on the attributes that presented the highest importance values based on the Weight by Tree Importance. This observation aimed to identify the attributes that can be associated with student performance in large-scale assessments in Brazil. It is worth mentioning that the edition analyzed, and the focus of this study, was the 2017 results of Secondary School students at a national level.

3.3 Design of the development of the Data Mining phase

Research design is a fundamental part of understanding all the steps in conducting the study. Thus, Figure 1 illustrates the steps needed to develop the Data Mining phase of this research.

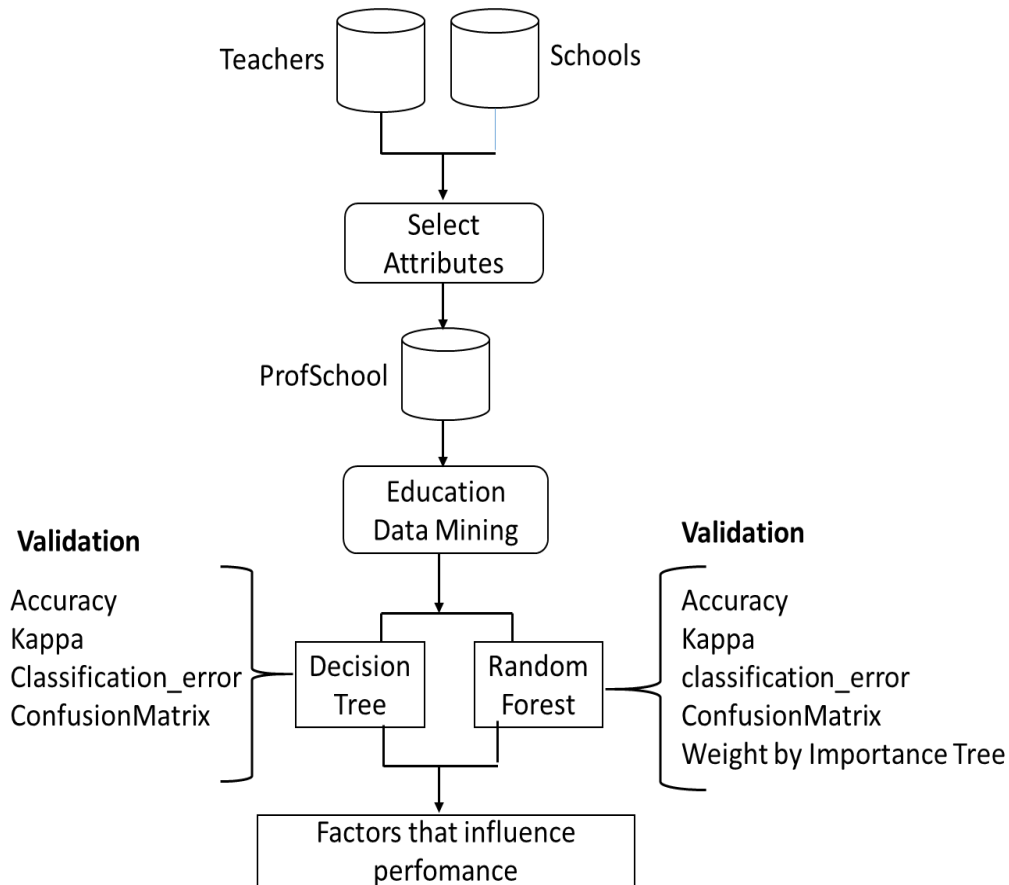


Figure 1. Steps for the development of the Data Mining phase

4 RESULTS AND DISCUSSIONS

The results will be analyzed in two stages, firstly, the Decision Trees, and then the results of Random Forest. In data mining, attributes are expected to be found for the school context with the greatest potential to predict good results in the "Scale level in Portuguese" or "Scale level in Mathematics" of students. As such, the trees will indicate the attributes that produce the highest proportion of students at the highest scales.

4.1 Decision tree

The results of the first phase of Data Mining, which was the construction of Decision Trees, illustrated that the main factor considered (main node) was the school's "SOCIOECONOMIC_LEVEL". This result was presented for both proficiencies in Portuguese and Mathematics.

Based on this indication, the schools analyzed in this study were grouped by socioeconomic level, illustrated in Table 2, to obtain a better view of the scenario.

Socioeconomic level	Total	%
Very High	0	0
High	209	1,66
Medium High	1030	8,18
Medium	4339	34,44
Medium Low	5597	44,43
Low	1111	8,82
Very Low	312	2,48
Total	12598	100

Table 2. Total schools by Socioeconomic Level

This result from the Decision Tree confirms the relationship between the socioeconomic level and young people's educational performance, widely cited in the literature (Laros et al., 2012; Gustafsson et al., 2018; Murphy, 2019). Unfortunately, according to the data in Table 2, it is clear that among schools analyzed, 44.43% are classified as socioeconomic level Medium Low and 24.44% as Medium, together adding up to 78.87% of schools. On the highest levels, only 1.66% of schools are classified with socioeconomic status High and 8.18% classified as Medium High, totaling 9.84% of the total schools.

Following the separation logic at the school's socioeconomic level, the results of the decision trees will also be analyzed in this manner. Firstly, this visualization model can facilitate a better understanding of the results for each situation of the socioeconomic level, in addition to allowing the extent to which the attributes found can be associated with the students' performance to be observed.

After several tests with different configurations for the Decision Trees, the best combination was depth = 12. In this configuration, for the Portuguese language proficiency, the accuracy was 94.21% +/- 0.22%, the Error Classification was 5.79% +/- 0.22%, and the Kappa Index was 0.687 +/- 0, 01 and the Confusion Matrix, Table 3, illustrates a very reasonable result for this model.

True:	Level4to6	Level1to3	LevelZero	Level7to10
Level4to6:	4487	1242	30	2
Level1to3:	1296	60893	1038	0
LevelZero:	17	429	546	0
Level7to10:	0	0	0	0

Table 3. Confusion matrix for Decision Tree - Portuguese Language

Regarding the attributes identified as possible predictors of good results in terms of school performance, Table 4 presents the fragments found for each socioeconomic level.

Socioeconomic level = High
SOCIOECONOMIC_LEVEL = High CORRECT_HOMEWORK = Weekly PRINC_TEAC_TEACHING_QUALITY = AlwaysAlmostAlways: Level4to6 {Level4to6 =308, Level1to3=0, LevelZero =0, Level7to10 =2}
Socioeconomic level = Medium High
SOCIOECONOMIC_LEVEL= MediumHigh EXPEC_ENTER_COLLEGE = AlmostAll PROPOSE_HOMEWORK = Weekly: Level4to6 {Level4to6 =154, Level1to3 =1, LevelZero =0, Level7to10 =0}
Socioeconomic level = Medium
SOCIOECONOMIC_LEVEL = Medium EXPEC_ENTER_COLLEGE = AlmostAll POSGRADE = DoctorateDegree PRINC_RESPECTS_ME = AlwaysAlmostAlways: Level4to6 {Level4to6 =42, Level1to3 =2, LevelZero =0, Level7to10 =0}
Socioeconomic level = Medium Low
SOCIOECONOMIC_LEVEL = MediumLow EXPEC_ENTER_COLLEGE = AlmostAll POSGRADE = MasterDegree CORRECT_HOMEWORK = Weekly USE_INTERNET_EDU_PURPOSES= Sometimes IMPROVE_ICT = HighLevel EXP_FUNCTION = >20years: Level4to6 {Level4to6 =18, Level1to3 =1, LevelZero =0, Level7to10 =0}
Socioeconomic level = Low
SOCIOECONOMIC_LEVEL = Low EXPEC_ENTER_COLLEGE = LittleLessThanHalf EXPEC_FINISH_HIGH_SCHOOL = AlmostAll EXP_FUNCTION = >20years DISCIPLINE = LP: Level1to3 {Level4to6 =0, Level1to3 =160, LevelZero =0, Level7to10=0} DISCIPLINE = MT: Level1to3 {Level4to6 =1, Level1to3 =151, LevelZero =0, Level7to10=0}
Socioeconomic level = Very Low
SOCIOECONOMIC_LEVEL = VeryLow GRADE = MathematicsDegree POSGRADE = Specialization180h: Level1to3 {Level4to6 =0, Level1to3 =311, LevelZero =5, Level7to10 =0}

Table 4. Fragments of the Decision Tree for Portuguese Language proficiency.

After an initial analysis, the results in Table 4 show that no occurrences were found for the “Level7to10” class, making the next “Level4to6” class the target. According to Table 4, the higher the socioeconomic level, the higher the present value for this class, which corroborates the findings of Gonçalves and França (2008) and Laros et al. (2012).

The socioeconomic level of a school can be a predominant factor for the students and teachers to have access to teaching and learning resources, as well as access to technological resources such as the internet and media tools. With this, they are able to develop teaching practices based on these resources, improving learning, and honing creativity in the teaching environment (Gustafsson et al., 2018).

Another important attribute, with a high predictive capacity for student performance, is the level of expectation of teachers in relation to the continuity of their students' studies, whether they are in higher education or finishing Secondary School. It is interesting to note that the level of expectation also accompanies the socioeconomic level, that is, the higher the socioeconomic level, the higher the level of expectation of teachers.

There are no values for the Low and Very Low socioeconomic levels for the "Level4to6" class; the highest level reached for this class is "Level1to3". It is observed that the level of expectation of teachers about their students concluding a higher education course also decreases, so that, for the low socioeconomic level, the first concern of teachers, observing a teaching time of more than 20 years, is that students complete Secondary School (Soares et al., 2010; Fonseca & Namen, 2016).

Teaching practices, for example, proposing homework and correcting students' tasks, are also relevant for the highest socioeconomic levels. According to the tree results, proposing homework and correcting these tasks weekly can improve students' school development. At this point, where school tasks become routine, family participation becomes paramount so that students can safely develop their tasks, avoiding levels of stress, tiredness, and demotivation (Zafani & Omote, 2016; Brito, 2017; Jovarini et al., 2018;).

The next step was to analyze the results for the Decision Tree for Mathematical proficiency. In the same manner, several tests were carried out and the best configuration found was with depth = 12. In this configuration, accuracy was 92.18% +/- 0.34%, Error Classification was 7.09% +/- 0.34% and the Kappa Index was 0.661 +/- 0.01, with the Confusion Matrix illustrated in Table 5.

True:	Level4to6	Level7to10	Level1to3	LevelZero
Level4to6:	4376	200	1666	40
Level7to10:	155	568	29	0
Level1to3:	1607	25	59628	862
LevelZero:	18	0	360	446

Table 5. Confusion Matrix for Decision Tree – Mathematics.

The fragments of the Decision Tree for Mathematics, Table 6, were analyzed following the same logic used to analyze the Decision Tree for Portuguese language proficiency.

Socioeconomic level = High
SOCIOECONOMIC_LEVEL= High IMPROVE_ICT = ThereIsNo EXP_FUNCTION = >20years USE_INTERNET_EDU_PURPOSES= AlwaysAlmostAlways PRINC_ATTEN_LEARNING_STUDENT= AlwaysAlmostAlways: Level7to10{Level4to6 =0, Level7to10=102 , Level1to3 =0, LevelZero =0}
Socioeconomic level = Medium High
SOCIOECONOMIC_LEVEL = MediumHigh EXPEC_ENTER_COLLEGE = AlmostAll POSGRADE = Specialization180h TEACHER DISSAT CAREER= No CORRECT_HOMEWORK = Diary IMPROVE_SPEC_CONTENT = LowLevel DISCIPLINE = MT PRINC_MOTIVATES_ME= Often: Level4to6 { Level4to6 =50 , Level7to10 =0, Level1to3 =1, LevelZero =0}
Socioeconomic level = Medium
SOCIOECONOMIC_LEVEL = Medium EXPEC_ENTER_COLLEGE = AlmostAll POSGRADE = DoctorateDegree DEV_GROUP_ACTIV = Monthly: Level4to6 { Level4to6 =26 , Level7to10 =0, Level1to3 =0, LevelZero =0} POSGRADE = MasterDegree IMPROV_CURR_NAT_PARAMETERS = ModerateLevel PROPOSE_HOMEWORK = Diary: Level4to6 { Level4to6 =28 , Level7to10 =0, Level1to3 =1, LevelZero =0}
Socioeconomic level = Medium Low
SOCIOECONOMIC_LEVEL = MediumLow EXPEC_ENTER_COLLEGE = AlmostAll POSGRADE = MasterDegree IMPROVE_EVALUAT_METH = ModerateLevel GRADE = PortugueseDegree IMPROVE_SPEC_CONTENT = LowLevel: Level4to6 { Level4to6 =39 , Level7to10 =0, Level1to3 =2, LevelZero =0}
Socioeconomic level = Low
SOCIOECONOMIC_LEVEL = Low EXPEC_ENTER_COLLEGE = Few DEV_GROUP_ACTIV = Weekly USE_BOOKS_EDU_PURPOSES = Never: Level1to3 {Level4to6 =0, Level7to10 =0, Level1to3 =147 , LevelZero =2} EXPEC_ENTER_COLLEGE = LittleLessThanHalf USE_INTERNET_EDU_PURPOSES= AlwaysAlmostAlways

POSGRADE = Specialization180h
IMPROVE_EVALUAT_METH = ModerateLevel: Level1to3 {Level4to6 =0, Level7to10 =0, Level1to3 =160, LevelZero =0}
Socioeconomic level = Very Low
SOCIOECONOMIC_LEVEL = VeryLow
GRADE = MathematicsDegree
IMPROVE_ICT = LowLevel: Level1to3 {Level4to6 =0, Level7to10 =0, Level1to3 =116, LevelZero =0}

Table 6. Fragments of the Decision Tree for Mathematical Proficiency.

Similarly, the school's socioeconomic level is also a crucial factor for school performance in mathematics. At the High level, there are values for the “Level7to10” class, with a total of 102 occurrences.

Another attribute also found in the Decision Tree for Mathematics and which follows the same logic as the Tree for the Portuguese Language, is the level of expectation of teachers about students attending higher education. The higher the socioeconomic level, the higher the expectations of teachers concerning their students.

The attributes related to the teaching activities and practices present in the Mathematics Decision Tree can reinforce that teachers who are more committed to proposing exercises and correcting them, including developing activities in groups with a certain frequency, could contribute to students' better performance. It is important to note that the teacher's postgraduate level also emerged as an important attribute associated with students' better performance.

The next phase of analysis was to observe how relevant each attribute can be to predict the target attributes, the Scale Level in Portuguese, and the Scale Level in Mathematics. This analysis was performed using the results of Random Forest.

4.2 Random Forest

The tests for the Random Forest algorithm were performed according to the same logic used to build Decision Trees. Several configurations were tested and Tables 7 and 8 illustrate the values for the first 10 attributes found based on the Weight by Importance of the Tree. In addition to these, the weight of impotence for the socioeconomic level attribute was also included, although it is not ranked in the top ten. This operator calculates the weights of each attribute observing the points of division of the trees, so that the attributes with the highest weight can be considered the most relevant.

In addition to the Weight for Tree Importance value, Precision, Error Classification and Kappa Index are also described. Table 7 illustrates the results for proficiency in Portuguese and Table 8 for Mathematics.

Portuguese Language						
N°	Bagging = 100		Bagging = 70		Bagging = 50	
1	EXP_FUNCTION	1,0	EXP_FUNCTION	1,0	EXP_FUNCTION	1,0
2	IMPROV_CURR_NAT_PARAMETERS	0,798	IMPROV_CURR_NAT_PARAMETERS	0,797	IMPROV_CURR_NAT_PARAMETERS	0,790
3	IMPROVE_SPEC_CONTENT	0,721	IMPROVE_SPEC_CONTENT	0,716	IMPROVE_SPEC_CONTENT	0,709
4	IMPROVE_EVALUAT_METH	0,660	IMPROVE_EVALUAT_METH	0,652	IMPROVE_EVALUAT_METH	0,647
5	GRADE	0,633	GRADE	0,643	GRADE	0,644
6	IMPROVE_ICT	0,613	IMPROVE_ICT	0,610	IMPROVE_ICT	0,601
7	GENDER	0,505	GENDER	0,509	GENDER	0,506
8	USE_BOOKS_EDU_PURPOSES	0,459	USE_BOOKS_EDU_PURPOSES	0,458	USE_BOOKS_EDU_PURPOSES	0,456
9	POSGRADE	0,414	POSGRADE	0,416	POSGRADE	0,414
10	USE_INTERNET_EDU_PURPOSES	0,408	USE_INTERNET_EDU_PURPOSES	0,400	USE_INTERNET_EDU_PURPOSES	0,405
18	SOCIOECONOMIC_LEVEL	0,134	EXPEC_ENTER_COLLEGE	0,122	SOCIOECONOMIC_LEVEL	0,127
accuracy: 96.86% +/- 0.13%		accuracy: 96.84% +/- 0.12%		accuracy: 96.82% +/- 0.14%		
classification_error: 3.14% +/- 0.13%		classification_error: 3.16% +/- 0.12%		classification_error: 3.18% +/- 0.14%		
kappa: 0.816 +/- 0.008		kappa: 0.816 +/- 0.008		kappa: 0.814 +/- 0.009		

Table 7. Results of Random Forest for the Portuguese Language, based on Weight by Tree Importance.

Mathematics						
N°	Bagging = 100		Bagging = 70		Bagging = 50	
1	EXP_FUNCTION	1,0	EXP_FUNCTION	1,0	EXP_FUNCTION	1,0
2	IMPROV_CURR_NAT_PARAMETERS	0,799	IMPROV_CURR_NAT_PARAMETERS	0,791	IMPROV_CURR_NAT_PARAMETERS	0,780
3	IMPROVE_SPEC_CONTENT	0,741	IMPROVE_SPEC_CONTENT	0,752	IMPROVE_SPEC_CONTENT	0,772
4	IMPROVE_EVALUAT_METH	0,682	IMPROVE_EVALUAT_METH	0,679	IMPROVE_EVALUAT_METH	0,689
5	GRADE	0,645	IMPROVE_ICT	0,653	IMPROVE_ICT	0,653
6	IMPROVE_ICT	0,642	GRADE	0,652	GRADE	0,642
7	GENDER	0,536	GENDER	0,521	GENDER	0,503
8	USE_BOOKS_EDU_PURPOSES	0,479	USE_BOOKS_EDU_PURPOSES	0,479	USE_BOOKS_EDU_PURPOSES	0,476
9	USE_INTERNET_EDU_PURPOSES	0,440	USE_INTERNET_EDU_PURPOSES	0,436	USE_INTERNET_EDU_PURPOSES	0,459
10	POSGRADE	0,416	POSGRADE	0,399	POSGRADE	0,387
17	SOCIOECONOMIC_LEVEL	0,149	SOCIOECONOMIC_LEVEL	0,149	SOCIOECONOMIC_LEVEL	0,149
accuracy: 96.26% +/- 0.17%		accuracy: 96.22% +/- 0.18%		accuracy: 96.24% +/- 0.18%		
classification_error: 3.74% +/- 0.17%		classification_error: 3.78% +/- 0.18%		classification_error: 3.76% +/- 0.18%		
kappa: 0.804 +/- 0.009		kappa: 0.802 +/- 0.010		kappa: 0.803 +/- 0.010		

Table 8. Results of Random Forest for Mathematics, based on Weight by Tree Importance.

Confusion Matrix Portuguese Language					Confusion Matrix Mathematics				
True	Level4to6	Level1to3	LevelZero	Level7to9	True	Level4to6	Level7to9	Level1to3	LevelZero
Level4to6	4601	171	2	2	Level4to6	4596	161	191	0
Level1to3	1199	62393	825	0	Level7to9	31	620	0	0
LevelZero	0	0	787	0	Level1to3	1529	12	61429	693
Level7to9	0	0	0	0	LevelZero	0	0	0	655
Bagging =100									

Table 9. Confusion matrices for Portuguese and Mathematics

The Random Forest is a classification method that uses decision trees as an output, and through the analysis of the set of variables classified by the model, it allows the user to predict, with greater precision, which attributes could influence the target attribute of the research (Bravo Sanzana et al., 2015). One of the most important properties in Random Forest is its ability to show the importance of the attribute to predict the target variable, calculated by the increase of the percentage of the prediction error, considering the exclusion of the attribute (Ahmed & Hikmat Sadiq, 2018). As such, the attributes with higher importance values contribute better to the correct prediction of the observed attribute.

As per the results in Tables 7 and 8, the first eight attributes with the highest values are identical for both the Portuguese and Mathematics proficiency. After this value, little variation was found, with:

- Teacher's Experience (EXP_FUNCTION)
- Need for Improvement in National Curriculum Parameters (IMPROV_CURR_NAT_PARAMETERS)
- Need for Improvement in Specific Content (IMPROVE_SPEC_CONTENT)
- Need for Improvement in Evaluation Methodology (IMPROVE_EVALUAT_METH)
- Type of Teacher Education (GRADE)
- Need for Improvement in Information and Communication Technology (IMPROVE_ICT)
- Gender of teacher (GENDER) and finally,
- if the Teacher uses books for Teaching purposes (USE_BOOKS_EDU_PURPOSES).

The tests for both proficiencies were performed using tree numbers (bagging) equal to 100, 70, and 50. Comparing the results, the test with 100 trees showed the better accuracy, which was 96.86% +/- 0, 13% for the Portuguese language proficiency and 92.26% +/- 0.17% for Mathematics, with a low error rating, which was 3.14% for the Portuguese. Moreover, the Kappa indices of 0.816 for Portuguese and 0.804 for mathematics were satisfactory.

Regarding the best evaluated attributes as predictors of school performance, it was found that the most important attribute is the teacher's experience (EXP_FUNCTION). It is worth mentioning that this attribute is also present in the results of the Decision Trees, in the first phase of this study. This attribute is also mentioned in the bibliography as relevant since teachers who have more experience in their role are more able to develop learning alternatives (Mesquita, 2018; Araújo et al., 2019).

Also, on the eight attributes selected as most valued in a Random Forest, four of them are directly related to the indication of the need to improve the teaching staff, they are IMPROV_CURR_NAT_PARAMETERS, IMPROVE ICT, IMPROVE_SPEC_CONTENT, and IMPROVE_EVALUAT_METH. Some of these attributes are also present in results from the Decision Trees, which reinforces these attributes' criticality as predictors of school performance.

The need for teacher training is also cited by authors who argue that the routine updating of teachers' skills and competences can provide a more creative and challenging teaching-learning environment (Symeonidis, 2019; Voinea, 2019; Nascimento & Gomes, 2020). It is interesting to note that these attributes appear more frequently in the Decision Tree, referring to Mathematical proficiency. The results showed that teachers indicated that there is no need for further training or there is a low level of need.

The teaching practices that emerged from the Random Forest that concern teachers using books or the internet for teaching purposes can be directly related to the experience of teachers.

In the current conditions, with the rapid growth of information technology applications for teaching, there is a diversity in the applications that reinforce the knowledge taught in the classroom. As such, the use of Information Technology has become an additional resource for interpretive and critical learning.

Teachers have already identified the need to improve this area (IMPROV_ICT), possibly because the use of new technologies in the teaching environment can enable the expansion and strengthening of cognitive possibilities during the process of knowledge development and creation (Budnyk, 2019; Lorenz et al., 2019; Roliak, 2019).

Indeed, offering better teaching-learning environments requires great investments, whether in the physical structure (classrooms, laboratories, computers, and the internet, etc.), or the training and development of the teaching staff (training and postgraduate courses such as master's and doctorate). Furthermore, the inclusion and development of the students and the community where they reside also demand specific actions (Fonseca, 2009; Fonseca & Namen, 2016; Limon, 2016; Brito, 2017; Calixto et al., 2017 ; Sebastian et al., 2019; Queiroz et al., 2020).

After analyzing the Decision Trees, it was found that the school's socioeconomic level is a predominant attribute for good school performance. This relationship has already been discussed in the literature, in analyzes focused on students' socioeconomic level, others on family income (Machebe et al., 2017; Castillo-Vergara et al., 2018; Queiroz et al., 2020).

The school's socioeconomic level relates to how these schools are classified based on the socioeconomic level of the students as a whole. The value reported by the government for each school is the average calculated from the socioeconomic levels of students who study at these schools. This value may reflect an important parameter relating to the students' purchasing power, and, with that, some conditions that can influence the students' performance, such as better access to computers, the internet, teaching resources, among others.

Firstly, in Decision Trees, this attribute (SOCIOECONOMIC_LEVEL) is evident as the main node. Secondly, in the Random Forest results, it was found that this attribute does not have a high importance value. This may have been due to the operator Weight by Tree Importance using the average of the information gain values for the trees to calculate the importance, while in the Decision Tree, the main node is chosen based on the information gain after calculating the entropy. Although this attribute was not evident in the results from the Random Forest, it is nevertheless of great importance for the quality of education, since students, teachers and schools with greater access to resources, can have a better learning gain, especially in developing countries (Silva et al., 2018; Damon et al., 2019).

Table 10 shows the results for the Weight by Information Gain of the test with 100 trees. This operator calculates the relevance of the attributes used based on the information gain and assigns weights to them, so the higher the weight, the more relevant this attribute can be considered.

The results of information gain in Table 10 show that, among the attributes that presented the most significant gain of information for each proficiency, the Socioeconomic Level is the attribute with the greatest value, converging to the Decision Trees results described in Tables 5 and 6.

Portuguese Language		Mathematics	
SOCIOECONOMIC_LEVEL	1,000	SOCIOECONOMIC_LEVEL	1,000
EXPEC_ENTER_COLLEGE	0,739	EXPEC_ENTER_COLLEGE	0,725
EXPEC_FINISH_HIGH_SCHOOL	0,139	EXPEC_FINISH_HIGH_SCHOOL	0,124
POSGRADE	0,115	POSGRADE	0,120

Table 10. Weight by Information Gain

The attributes directly related to the preparation and expectation of teachers, such as the level of expectation of teachers in relation to their students entering an undergraduate course, together with their educational background and the routine use of teaching activities, can provide a more favorable environment so that students obtain better results in large-scale assessments. These conditions have already been observed by Marturano & Pizato (2015), Pettersson et al., (2016), Brito (2017), Jovarini et al. (2018) and Schaik et al. (2018) and which are also evident in the Decision Tree and Random Forest results.

Also, the Random Forest pointed out that the teacher's experience, as well as the updates in skills and competences related to the teaching function, are strongly associated with the students' school performance. These conditions can be aligned with the studies by Hanushek & Woessmann (2017), Lauermann, (2017), Schildkamp et al. (2020) and Zakariya et al. (2020).

Both results from the Decision Trees and Random Forest could clarify how the teachers' profile attributes can be associated to the students' performance. It is important to note that the socioeconomic level of the school can be fundamental for students to perform better in large-scale assessments. In addition, issues involving teacher training, level of expectation about their students, teaching practices, and the

need for improvement can also contribute to developing a more conducive environment for learning.

Confirming previous studies, the decision trees revealed a clear association between the attributes socio-economic level and school performance. However, it didn't generate enough nodes to reveal other important attributes. To that extent, Random Forest proved to be a crucial tool to deepen data analysis. This ability is due to the method used, which creates many trees, each with few attributes chosen at random. Then, the results are aggregated, each tree "voting" to choose the attributes with greater predictive capacity. This technique reduces the error of classification of trees and according to Segal (2004), it reduces the impact of the correlation of variables.

Regarding the use of the algorithms, it is important to point out that both Decision Trees and Random Forest can predict performance in the educational area with an adequate level of resolution, and Random Forest can be considered safer because it has a more robust methodology.

5 FINAL CONSIDERATIONS

The objective of this study was to investigate which factors related to the teaching staff could be associated to the performance of Secondary School students in large-scale assessments of the 2017 Basic Education Assessment System in Brazil. Decision and Random Forests were the techniques used as well as factors already discussed in the literature, such as the Socioeconomic Level. The results show that this attribute plays a fundamental role in influencing students' performance.

However, the isolated importance of teacher training that appears in the Random Forest technique, overcoming other factors considered to be more important, contributes to reflect on the importance of school management activities that can improve teachers' skills and give them more tools to adjust teaching practices when deemed necessary.

Teachers are key players in educational environments, as they can encourage students to advance in their development. The improvement of their skills in the classroom can come from postgraduate courses as well as from short courses of improvement, such as ICT courses, mentioned among the attributes that impact students' performance. This kind of training is a clear example of a short course that can give teachers an arsenal of tools to be used when they think these tools will lead

to better results. When motivated and committed, and if they join forces with the community and families, teachers can pave the way for young people to seek better opportunities, enabling social inequality to reduce.

The fact that the Decision Tree-based algorithms used in this study were able to predict, with a good level of confidence, the attributes that best are associated to the performance of students on large-scale assessments is an important contribution. Based on the results presented, educational managers can develop training plans, motivational training and other measures needed for the learning environment to become more efficient and motivating.

Only with a well-designed educational plan, involving all segments of education such as government, school, teachers, students and the community, can the level of quality in education evolve positively. Quality education should not be seen as a benefit to the students involved but an ambition of the society, as stated in the UN's Sustainable Development Goals in the fourth objective, Quality Education: Ensuring inclusive and equitable and quality education, and promoting learning opportunities throughout life for everyone.

The development of this study can contribute to the literature related to educational data mining, in the case of the use of the Random Forest algorithm, as already suggested by Martínez-Abad (2000), since few publications using this algorithm were found.

A limitation for the analysis of school assessment data on a large scale is that many teachers and schools do not complete their forms correctly, thus reducing the universe of analysis. A suggestion for further studies would be analyzing future editions of the large-scale assessment, when relieved. As another suggestion, an analysis of municipal schools involving the level of investment and the results of large-scale assessments could provide a snapshot of the efficiency of the use of investments in education.

FUNDING SOURCES

Funding: This work was supported by the Coordination of Improvement of Higher Level Personnel - Brazil (CAPES) - Financing Code 001

REFERENCES

- ALDRUP, K. et al. Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. **Journal of Educational Psychology**, v. 110, n. 8, p. 1066–1083, 2018.
- ALVES, T.; SILVA, R. M. DA. Estratificação das oportunidades educacionais no Brasil: contextos e desafios para a oferta de ensino em condições de qualidade para todos. **Educação & Sociedade**, v. 34, n. 124, p. 851–879, set. 2013.
- AMÂNCIO-VIEIRA, S. F. et al. A Relação entre Custo Direto e Desempenho Escolar: Uma Análise Multivariada nas Escolas de Ensino Fundamental de Londrina/PR. **Educação em Revista**, v. 31, n. 1, p. 169–194, mar. 2015.
- AMÂNCIO-VIEIRA, S. F.; PEREIRA, R. S. Custos diretos, nível socioeconômico, estrutura escolar e desempenho educacional: uma análise das escolas municipais de São José dos Pinhais/PR. **Anais do Congresso Brasileiro de Custos - ABC**, v. 0, 2016.
- AMARAL, N. C. Os 10% do PIB como promotor da qualidade da educação: uma análise considerando os resultados do PISA e os valores aplicados por estudante em diversos países. **Rede Nacional de Ciência para a Educação (Rede CpE)**, 2015.
- ANDRADE, J. M. DE; LAROS, J. A. Fatores associados ao desempenho escolar: estudo multinível com dados do SAEB/2001. **Psicologia: Teoria e Pesquisa**, v. 23, n. 1, p. 33–41, mar. 2007.
- ANDRADE, M.; FRANCO, C.; DE CARVALHO, J. P. Gênero e Desempenho em Matemática ao final do Ensino Médio: Quais as relações? **Anais**, p. 1–16, 2016.
- BACK, L. T. et al. Classroom management, school staff relations, school climate, and academic achievement: testing a model with urban high schools. **Learning Environments Research**, v. 19, n. 3, p. 397–410, 1 out. 2016.
- BAKER, R. S. J. D; YACEF, K. The State of Educational Data Mining in 2009: A Review and Future Visions. **JEDM | Journal of Educational Data Mining**, v. 1, n. 1, p. 3–17, 1 out. 2009.
- BAKER, R. S.; INVENTADO, P. S. Educational Data Mining and Learning Analytics. In: LARUSSON, J. A.; WHITE, B. (Eds.). . **Learning Analytics: From Research to Practice**. New York, NY: Springer, 2014. p. 61–75.
- BARBOSA, S. G.; WILHELM, V. E. Evaluation of the performance of public schools through Data Envelopment Analysis/Avaliação do desempenho das escolas públicas por meio de Data Envelopment Analysis. **Acta Scientiarum. Technology**, 2009.
- BARROS, Á. G. DE et al. A RELEVÂNCIA DA FORMAÇÃO CONTINUADA PARA OS GESTORES ESCOLARES DA REDE PÚBLICA. **LINKSCIENCEPLACE - Interdisciplinary Scientific Journal**, v. 6, n. 3, 6 nov. 2019.

BEZERRA, C. et al. Evasão Escolar: Aplicando Mineração de Dados para Identificar Variáveis Relevantes. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 27, n. 1, p. 1096, 7 nov. 2016.

BEZERRA, L. N. M.; SILVA, M. T. Educational Data Mining Applied to a Massive Course. **International Journal of Distance Education Technologies (IJDET)**, v. 18, n. 4, 1 out. 2020.

BEZERRA, L. N. MENDES. **Mineração de Dados Educacionais para a gestão de cursos massivos**. Tese de Doutorado—São Paulo: Universidade Paulista - UNIP, 2017.

BEZERRA, M. G.; KASSOUF, A. L. **Análise de fatores que afetam o desempenho escolar nas escolas das áreas urbanas e rurais do Brasil**. XLIV CONGRESSO DA SOBER - Questões agrárias, educação no campo e desenvolvimento. **Anais...** In: XLIV CONGRESSO DA SOBER. Fortaleza-CE: Sociedade Brasileira de Economia e Sociologia Rural, 2006Disponível em: <<https://bdpi.usp.br/item/001536729>>. Acesso em: 28 set. 2019

BOUZADA, M. A. C.; RIBEIRO, L. O. M.; PEIXE, J. B. **Métodos Quantitativos Aplicados a Casos Reais - 1st Edition**. Rio de Janeiro: Elsevier Editora Ltda, 2013.

BRASIL. **Lei nº 9.394, de 20 de dezembro de 1996, Estabelece as diretrizes e bases da educação nacional**. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L9394.htm>.

BRASIL. **Lei nº 13005, de 25 de julho de 2014. Estabelece o Plano nacional de educação - PNE**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l13005.htm>. Acesso em: 28 abr. 2019.

BRASIL. **Tesouro divulga o relatório Aspectos Fiscais da Educação no Brasil**. Tesouro Nacional, , 2018. Disponível em: <<http://www.tesouro.fazenda.gov.br/-/tesouro-divulga-o-relatorio-aspectos-fiscais-da-educacao-no-brasil>>. Acesso em: 5 out. 2019

BRITO, M. V. S. DE. **Determinantes do sucesso escolar : uma análise multinível a partir dos dados do Pisa 2015**. Dissertação de Mestrado—Brasília: Universidade de Brasília, 23 jun. 2017.

BROOKE, N.; SOARES, J. F. Pesquisa em eficácia escolar: origem e trajetória. **Est. Aval. Educ**, v. 22, p. 593–598, 2011.

CALIXTO, K.; SEGUNDO, C.; GUSMÃO, R. P. DE. Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 28, n. 1, p. 1447, 27 out. 2017.

CAMPOS, H. R.; JORGE, S. D. C. Violência na escola: uma reflexão sobre o bullying e a prática educativa. **Em Aberto**, v. 23, n. 83, p. 22, 2010.

CÁRIA, N. P.; LAMBERT-DE-ANDRADE, N. Democratic Management at school: in search of participation and leadership. **Revista Eletrônica de Educação**, v. 10, n. 3, p. 25–40, 30 nov. 2016.

CARVALHO, L. A. V. D. **Datamining - A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Rio de Janeiro: Editora Ciência Moderna, 2005.

CASTILLO-VERGARA, M. et al. Does socioeconomic status influence student creativity? **Thinking Skills and Creativity**, v. 29, p. 142–152, 1 set. 2018.

CASTRO, J. A.; CARVALHO, C. H. A. Necessidades e possibilidades para o financiamento da educação brasileira no plano nacional de educação. **Educação & Sociedade**, v. 34, n. 124, p. 829–849, set. 2013.

CASTRO, L. N.; FERRARI, D. G. **Introdução a mineração de dados**. 1ª edição ed. [s.l.] SARAIVA, 2017.

CHIRINÉA, A. M.; BARREIRO, I. M. F. Qualidade da educação: eficiência, eficácia e produtividade escolar. **Revista on line de Política e Gestão Educacional**, v. 0, n. 7, 2009.

CILIATO, F. L. G.; SARTORI, J. Pluralidade cultural: os desafios aos professores em frente da diversidade cultural. **Revista Monografias Ambientais**, v. 14, n. 0, p. 65–78, 7 dez. 2015.

COLPANI, R. Mineração de Dados Educacionais: um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar. **Informática na educação: teoria & prática**, v. 21, n. 3, 2018.

CORREIA-ZANINI, M. R. G. **Um estudo prospectivo sobre o percurso escolar de crianças nos primeiros anos do Ensino Fundamental**. Tese de Doutorado—São Paulo: Universidade de São Paulo, 6 dez. 2013.

CUNHA, J. M. P. DA et al. Social segregation and academic achievement in state-run elementary schools in the municipality of Campinas, Brazil. **Geoforum**, Themed Issue: Land, Labor, Livestock and (Neo)Liberalism: Understanding the Geographies of Pastoralism and Ranching. v. 40, n. 5, p. 873–883, 1 set. 2009.

DALLAZEM, A.; COELHO, V. R. O Desempenho Escolar na voz dos atores de escolas públicas catarinenses. **Revista Teias**, v. 20, n. 56, p. 398–417, 26 mar. 2019.

DAMON, A. et al. What education policies and programmes affect learning and time in school in developing countries? A review of evaluations from 1990 to 2014. **Review of Education**, v. 7, n. 2, p. 295–387, 2019.

DÍAZ, L. M. B.; GARCÍA, M. L. S.; CANO, E. V. Effects on academic performance in secondary students according to the use of ICT. **IJERI: International Journal of Educational Research and Innovation**, n. 12, p. 90–108, 14 jun. 2019.

DOURADO, L. F.; OLIVEIRA, J. F.; SANTOS, C. DE A. A Qualidade da Educação: conceitos e definições. **Textos para Discussão**, 2007.

DRABACH, N. P. **Processos de provimento, exigências e atribuições para os dirigentes escolares em contextos de reformas gerencialistas**. Tese de Doutorado—Campinas-São Paulo: UNICAMP, 2018.

DWYER, T. et al. Desvendando mitos: os computadores e o desempenho no sistema escolar. **Educação & Sociedade**, v. 28, n. 101, p. 1303–1328, dez. 2007.

FANTINATO, A. C.; CIA, F. Envolvimento parental, competência social e o desempenho acadêmico de escolares. **Psicologia Argumento**, v. 29, n. 67, 24 nov. 2017.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Advances in Knowledge Discovery and Data Mining. In: FAYYAD, U. M. et al. (Eds.). . Menlo Park, CA, USA: American Association for Artificial Intelligence, 1996. p. 1–34.

FELÍCIO, F. DE; FERNANDES, R. O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental no Estado de São Paulo. **Anais - ANPEC**, 2005.

FERNANDES, L. DE M. et al. Preditores do Desempenho Escolar ao final do Ensino Fundamental: Histórico de Reprovação, Habilidades Sociais e Apoio Social. **Trends in Psychology**, v. 26, n. 1, p. 215–228, mar. 2018.

FERNANDES, R. Índice de Desenvolvimento da Educação Básica (Ideb). **Textos para Discussão**, p. 29, 2007.

FILHO, N. M.; NUÑEZ, D. F.; RIBEIRO, F. P. Comparando as escolas paulistas com melhor e pior desempenho no SARESP e na Prova Brasil. **São Paulo em Perspectiva**, v. 23, n. 1, p. 20, 2009.

FINI, M. INÊS. Currículo e Avaliação, articulação necessária em favor da aprendizagem dos alunos da rede pública de São Paulo. **Revista São Paulo em Perspectiva - Fundação Seade**, 2009.

FONSECA, S. O. DA; NAMEN, A. A. Mineração em Bases de Dados do INEP: Uma Análise Exploratória para Nortear Melhorias no Sistema Educacional Brasileiro. **Educação em Revista**, v. 32, n. 1, p. 133–157, mar. 2016.

FORD, T. G. Can the use of informal control mechanisms increase trust among teachers? An evaluation of the Accelerated Schools intervention. **Studies in Educational Evaluation**, v. 63, p. 59–71, 1 dez. 2019.

FRANCO, A. M. P.; MENEZES FILHO, N. Uma análise de rankings de escolas brasileiras com dados do SAEB. **Estudos Econômicos (São Paulo)**, v. 42, n. 2, p. 263–283, jun. 2012.

FREITAS, E. D. S. DE. **Práticas de gestão do conhecimento como variáveis intervenientes nos resultados do IDEB**. Dissertação de Mestrado—Maringá-PR: Centro Universitário de Maringá - UNICESUMAR, 2019.

GARCIA, P. S. et al. A infraestrutura das escolas de ensino fundamental da Região do Grande ABC paulista. **Revista Ibero-Americana de Estudos em Educação**, v. 9, n. 3, p. 614–631, 2014.

- GARCIA, P. S. et al. Desempenho escolar: uma análise do IDEB dos municípios da região do ABC. **Revista Eletrônica de Educação**, v. 10, n. 2, p. 95–114, 31 ago. 2016.
- GATTI, B. Avaliação de sistemas educacionais no Brasil. **Sísifo. Revista de Ciências da Educação**, p. 7–18, 2009.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2010.
- GIVISIEZ, G. H. N.; OLIVEIRA, E. L. DE. Royalties do petróleo e educação: análise da eficiência da alocação. **Revista Brasileira de Pós-Graduação**, v. 8, n. 0, p. 1–13, 2011.
- GÓIS, L. S.; ROCHA, G. S. A atuação do Gestor Escolar no enfrentamento da evasão no Instituto Federal da Bahia(IFBA). **Revista Eletrônica Científica Ensino Interdisciplinar**, v. 5, n. 14, p. 340–361, 7 out. 2019.
- GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data mining: conceitos, técnicas, algoritmos, orientações e aplicações**. 2ª ed. Rio de Janeiro: Elsevier, 2015.
- GONÇALVES, F. DE O.; FRANÇA, M. T. A. Transmissão intergeracional de desigualdade e qualidade educacional: avaliando o sistema educacional brasileiro a partir do SAEB 2003. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 16, n. 61, p. 639–662, dez. 2008.
- GUSTAFSSON, J.-E.; NILSEN, T.; HANSEN, K. Y. School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. **Studies in Educational Evaluation**, Research on Equity in Education: Implications for School Evaluation. v. 57, p. 16–30, 1 jun. 2018.
- HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)**. 3ª ed. [s.l.] Morgan Kaufmann, 2012.
- HANUSHEK, E. A.; WOESSMANN, L. School Resources and Student Achievement: A Review of Cross-Country Economic Research. In: ROSÉN, M.; YANG HANSEN, K.; WOLFF, U. (Eds.). **Cognitive Abilities and Educational Outcomes: A Festschrift in Honour of Jan-Eric Gustafsson**. de Metodologia de Medição e Avaliação Educacional. Cham: Springer International Publishing, 2017. p. 149–171.
- HOLFMANN, M.; KLINKENBERG, R. **RapidMiner: Data Mining Use Cases and Business Analytics Applications**. 1ª ed. London: Chapman and Hall/CRC, 2013.
- INEP. **IDEB - SAEB**. Disponível em: <<http://portal.inep.gov.br/web/guest/educacao-basica/saeb>>. Acesso em: 29 mar. 2020a.
- INEP. **Censo Escolar**. Disponível em: <<http://portal.inep.gov.br/web/guest/censo-escolar>>. Acesso em: 29 abr. 2020b.
- INEP. **SAEB - Microdados**. Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 29 mar. 2020c.

- INEP. **PNE - Monitoramento das Metas**. Disponível em: <http://simec.mec.gov.br/pde/grafico_pne.php>. Acesso em: 29 mar. 2020d.
- INEP. **Indicadores Financeiros Educacionais**. Disponível em: <<http://inep.gov.br/indicadores-financeiros-educacionais>>. Acesso em: 29 mar. 2020e.
- INEP. **Pisa em Foco**. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/pisa-em-foco>>. Acesso em: 14 nov. 2020f.
- INEP. **PISA**. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa>>. Acesso em: 29 mar. 2020g.
- INEP. **Dados Abertos - Educação Brasileira**. Disponível em: <<http://portal.inep.gov.br/web/guest/dados>>. Acesso em: 14 nov. 2020h.
- INEP. **Indicadores Educacionais**. Disponível em: <<http://portal.inep.gov.br/indicadores-educacionais>>. Acesso em: 29 mar. 2020i.
- INEP. **Sinopses Estatísticas - INEP**. Disponível em: <<http://portal.inep.gov.br/web/guest/sinopses-estatisticas>>. Acesso em: 14 nov. 2020j.
- INEP. **IDEB**. Disponível em: <<http://portal.inep.gov.br/web/guest/ideb>>.
- JOVARINI, N. V.; LEME, V. B. R.; CORREIA-ZANINI, M. R. G. Influência das Habilidades Sociais e Estressores sobre o Desempenho Escolar no 6º ano. **Paidéia (Ribeirão Preto)**, v. 28, 2018.
- JÚNIOR, O. DE G. F. et al. Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de Maceió. **RENOTE**, v. 17, n. 1, p. 296–305, 28 jul. 2019.
- LACRUZ, A. J. et al. Indicadores de qualidade na educação: análise discriminante dos desempenhos na Prova Brasil. **Revista Brasileira de Educação**, v. 24, 2019.
- LAKATOS, E. M.; MARCONI, M. DE A. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2017.
- LAROS, J. A.; MARCIANO, J. L.; ANDRADE, J. M. DE. Fatores associados ao desempenho escolar em Português: um estudo multinível por regiões. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 20, n. 77, p. 623–646, dez. 2012.
- LAROS, J. A.; MARCIANO, J. L. P.; ANDRADE, J. M. DE. Fatores que afetam o desempenho na prova de Matemática do SAEB: um estudo multinível. **Aval. psicol.**, v. 9, p. 14, 2010.
- LASATER, K.; BENGTON, E.; ALBILADI, W. S. Data use for equity?: How data practices incite deficit thinking in schools. **Studies in Educational Evaluation**, p. 100845, 15 fev. 2020.

LAUERMANN, F. Teacher motivation, responsibility, pedagogical knowledge and professionalism: a new era for research. **Pedagogical Knowledge and the Changing Nature of the Teaching Profession**, p. 171–191, 21 fev. 2017.

LOBO, G. D. Avaliação do Desempenho Escolar dos Estudantes da Região Nordeste que Realizaram o ENEM: Uma Análise com Modelos Hierárquicos. **Revista Espacios**, v. 38, p. 13, 2017.

LUZ, L. S. **Os Determinantes do Desempenho Escolar: A Estratificação Educacional e o Efeito Valor Adicionado**. . In: XV ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, ABEP. Caxambu-MG: 2006Disponível em: <<http://www.abep.org.br/publicacoes/index.php/anais/article/view/1693>>

MACHEBE, C. H.; EZEGBE, B. N.; ONUOHA, J. C. The Impact of Parental Level of Income on Students' Academic Performance in High School in Japan. **Universal Journal of Educational Research**, v. 5, p. 1614–1620, 2017.

MARANGONI, R. A. O Trabalho do Diretor de Escola: Análise a partir de uma perspectiva histórica. **Jornal de Políticas Educacionais**, v. 14, n. 0, 26 abr. 2020.

MARTURANO, E. M.; PIZATO, E. C. G. Preditores de Desempenho Escolar no 5º Ano do Ensino Fundamental. **PSICO**, v. 46, 2015.

MASCI, C.; DE WITTE, K.; AGASISTI, T. The influence of school size, principal characteristics and school management practices on educational performance: An efficiency analysis of Italian students attending middle schools. **Socio-Economic Planning Sciences**, Recent developments on the use of DEA in the public sector. v. 61, p. 52–69, 1 mar. 2018.

MEC. **PNE - Plano Nacional de Educação - Plano Nacional de Educação - Lei nº 13.005/2014**. Disponível em: <<http://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>>.

MESQUITA, S. S. DE A. Referenciais do “bom professor” de ensino médio: exercício de articulação teórica. **Caderno Pesquisa**, v. 48, 2018.

MINAYO, M. C. DE S. (OR). **Pesquisa social: teoria, método e criatividade**. 29. ed. Petrópolis/RJ: Vozes, 2010.

MOHR, A. C.; NAUJORKS, M. I. Políticas Educacionais de correção de fluxo no RS: Efeitos da racionalidade neoliberal na Gestão Escolar. **Reflexão e Ação**, v. 25, n. 1, p. 160–177, 2 jan. 2017.

MORAIS, J. DE S.; NASCIMENTO, F.-L. S. C.; MAGALHÃES, N. R. S. Evocações narrativas de gestores escolares acerca das políticas de saberes e fazeres da gestão. **Educação, Ciência e Cultura**, v. 25, n. 2, p. 221–237, 5 ago. 2020.

MURPHY, S. School location and socioeconomic status and patterns of participation and achievement in senior secondary mathematics. **Mathematics Education Research Journal**, v. 31, n. 3, p. 219–235, 1 set. 2019.

NASCIMENTO, M. C. DO; GOMES, G. R. R. Formação continuada docente para a utilização das TIC no processo de ensino e aprendizagem. **Research, Society and Development**, v. 9, n. 2, p. e33921998–e33921998, 1 jan. 2020.

NASCIMENTO, V. P.; CHIUSOLI, C. L. O papel do gestor escolar: estudo de caso sobre os desafios da educação pública. **Revista on line de Política e Gestão Educacional**, v. 23, n. 2, p. 238–254, 6 mar. 2019.

NICHELE, P. T.; MELLO, M. A. S. Gestão Escolar na perspectiva da educação democrático-participativa e a função social da escola. **Revista Saberes Pedagógicos**, v. 4, n. 3, p. 323–343, 29 set. 2020.

OCDE. **PISA - PISA**. Disponível em: <<http://www.oecd.org/pisa/>>. Acesso em: 29 set. 2018.

OLIVEIRA, A. C. P. DE; WALDHELM, A. P. S. Liderança do diretor, clima escolar e desempenho dos alunos: qual a relação? **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 24, n. 93, p. 824–844, dez. 2016.

ONU. **Agenda 2030**. Disponível em: <<https://nacoesunidas.org/pos2015/agenda2030/>>.

ONU. **Plataforma Agenda 2030**. Disponível em: <<http://www.agenda2030.com.br/>>. Acesso em: 28 set. 2019.

ORTIGÃO, M. I. R. Avaliação e Políticas Públicas: possibilidades e desafios para a Educação Matemática. **BOLEMA: Boletim de Educação Matemática**, 2008.

PAULA, J. S.; FRANCO, A. M. P.; SILVA, J. W. Fatores relacionados ao atraso escolar no estado de Minas Gerais. **Estudos em Avaliação Educacional**, v. 29, n. 72, p. 886–917, 20 dez. 2018.

PEÑA-AYALA, A. (ED.). **Learning Analytics: Fundaments, Applications, and Trends : A View of the Current State of the Art to Enhance e-Learning: 94**. 1. ed. New York, NY: Springer, 2017.

PEREIRA, I. A. Tipos de liderança dentro do ambiente educativo: uma discussão sobre sua identidade organizacional. **Educationis**, v. 8, n. 1, p. 8–22, 28 fev. 2020.

PEREIRA, M. J.; MORI, N. N. R. Diretrizes curriculares e o desempenho de alunos paranaenses da 8ª série do ensino fundamental na Prova Brasil. **RBPG. Revista Brasileira de Pós-Graduação**, v. 8, p. p121- 145. 25p., 2012.

PEREIRA, P. R.; GASQUE, K. C. G. D. Tomada de decisão do gestor escolar das escolas públicas de ensino médio no Distrito Federal e a interface com o letramento informacional. **Em Questão**, v. 25, n. 3, p. 40–60, 7 ago. 2019.

PHILIPPSEN, M. R. **Avaliação externa do desempenho escolar e seus impactos na prática pedagógica e na gestão escolar: estudo de caso em uma escola pública de Porto Alegre (RS)**. Monografia—Rio Grande do Sul: Universidade Federal de Santa Maria, 1 dez. 2018.

PINTO, J. M. R. Remuneração adequada do professor: desafio à educação brasileira. **Retratos da Escola**, v. 3, n. 4, 6 maio 2012.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. 1. ed. California-EUA: O'Reilly Media, 2013.

PROVOST, F.; FAWCETT, T. **Data Science Para Negócios**. 1ª Edição ed. Rio de Janeiro: Editora Alta Books, 2016.

QUEIROZ, M. V. A. B.; SAMPAIO, R. M. B.; SAMPAIO, L. M. B. Dynamic efficiency of primary education in Brazil: Socioeconomic and infrastructure influence on school performance. **Socio-Economic Planning Sciences**, v. 70, p. 100738, 1 jun. 2020.

RAZAK, N. A. et al. Successful implementation of information and communication technology integration in Malaysian public schools: An activity systems analysis approach. **Studies in Educational Evaluation**, v. 58, p. 17–29, 1 set. 2018.

REZENDE PINTO, J. M. DE; AMARAL, N. C.; CASTRO, J. A. DE. O financiamento do Ensino Médio no Brasil: de uma escola boa para poucos à massificação barata da rede pública. **Educação & Sociedade**, v. 32, n. 116, p. 639–665, set. 2011.

RIBEIRO, M. D.; SÍVERES, L.; BRITO, R. O. A Formação de gestores escolares: a dimensão ética em questão. **Educação Por Escrito**, v. 10, n. 2, p. e36685–e36685, 2019.

RIDDLE, S.; APPLE, M. W. **Re-imagining Education for Democracy**. 1. ed. London: Routledge, 2019.

RODRIGUES, C. G.; RIOS-NETO, E. L. G.; PINTO, C. C. DE X. Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico, 1997 a 2005. **Revista Brasileira de Estudos de População**, v. 28, n. 1, p. 5–36, jun. 2011.

ROMERO, C.; VENTURA, S. Data mining in education. **WIREs Data Mining and Knowledge Discovery**, v. 3, n. 1, p. 12–27, 2013.

SALAM, S. et al. ICT and students' performance in Pakistan. **Human Systems Management**, v. 36, n. 4, p. 277–284, 1 jan. 2017.

SALES, F. et al. Evasão no Ensino Básico da Rede Pública Municipal de Juiz de Fora: uma Análise com Mineração de Dados. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 30, n. 1, p. 1371, 11 nov. 2019.

SANTOS, A. F. DOS. Os Questionários Contextuais do SAEB 2013 e do PISA 2012: uma comparação dos elementos constituintes do Ensino. **Reuniões da ABAVE**, v. 0, n. 8, p. 453–486, 30 out. 2015.

SANTOS, R. S. DOS; MONTANUCI, R.; PACHECO, E. J. Educação de qualidade e avaliação: Raízes históricas de uma correlação contemporânea. **Cadernos da FUCAMP**, v. 18, n. 35, 29 ago. 2019.

- SCHAIK, P. VAN et al. Barriers and conditions for teachers' utilisation of academic knowledge. **International Journal of Educational Research**, v. 90, p. 50–63, 1 jan. 2018.
- SCHILDKAMP, K. et al. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. **International Journal of Educational Research**, v. 103, p. 101602, 1 jan. 2020.
- SCHNEIDER, G.; GOUVEIA, A. B. Qualidade da escola: uma proposta de índice para as condições materiais de escolas a partir de dados contextuais do Saeb. **Revista Brasileira de Pós-Graduação**, v. 8, n. 1, 31 dez. 2011.
- SEBANC, A. M.; GUIMOND, A. B.; LUTGEN, J. Transactional Relationships Between Latinos' Friendship Quality and Academic Achievement During the Transition to Middle School. **The Journal of Early Adolescence**, v. 36, n. 1, p. 108–138, 1 jan. 2016.
- SILVA, L. A. DA; PERES, S. M.; BOSCARIOLI, C. **Introdução á Mineração de Dados**. Rio de Janeiro: Elsevier, 2016.
- SILVA, R. J. DA. **Mineração de dados aplicada a análise de desempenho de alunos no 5º ano do ensino fundamental**. Monografia—Serra Talhada - PE: Universidade Federal Rural de Pernambuco, 2019.
- SIMIELLI, L. E. R. Equidade e Oportunidades Educacionais: O Acesso a Professores no Brasil. **Arquivos Analíticos de Políticas Educativas**, v. 25, 2017.
- SIMÕES, M. DE F.; FERRÃO, M. E. Competência percebida e desempenho escolar em Matemática. **Estudos em Avaliação Educacional**, v. 16, n. 32, p. 25–42, 2005.
- SIMON, A.; CAZELLA, S. Mineração de Dados Educacionais nos Resultados do ENEM de 2015. **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**, v. 6, n. 1, p. 754, 27 out. 2017.
- SOARES, C. S.; ROSA, F. S. DA. Gastos públicos em educação e desempenho escolar: Análise qualitativa dos resultados do IDEB em Santa Maria/RS. **Revista Eletrônica de Estratégia & Negócios**, v. 13, n. 0, p. 38–62, 8 maio 2020.
- SOARES, J. F.; ALVES, M. T. G. Desigualdades raciais no sistema brasileiro de educação básica. **Educação e Pesquisa**, v. 29, n. 1, p. 147–165, jun. 2003.
- SOARES, T. M. et al. A expectativa do professor e o desempenho dos alunos. **Psicologia: Teoria e Pesquisa**, v. 26, n. 1, p. 157–170, mar. 2010.
- SOUZA, L. D. M.; RIBEIRO, M. S. DE S. O Perfil do Gestor Escolar Contemporâneo: das permanências as incorporações para exercício da função. **Revista Espaço do Currículo**, v. 10, n. 1, 28 abr. 2017.
- SOUZA, M. I. M. O fazer do gestor escolar: desafios e possibilidades de sua atuação profissional, enquanto facilitador do processo de ensino-aprendizagem. **Research, Society and Development**, v. 9, n. 7, p. e335973900–e335973900, 14 maio 2020.

SYMEONIDIS, V. Teacher competence frameworks in Hungary: A case study on the continuum of teacher learning. **European Journal of Education**, v. 54, n. 3, p. 400–412, 2019.

TARDA, K. B.; RODRIGUES, R. V. A Influência dos Gastos em Educação no Crescimento Econômico da Região Administrativa de Campinas. **Revista Estudo & Debate**, v. 22, n. 2, 23 dez. 2015.

TRAVITZKI, R.; FERRÃO, M. E.; COUTO, A. P. Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: Uma visão a partir da análise de dados do ENEM. **Education Policy Analysis Archives**, v. 24, p. 74, 11 jul. 2016.

VASCONCELOS, J. C. et al. School infrastructure and public investment in Education in Brazil: the importance of educational performance. **Ensaio: Avaliação e Políticas Públicas em Educação**, n. AHEAD, 2020.

VIDAL, E. M.; VIEIRA, S. L. Professores da educação básica: perfil e percepções sobre sucesso dos alunos. **Estudos em Avaliação Educacional**, v. 28, p. 64–101, 2017.

VIEIRA, A. E. R.; BUSSOLOTI, J. M. Gestão Escolar: um estudo de caso sobre Escolas Técnicas. **Interação - Revista de Ensino, Pesquisa e Extensão**, v. 20, n. 1, p. 45–70, 2018.

VOINEA, M. Rethinking Teacher Training According to 21st Century Competences. **European Journal of Multidisciplinary Studies**, v. 4, n. 3, p. 20–26, 25 set. 2019.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3ª ed. [s.l.] Morgan Kaufmann, 2011.

ZAFANI, M. D.; OMOTE, S. Atribuições Da Família Na Percepção Do Professor. **Journal of Research in Special Educational Needs**, v. 16, n. S1, p. 221–224, 2016.

ZAKARIYA, Y. F.; BJØRKESTØL, K.; NILSEN, H. K. Teacher job satisfaction across 38 countries and economies: An alignment optimization approach to a cross-cultural mean comparison. **International Journal of Educational Research**, v. 101, p. 101573, 1 jan. 2020.

ZANBELLO, B. L. et al. **Administração Escolar: do Diretor ao Gestor**. Anais Eletrônico do XI EPCC - Encontro Internacional de Produção Científica. **Anais...** In: ENCONTRO INTERNACIONAL DE PRODUÇÃO CIENTÍFICA UNICESUMAR. Maringá-PR: 29 out. 2019 Disponível em: <<http://rdu.unicesumar.edu.br/handle/123456789/3752>>. Acesso em: 3 set. 2020

6.5 ARTIGO 5. Fatores preditivos de desempenho escolar em avaliações do SAEB: influência da Gestão Escolar.

Predictive factors of school performance in SAEB assessments: influence of School Management.

Factores predictivos del desempeño escolar en las evaluaciones SAEB: influencia de la Gestión Escolar.

Ivonaldo Vicente da Silva

Márcia Terra da Silva

Nilsa Duarte da Silva Lima

Resumo

As atividades desenvolvidas no ambiente escolar, pelo Diretor de Escola, podem contribuir para que a instituição de ensino alcance níveis diferenciados em termos de qualidade educacional, e esse diferencial pode ser fruto do nível de experiência do Diretor. Dessa forma, o objetivo desse artigo foi observar quais fatores referente às atividades dos Diretores de Escola puderam contribuir para o desempenho dos alunos do Ensino Médio, nas avaliações em larga escala do SAEB no ano de 2017. Após a coleta dos dados no site do governo, foram executadas as tarefas relacionadas a seleção e tratamento desses dados e logo após, a fase de Mineração de Dados Educacionais, que contemplou a execução dos algoritmos Árvore de Decisão, Random Forest e Gradient Boosted Tree. Os resultados encontrados pela fase de Mineração de Dados indicaram que tanto para Língua Portuguesa quanto para Matemática, o Tempo de Experiência do Diretor de Escola, no exercício da função, pode contribuir para que os alunos do Ensino Médio obtivessem melhores desempenhos nas avaliações do SAEB, além desse fator outros como nível de pós-graduação, ações para controle de reprovações e Nível Socioeconômico da escola também contribuíram para o desempenho escolar dos alunos. Um estudo mais detalhado dos fatores influenciadores do desempenho pode contribuir para elaboração de projetos, que envolvam políticas públicas, e que promovam melhorias em todo ambiente escolar, bem como nos desempenhos dos alunos.

Palavras Chave: Gestão Escolar; Desempenho Escolar; Mineração de Dados.

Abstract

The activities developed in the school environment, by the School Director, can contribute for the educational institution to reach different levels in terms of educational quality, and this differential can be the result of the level of experience of the Director. Thus, the objective of this article was to observe which factors referring to the activities of the School Directors could contribute to the performance of high school students, in the large scale assessments of SAEB in 2017. After data collection on the government website , the tasks related to the selection and treatment of these data were carried out and soon after, the Educational Data Mining phase, which included the execution of the Decision Tree, Random Forest and Gradient Boosted Tree algorithms. The results found by the Data Mining phase indicated that for both Portuguese and Mathematics, the Time of Experience of the School Director, in the exercise of the function, can contribute for the high school students to obtain better performances in the SAEB evaluations, in addition to this factor, others such as postgraduate level, actions to control failures and the socioeconomic level of the school also contributed to the students' academic performance. A more detailed study of the factors influencing performance can contribute to the elaboration of projects, which involve public policies, and which promote improvements in the entire school environment, as well as in the performance of students.

Keywords: School Management; School performance; Data Mining

Resumen

Las actividades realizadas en el ámbito escolar, por parte del Director de la Escuela, pueden contribuir a que la institución educativa alcance distintos niveles en cuanto a calidad educativa, y este diferencial puede ser resultado del nivel de experiencia del Director. Así, el objetivo de este artículo fue observar qué factores referidos a las actividades de los Directores Escolares podrían contribuir al desempeño de los estudiantes de secundaria, en las evaluaciones a gran escala del SAEB en 2017. Después de la recolección de datos en el sitio web del gobierno, se llevaron a cabo las tareas relacionadas con la selección y tratamiento de estos datos y poco después, la fase de Minería de Datos Educativos, que incluyó la ejecución de los algoritmos Decision Tree, Random Forest y Gradient Boosted Tree. Los resultados encontrados por la fase de Data Mining indicaron que tanto para Portugués como para Matemáticas, el Tiempo de Experiencia del Director de la Escuela, en el ejercicio de la función, puede contribuir para que los estudiantes de secundaria obtengan mejores desempeños en las evaluaciones SAEB, Además de este factor, otros como el nivel de posgrado, las acciones de control de fallas y el nivel socioeconómico de la escuela también contribuyeron al rendimiento académico de los estudiantes. Un estudio más detallado de los factores que influyen en el desempeño puede contribuir a la elaboración de proyectos, que involucran políticas públicas, y que promueven mejoras en todo el entorno escolar, así como en el desempeño de los estudiantes.

Palabras clave: Gestión escolar; El rendimiento escolar; Procesamiento de datos

1. Introdução

O sistema de ensino brasileiro apresenta necessidades específicas para atender a população com relação a uma educação de qualidade, assegurada pela Constituição Brasileira e pela Lei de Diretrizes e Bases da Educação-LDB (BRASIL, 1988, 1996). Diversas são as carências na educação, desde a estrutura física e acesso a recursos tecnológicos até formação e capacitação do corpo docente e administrativo (Brito, 2017; Calixto et al., 2017; Mesquita, 2018; Vasconcelos et al., 2020).

No Brasil, o ciclo básico da educação é composto pelos cursos do Ensino Fundamental em nove anos e o Ensino Médio, normalmente ofertado em três anos. Além disso, algumas escolas ainda oferecem cursos da Educação Infantil, Cursos Técnicos e os cursos da Educação de Jovens e Adultos-EJA, antigo supletivo. A oferta desses cursos depende de estratégias desenvolvidas pelas Secretárias de Educação de cada região.

Para garantir o funcionamento das escolas públicas é necessário que todos os recursos financeiros estejam previstos em orçamentos, que são elaborados pelo governo. Já as escolas particulares, os recursos necessários para a manutenção das atividades são garantidos pelo pagamento das mensalidades. A maior oferta de vagas em escolas no Brasil é feita pela rede pública. Para se ter uma noção, em 2017, do total de 48.608.093 de matrículas efetuadas no Sistema de Educação Básica brasileiro, apenas 8.887.061 foram efetuadas em escolas particulares (INEP, 2019).

No ambiente escolar, normalmente o organograma apresenta como maior nível hierárquico, a função de Diretor de Escola, logo após aparecem as funções administrativas, pedagógicas e de manutenção. Portanto, o Diretor é responsável por conduzir a escola, e por meio de suas ações, dar suporte necessário para que todos os departamentos funcionem perfeitamente, evitando que as atividades pedagógicas sejam prejudicadas.

A quantidade de alunos que cada escola pode suportar também se torna um grande desafio para os Diretores, pois quanto maior a oferta de vagas, maior é a necessidade de recursos. Para acompanhar a eficiência da escola em relação ao modelo de gestão, incluindo a oferta de vagas, o governo criou o Índice de Complexidade de Gestão Escolar. Esse índice possui seis níveis de complexidade

que envolvem a análise sobre quantidade de alunos, cursos e turnos que cada escola disponibiliza para sua comunidade.

Quanto maior o nível de complexidade maior é o porte da escola, como exemplo de escola classificada como Nível 6 em complexidade de gestão, ela possui porte superior a 500 matrículas, oferta cursos em 3 turnos com 4 ou mais etapas relacionadas às idades dos alunos, e têm o curso Educação de Jovens e Adultos-EJA, como etapa mais elevada.

Ainda em relação a complexidade de gestão, escolas com porte para muitos alunos também absorve mais recursos, sejam eles financeiros, de material e de pessoal. A merenda escolar, em escolas públicas, já é uma demanda importantíssima para a gestão escolar, pois existem regiões em que os alunos precisam dessa merenda para complementar sua alimentação diária (Santos et al., 2018; Fernandes et al., 2020; Ramos et al., 2020;).

Já no ambiente escolar, cabe aos Diretores de Escola e aos Professores coordenarem as atividades pedagógicas, cada um na sua área de atuação, de forma a garantir a perfeita execução do Plano Pedagógico proposto para a escola (MEC, 2014).

Em uma escola, com base na gestão democrática, as tarefas e atividades que o Diretor deve realizar são semelhantes a qualquer outro tipo de empresa, pois incluem metas de desempenho, gestão de equipes, gestão de conflitos, controle orçamentário e gerenciamento do relacionamento com todos os envolvidos no ambiente escolar (Cária & Lambert-de-Andrade, 2016; Riddle & Apple, 2019). Dentre as atividades inerentes ao Diretor de Escola, que foram citadas, algumas tem como foco o desempenho escolar dos alunos.

O desempenho dos alunos é um fator importante para que o governo e a sociedade possam acompanhar e avaliar se o modelo utilizado para o sistema de ensino está alcançando os objetivos propostos, em termos de qualidade. No Brasil, o Governo Federal criou um índice específico para acompanhar o desenvolvimento da qualidade da educação, que é o Índice de Desenvolvimento da Educação Básica - IDEB. A proposta do governo é que o valor desse índice, numa escala que varia entre zero e dez, alcance 6 (seis pontos) até o ano de 2021 (INEP, 2020a).

Sabendo dessa condição, os Diretores utilizam suas habilidades e competências para conduzir suas equipes em prol da melhoria do ensino aprendizagem, com foco no desempenho dos alunos. Algumas ações comuns no

ambiente escolar, podem ser consideradas como atividades rotineiras da gestão escolar, como exemplo são os treinamentos e capacitações das equipes, reuniões rotineiras para debater propostas de melhorias envolvendo as equipes pedagógicas e administrativa, reuniões com conselho de escola para tratar de assuntos relacionados ao modelo de gestão e planos pedagógicos, reuniões com representantes do governo para debater assuntos sobre a eficiência escolar e finalmente reuniões com representantes das comunidades para tratar de projetos regionais (Back et al., 2016; Aldrup et al., 2018; Masci et al., 2018).

Alguns estudos já publicados citam o estilo de liderança do Diretor de Escola como importante influenciador no desempenho da equipe, da melhoria do clima organizacional referente ao ambiente escolar, de motivação do corpo docente e administrativo, e de melhoria no nível de relacionamento com pais e responsáveis pelos alunos e comunidade (Nascimento & Chiusoli, 2019; Nichele & Mello, 2020; Pereira, 2020)

Diversas são as atividades pertinentes à função de Diretor de Escola, e a análise de como a execução dessas tarefas podem contribuir para a melhoria do ensino e do desempenho escolar, pode trazer à luz, necessidades de aperfeiçoamentos na maneira de como essas tarefas estão sendo conduzidas. Uma fonte de dados importante para esse tipo de análise são os questionários do SAEB, pois envolvem perguntas sobre o cotidiano da gestão escolar.

Dessa forma, com base nas atividades dos gestores escolares e amparado pelos resultados do SAEB, esse estudo tem como objetivo central observar quais fatores referentes às atividades dos Diretores de Escola puderam contribuir para o desempenho dos alunos do Ensino Médio, nas avaliações em larga escala do SAEB no ano de 2017.

Para o desenvolvimento dessa pesquisa, logo após a introdução será apresentado o capítulo de Metodologia, que abordou as publicações que fundamentaram o estudo, bem como o método utilizado para o desenvolvimento da pesquisa. Em seguida, os Resultados são apresentados para debate e reflexões e finalmente o capítulo de considerações finais.

2. Metodologia

A metodologia para o desenvolvimento desse estudo está estruturada em Revisão da Literatura, que teve como objetivo observar publicações realizadas sobre o tema dessa pesquisa e complementarmente, Desenvolvimento do Método da Pesquisa, que envolveu a fase de coleta, tratamento e Mineração de Dados Educacionais.

2.1 Revisão da Literatura

As questões relacionadas com os desafios enfrentados pelos Diretores no cotidiano escolar, já foram observadas em pesquisas na área da educação. Fatores como modelos de gestão, dificuldades de acesso a vários tipos de recursos, formação e capacitação da equipe e apoio a comunidade, são exemplos de abordagens encontradas em objetivos de pesquisas já publicadas.

O estudo de Souza & Ribeiro (2017) procurou observar questões teóricas e práticas em relação as atividades do gestor escolar de escolas públicas, procurando delinear um perfil para o gestor contemporâneo, baseado entre a sua formação e os modelos de gestão já existentes no ambiente escolar. Os resultados indicaram que o modelo de gestão vigente nas escolas públicas exige que o gestor seja um profissional multitarefa, além disso, os autores observaram que falta incentivo financeiro e de valorização profissional da função.

Com intuito de observar se as políticas de provimento do cargo de direção da escola, em conjunto com o estilo de liderança do diretor, influenciam o desempenho dos alunos do 5º do Ensino Fundamental na proficiência matemática, Oliveira et al. (2018) analisaram os resultados das edições de 2007, 2009 e 2011 do SAEB e concluíram que existe uma associação positiva entre o estilo de liderança dos diretores com o desempenho dos alunos em matemática. Além disso, os resultados da pesquisa encontraram uma associação negativa entre o desempenho dos alunos em matemática com o estilo de liderança de Diretores que tiveram acesso ao cargo por meio de nomeação.

Com o intuito de observar quais fatores contribuem para a evasão escolar em alunos do Ensino Médio, Colpani (2018) utilizou a técnica de Mineração de Dados Educacionais e a regressão linear nos dados relativos ao Censo Escolar de 2017, e concluiu que o fator que mais contribui para a evasão escolar é taxa de distorção

idade-série, indicando com isso que os alunos que estão acima da idade tendem a não concluir o Ensino Médio. Cabe então ao gestor, observar no ambiente escolar alunos que apresentam sinais que possam indicar dificuldades para prosseguir com os seus estudos, seja por falta de motivação ou ainda pela necessidade de exercer uma atividade laboral, para com isso, desenvolver ações preventivas para esse tipo de evasão,

Amparados pela Mineração de Dados Educacionais e do algoritmo Naive Bayes, Fonseca et al. (2018) investigaram os resultados da Prova Brasil do ano de 2013, do estado do Rio de Janeiro, na tentativa de encontrar os fatores que contribuíram para o desempenho dos alunos na proficiência matemática. Como resultados os autores indicaram que 61,43% dos alunos com notas altas nas avaliações estão em escolas que informaram não haver problemas significantes para o exercício das atividades. Além disso, os resultados demonstraram que nessas escolas existe um baixo índice de absenteísmo do corpo docente, fato esse que pode estar atrelado diretamente com o modelo de gestão escolar.

Também com o uso da técnica de Mineração de Dados Educacionais, Júnior et al., (2019) analisaram a evolução do Índice de Desenvolvimento da Educação Básica (IDE) das escolas públicas do estado de Maceió. O intuito da pesquisa foi auxiliar os gestores no processo de tomada de decisão para adoção de melhorias das ações que envolvem a gestão escolar. Os resultados obtidos pelo método de regressão linear e Árvore de Decisão, indicaram que os fatores com maior capacidade de influenciar o desempenho dos alunos são o nível de escolaridade dos pais do aluno, o nível de incentivo para os estudos e o nível de compromisso do corpo docente.

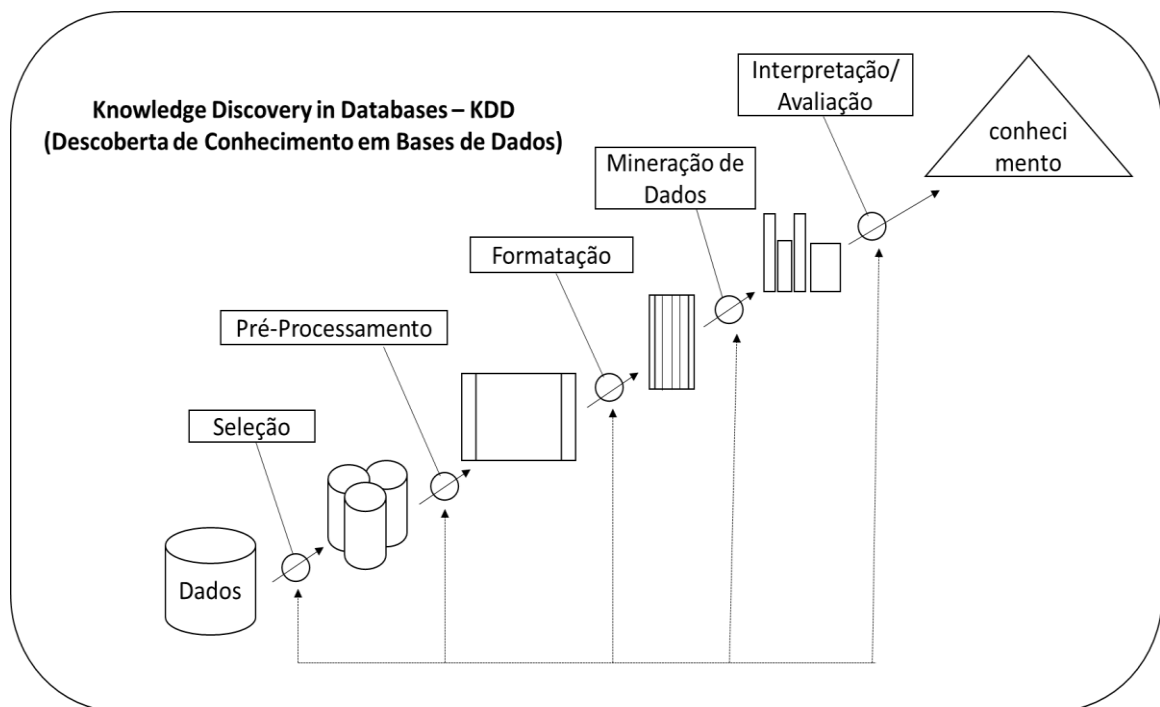
Além desses exemplos citados, outras publicações que utilizaram Mineração de Dados Educacionais foram coletadas e os seus resultados puderam contribuir de forma significativa para o desenvolvimento desse estudo.

2.1.2 Técnica de Mineração de Dados Educacionais

Uma das alternativas para análise de grande volume de dados é a Mineração de Dados Educacionais, pois utiliza algoritmos e aprendizado de máquina, para identificar padrões de comportamento que poderão contribuir no processo de tomada de decisão (Provost & Fawcett, 2016).

A Mineração de Dados faz parte do processo denominado Knowledge Discovery in Databases – KDD, (Descoberta de Conhecimento em Bases de Dados), e pode ser entendido como um processo de várias etapas para identificar padrões em grande conjunto de dados (Fayyad et al., 1996; Goldschmidt et al., 2015; Bezerra & Silva, 2020). A Figura 1 ilustra o processo de KDD.

Figura 1. Processo de Descoberta de Conhecimento em Bases de Dados



Fonte: Adaptado de Fayyad et al. (1996).

De acordo com a Figura 1, as tarefas que envolvem Seleção, Pré-Processamento e Formatação dos dados devem ser realizadas antes do processo de Mineração de Dados. Essa sequência se torna necessária por que os dados quando coletados na origem, não estão preparados para a fase de mineração, então, é primordial que os dados brancos, nulos ou preenchidos erroneamente sejam retirados da seleção ou tratados por técnicas apropriadas de mineração. Além disso, e se for preciso, esses dados devem ser formatados para proporcionar um melhor entendimento na fase de análise.

Finalizadas essas etapas, o conjunto de dados devidamente preparado poderá ser utilizado para a Mineração de Dados, e nessa fase, inicialmente deve-se

definir qual ou quais algoritmos serão utilizados. *Naive Bayes*, *Decision Tree*, *Random Forest* e *Gradient Boosted Tree*, Redes Neurais, Regressão, são exemplos de algoritmos utilizados em estudos que envolvem predição de comportamento, perfis de consumo, análise de crédito e risco, inclusive análise de desempenhos (Romero & Ventura, 2013; Provost & Fawcett, 2016; Puyalnithi et al., 2016).

Para esse estudo foram utilizados os algoritmos Naive Bayes, Decision Tree, Random Forest e Gradient Boosted Tree, para observar se os fatores referentes às atividades dos Diretores de Escola puderam contribuir para o desempenho dos alunos do Ensino Médio, nas avaliações em larga escala do SAEB no ano de 2017.

2.2 Desenvolvimento do Método do Estudo

Como fase inicial desse estudo, foi necessário observar publicações relacionadas ao tema, com objetivo de identificar fatores do ambiente escolar e relacionados com a função de gestão, que já foram analisados por meio da Mineração de Dados. O resultado dessa fase contribuiu para o desenvolvimento dessa pesquisa, no sentido que proporcionou um rol de fatores considerados relevantes para o exercício da função de Diretor de Escola. A partir desse rol de fatores o processo de coleta de dados pode ser definido e planejado.

Como etapa seguinte, foram coletados os dados da avaliação em larga escala do Sistema de Avaliação da Educação Básica-SAEB, da edição de 2017, no site do governo. No pacote de dados publicados pelo governo estavam inclusos os resultados dos alunos nas avaliações, além dos questionários preenchidos pelos professores, pelos Diretores e Representantes de Escola (INEP, 2020b).

O banco de dados completo referente as avaliações do SAEB, da edição de 2017, continha 1.966.507 de alunos do Ensino Médio, 73.674 de Diretores de Escola, 753.668 de Professores e finalmente 73.674 de registros provenientes de Escolas. Importante salientar que não foi possível identificar os respondentes, pois esses não eram identificados por nome, recebem apenas um código identificador referente a escola (IDESCOLA). A partir dos dados coletados a próxima fase foi a de pré-processamento dos dados.

2.2.1 Fase de pré-processamento dos dados

Esse estudo teve como foco os questionários preenchidos pelos Diretores e pelas Escolas que ofertaram o curso Ensino Médio em 2017. Para compor a base de

dados inicial para esse estudo, foram relacionados os questionários dos Diretores com as respectivas escolas onde atuavam. Como resultado dessa tarefa, foi criada uma base de dados com o nome de “DiretoresEscola”, que serviu para a análise dos dados e conseqüentemente para a fase de Mineração de Dados.

Com a base de dados já separada, o próximo passo foi separar os atributos necessários para essa pesquisa, com base na literatura observada, em conjunto com os formulários disponibilizados no pacote de dados coletados do site do governo. Os formulários direcionados aos Diretores de Escola possuíam 111 questões enquanto que para as Escolas, os formulários possuíam 74 questões. Do conjunto de questões disponíveis desses formulários foram separados 27 atributos com base nos fatores mencionados na bibliografia. A lista de fatores e questões separadas dos formulários está descrita no Quadro 1.

Quadro 1. Relação de Fatores presentes nos questionários dos Diretores e das Escolas.

Atributos	Questões dos formulários
Questões relacionadas aos Formulários aplicados aos Diretores	
FORMAÇÃO E EXPERIÊNCIA NA FUNÇÃO	
POSGRADUACAO	8. Indique o curso de pós-graduação de mais alta titulação que você possui.
EXP_FUNCAO	16. Há quantos anos você exerce funções de direção?
ORG_TREINAMENTO	26. Nos últimos dois anos, você organizou alguma atividade de formação continuada (atualização, treinamento, capacitação, etc.) nesta escola?
CARACTERÍSTICAS DA EQUIPE ESCOLAR	
CONSELHO_ESCOLA	29. O Conselho Escolar é um colegiado geralmente constituído por representantes da escola e da comunidade que tem como objetivo acompanhar as atividades escolares. Neste ano, quantas vezes se reuniu o conselho escolar?
CONSELHO_CLASSE	31. O Conselho de Classe é um órgão formado por todos os professores que lecionam em cada turma/série. Neste ano e nesta escola, quanta vezes se reuniu o Conselho de Classe?
ABANDONO_ESCOLAR	41. Nesta escola, há alguma ação para redução das Taxas de Abandono?
REPROVACAO_ESCOLAR	42. Nesta escola, há alguma ação para redução das Taxas de Reprovação?
MELHOR_ENSINO_ALUNOS	44. Nesta escola, indique com que frequência você discute com os professores medidas com o objetivo de melhorar o ensino e a aprendizagem dos alunos.
INDIQUE COM QUAL FREQUÊNCIA SÃO DESENVOLVIDAS AS SEGUINTE	

ATIVIDADES PARA MINIMIZAR AS FALTAS DOS ALUNOS NESTE ANO E NESTA ESCOLA:	
FALTAS_REUNIAO_PAIS	47. Os pais/responsáveis são chamados à escola para conversar sobre o assunto em reunião de pais.
FALTA_INDIVIDUAL_PAIS	48. Os pais/responsáveis são chamados à escola para conversar sobre o assunto individualmente.
EM RELAÇÃO À MERENDA ESCOLAR, COMO VOCÊ AVALIA OS SEGUINTE ASPECTOS:	
MERENDA_REC_FINANCEIROS	62. Recursos financeiros.
MERENDA_QUANT_ALIMENTOS	63. Quantidade de alimentos.
MERENDA_QUALI_ALIMENTOS	64. Qualidade de alimentos.
O FUNCIONAMENTO DA ESCOLA FOI DIFICULTADO POR ALGUM DOS SEGUINTE PROBLEMAS?	
FUNC_ESCOLA_REC_FINANCEIROS	67. Insuficiência de recursos financeiros.
FUNC_ESCOLA_INEX_PROF	68. Inexistência de professores para algumas disciplinas ou séries.
FUNC_ESCOLA_FALTA_ADM	69. Carência de pessoal administrativo.
FUNC_ESCOLA_FALTA_PED	70. Carência de pessoal de apoio pedagógico (supervisor, coordenador, orientador educacional).
FUNC_ESCOLA_FALTA_REC_PED	71. Falta de recursos pedagógicos.
FUNC_ESCOLA_INTER_ATIV	72. Interrupção das atividades escolares.
FUNC_ESCOLA_FALTA_PROF	73. Alto índice de faltas por parte dos professores.
FUNC_ESCOLA_FALTA_ALUNOS	74. Alto índice de faltas por parte dos alunos.
FUNC_ESCOLA_ROTAT_PROF	75. Alta rotatividade do corpo docente.
FUNC_ESCOLA_INDISC_ALUNOS	76. Indisciplina por parte dos alunos.
CONSIDERE AS CONDIÇÕES EXISTENTES PARA O EXERCÍCIO DO CARGO DE DIRETOR NESTA ESCOLA:	
EXER_CARGO_INTERF_EXTERNA	77. Há interferência de atores externos em sua gestão?
EXER_CARGO_APOIO_SUPERIORES	78. Há apoio de instâncias superiores?
CARACTERÍSTICAS DA ESCOLA – FORMULÁRIO APLICADO ÀS ESCOLAS	
NIVEL_SOCIOECONOMICO	Nível Socioeconômico da escola
MEDIA_3EM_LP	Valor da Média atribuída pelo governo para Língua Portuguesa
MEDIA_3EM_MT	Valor da Média atribuída pelo governo para Matemática.
Escala_LP	Nível da escala de desempenho referente a Língua Portuguesa
Escala_MT	Nível da escala de desempenho referente a Matemática.

Fonte: Elaborado pelos Autores (2020).

Com a base de dados “DiretoresEscola” pronta, o próximo passo foi a tarefa de tratamento desses dados, ou seja, todos os dados presentes nessa base de dados foram analisados e os dados brancos, nulos ou preenchidos erroneamente

foram eliminados. Essa tarefa foi primordial para que não ocorressem erros de interpretação e análise, na fase de Mineração de Dados. Após realizada essa tarefa, do total de 56.614 registros disponíveis na base de dados, só puderam ser utilizados 14.006.

Como próxima fase, foi realizada a formatação dos dados que consistiu na substituição das respostas presentes nos registros, por uma palavra ou sigla que proporcionasse maior ganho de informação, na fase de análise. Com isso, e utilizando o atributo “PosGraduacao” como exemplo, onde estava presente a letra “A”, foi inserido o valor correspondente ao formulário, para esse caso a palavra utilizada foi “NaoFez”.

Em relação ao desempenho da escola, ou seja, as médias atribuídas para Língua Portuguesa e Matemática, foram criados dois Atributos denominados “Escala_LP” e “Escala_MT”. Para observar o nível de desempenho da escola utilizou-se a escala de desempenho atribuída aos alunos do Ensino Médio, dessa forma, a partir das médias atribuídas para as escolas, pode se ter noção sobre em que nível da escala de desempenho a escola está classificada.

Em relação a escala de desempenho dos alunos do Ensino Médio, existe uma variação entre 0 e 500 pontos, sendo que o Nível 1 da escala inicia em 225 pontos e vai progredindo a cada 25 pontos formando intervalos regulares. Assim, o Nível 1 compreende os desempenhos entre 225 e 250 pontos, o Nível 2 de 250 a 275 pontos e assim sucessivamente. A avaliação pressupõe que os alunos que não obtiveram 225 pontos não conseguiram apresentar ainda as habilidades elementares e essenciais compatíveis com sua etapa escolar.

A partir das médias das escolas e das escalas de desempenho dos alunos, foi possível efetuar uma análise preliminar sobre como foi classificado o nível de aprendizado dos alunos dessas escolas. Ressaltando que a média atribuída pelo governo para a escola, foi calculada a partir dos resultados do conjunto de alunos daquela escola, em cada proficiência analisada.

Com isso, os atributos “Escala_LP” e “Escala_MT” foram preenchidos com o nível da escala de desempenho, observando o valor da média escolar de cada proficiência. Apenas para efeitos ilustrativos, para os valores de médias entre 225 e 250 pontos foi atribuído a classificação “Nível 1” no atributo “Escala_LP”, para o intervalo entre 250 e 275 foi atribuído o “Nível 2” e assim sucessivamente, tanto para Língua Portuguesa (Escala_LP) quanto para Matemática (Escala_MT).

Como última fase no processo de tratamento dos dados, foi necessário observar o volume de dados e se esses apresentavam dados isolados, considerados como outliers. Após efetuadas as análises necessárias, foram separados os dados que apresentavam valores de médias iguais ou superiores a 225 pontos e inferiores a 350, respectivamente os níveis entre 1 e 5 da escala de desempenho.

Não foram utilizados os dados com médias inferiores a 225 pontos por que essas escolas apresentaram resultados compatíveis com o Nível Zero da escala de desempenho. Já os dados iguais ou superiores a 350 pontos não foram utilizados por que representavam aproximadamente 0,5% de registros em Língua Portuguesa e 2,2% de registros em Matemática, podendo com isso interferir nos resultados apresentados no processo de Mineração de Dados.

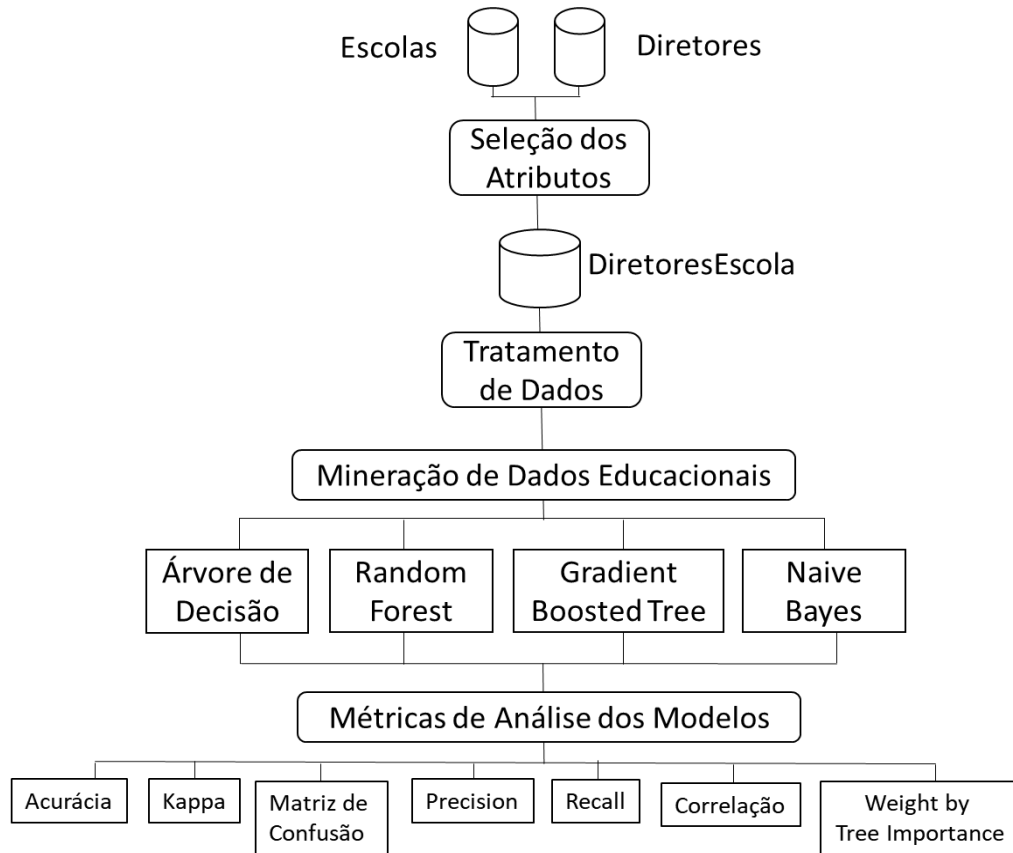
Após realizado esse filtro, o total de registros utilizados para a fase de Mineração de Dados foi de 13.549 para a proficiência Língua Portuguesa e de 13.381, para Matemática, correspondendo respectivamente a 96,7% e 95,5% do total de registros separados (14.006).

Finalmente com a base devidamente preparada a próxima fase consistiu no processo de Mineração de Dados e os resultados alcançados foram analisados com base na bibliografia inicialmente observada.

2.2.2 Fase de Mineração de Dados Educacionais

Inicialmente se torna necessário observar como o processo de Mineração foi conduzido, portanto o fluxo de tarefas foi desenvolvido e o resultado está ilustrado na Figura 2.

Figura 2. Etapas para o desenvolvimento da fase de Mineração de Dados



Fonte: Desenvolvido pelos Autores (2020).

Com o fluxo de tarefas desenvolvido o próximo passo foi executar os algoritmos de mineração, para isso foi utilizado o software RapidMiner versão estudante. Os algoritmos escolhidos, conforme já descrito, foram Naive Bayes, Árvore de Decisão, Random Forest e Gradient Boosted Tree.

O algoritmo Naive Bayes é considerado um classificador probabilístico baseado no Teorema de Bayes criado pelo matemático Thomas Bayes, que tem como lógica a probabilidade condicional, ou seja, qual a probabilidade do evento X ocorrer dado o evento Y. A fórmula que esse algoritmo utiliza para prever os resultados é:

$$p\left(\frac{A}{B}\right) = P\left(\frac{B}{A}\right) P(A)/P(B)$$

O cálculo da densidade de probabilidade, utilizada na representação dos resultados desse algoritmo, é a soma das probabilidades de cada classe multiplicada por sua função densidade de probabilidade, sua fórmula é:

$$P(x) = \sum_i P(C_i)P\left(\frac{X}{C_i}\right) \quad (2)$$

Onde:

$P(x)$: densidade de probabilidade

$P(C_i)$: probabilidade a priori da classe C_i

$P(X/C_i)$: função densidade de probabilidade da classe C_i

As Árvores de Decisão são normalmente utilizadas em modelos de classificação e regressão, e podem ser baseadas em algoritmos como J48, C.45, ID3, CART, *Random Forest*, *Gradient Boosted Tree*. Outro ponto em comum em Árvores de Decisão é que podem ser supervisionadas ou não. Quando utilizado o método supervisionado, é necessário que exista um atributo alvo de predição para o desenvolvimento das árvores, e nesse modelo, o algoritmo se encarrega de encontrar os melhores pontos de poda para montar a Árvore, com os atributos que apresentaram os melhores ganhos de informação.

A partir do conjunto de dados o algoritmo utilizado para Árvore de Decisão executa o cálculo com base na Entropia, para encontrar o melhor ponto de corte para os nós. Após efetuados todos os cálculos é indicado como nó principal aquele que apresentou maior ganho de informação, e a partir desse nó todas as outras ramificações ou folhas são acrescentadas. Para calcular o ganho de informação é necessário escolher um dos critérios que são: Accuracy, Information_gain, Gain_ratio, Least_Square. Dessa forma pode-se entender que o atributo mais importante, com base na Entropia e no Ganho de Informação, está no topo da árvore. A fórmula utilizada para o cálculo da Entropia é:

$$Entropy(Y) = \sum_i p(c/Y) \log_2 p(c/Y)$$

2)

Onde Y é o conjunto de casos e $p(c/Y)$ o valor de Y é a proporção da classe c .

Para esse estudo foi escolhido para o cálculo de ganho de informação o critério “Gain ratio”, que ajusta o valor do critério “Information_gain” para cada atributo, permitindo uniformidade e amplitude dos valores do atributo. A fórmula utilizada para cálculo do ganho de informação pelo critério Gain_ratio é:

$$\text{Gain Ratio} = \text{New Ratio} - \text{Old Ratio}$$

3)

As Árvores de Decisão Random Forest e Gradient Boosted Tree são consideradas como algoritmos mais robustos e os seus resultados normalmente são por meio de um conjunto de árvores, que podem ser analisadas com base nos fatores de maior ganho de informação ou de importância da árvore.

As *Random Forest* ou Floresta de Árvores são consideradas um método de aprendizado em conjunto (Ensemble). O algoritmo utiliza para criar a floresta de árvores o método de *Bootstrap Agregating*, ou *Bagging*, que gera um conjunto de dados por amostragem *bootstrap* da base de dados utilizado no processo, e o conjunto de resultados pode proporcionar um melhor resultado geral. São criadas várias pequenas árvores com baixo desempenho, mas que quando analisadas em conjunto se tornam uma boa alternativa para predição de atributos, pois utiliza a média entre as árvores para indicar os atributos de maior importância.

Já o *Gradient Boosted Tree* também utiliza a lógica de aprendizado em conjunto, podendo ser comparado com o *Random Forest* e em alguns aspectos até pode ser semelhante, mas esse algoritmo utiliza como critério para a criação do conjunto de árvores, a otimização da função de perda, com isso sempre que ele cria uma nova árvore ele tenta corrigir os erros da árvore anterior no conjunto, utilizando para isso a taxa de aprendizagem.

O classificador bayesiano, bem como as Árvores de Decisão, pode fornecer informações e conhecimento suficientes para auxiliar o processo de tomada de decisão por parte dos gestores escolares. Nesse estudo as tarefas de Mineração de Dados Educacionais foram amparadas pelo software RapidMiner. E os modelos de classificação foram conduzidos pelo método supervisionado, ou seja, os algoritmos foram executados utilizando um atributo alvo que são as escalas de desempenho, “Escala_LP” ou “Escala_MT”.

Para executar os algoritmos, os dados foram separados por meio de Split de dados na proporção de 70/30, além disso foi utilizado o operador Cross Validation que separa o conjunto de dados em 10 partes iguais. Desse total, o operador separa uma parte para efetuar os testes e as demais partes, para treinamento do modelo. O uso desse operador pode produzir resultados com boa estimativa de desempenho em conjunto de dados.

Na execução dos algoritmos Árvore de Decisão, Random Forest e Gradient Boosted Tree foram realizados diversos testes utilizando combinações de quantidade de ramificações, e o valor que apresentou melhor desempenho para os três algoritmos foi com o total de 10 ramificações, dessa forma o critério depth ficou igual a 10.

No Random forest e Gradient Boosted Tree também foram testadas várias combinações para quantidade de árvores, e a combinação que apresentou melhor ganho de informação foi com o total de 100 árvores. Os critérios utilizados para cada algoritmo estão descritos no Quadro 2.

Quadro 2. Critérios para execução dos algoritmos na fase de Mineração de Dados Educacionais.

Critérios	Algoritmos			
	Árvore de Decisão	<i>Random Forest</i>	<i>Gradient Boosted Tree</i>	<i>Naive Bayes</i>
Split de dados	70/30	70/30	70/30	70/30
Cross Validation	k=10	k=10	k=10	K=10
Depth	n=10	n=10	n=10	Não se aplica
Total de Árvores	1	100	100	Não se aplica

Fonte: Desenvolvido pelos autores (2020).

A análise dos modelos foi realizada por meio das métricas de validação sendo elas a Acurácia, Índice Kappa, Matriz de Confusão, Recall, Precision e Importância da Árvore (Weight by Tree Importance). Os resultados encontrados nessa fase de mineração são apresentados a seguir.

3. Resultados e Discussões

O objetivo desse estudo foi observar se os fatores referentes às atividades dos Diretores de Escola puderam contribuir para o desempenho dos alunos do Ensino Médio, nas avaliações em larga escala do SAEB no ano de 2017. Após a

coleta e tratamento de dados, a fase de mineração de dados educacionais foi realizada e os resultados estão descritos no Quadro 3.

Quadro 3. Resultados da Mineração de Dados para Língua Portuguesa e Matemática

Critérios de Avaliação	Língua Portuguesa				Matemática			
	DT	RF	GBT	NB	DT	RF	GBT	NB
Acurácia	70,36%	73,40%	73,25%	68,70%	69,35%	71,89%	71,56%	66,39%
Kappa	0,559	0,610	0,609	0,539	0,544	0,587	0,583	0,508
Recall	65,72%	69,83%	70,67%	64,24%	60,78%	64,31%	64,26%	60,30%
Precision	68,79%	72,52%	71,97%	63,38%	64,65%	69,70%	67,47%	58,75%
Correlation	68,40%	71,00%	70,90%	64,60%	65,40%	68,50%	68,50%	62,10%

DT= Árvore de Decisão - RF= Random Forest - GBT= Gradient Boosted Tree - NB =Naive Bayes

Fonte: Desenvolvido pelos Autores (2020).

Dentre os algoritmos utilizados o que apresentou o melhor desempenho foi o Random Forest, tanto quando utilizado como atributo alvo o nível da escala de Língua Portuguesa (Escala_LP) quanto para o nível da escala em Matemática (Escala_MT). Observa-se que de acordo com os resultados descritos no Quadro 3, que os desempenhos apresentados pelos algoritmos obtiveram bons índices de desempenho, no entanto o Random Forest foi levemente superior.

Complementando a fase de análise dos desempenhos apresentados pela mineração, outro critério que pode oferecer bom nível de informação para análise desses desempenhos, é a Matriz de Confusão. Essa matriz é uma técnica que utiliza estatística para avaliar o nível de eficiência das predições realizadas por um determinado modelo. No modelo produzido pela mineração pode ser que uma ocorrência verdadeiramente negativa pode ter sido classificada como positiva, ou vice e versa, com isso a matriz pode auxiliar na análise dos resultados produzidos. Para produzir a matriz são avaliadas quatro condições:

- TP (True Positive): número de ocorrências positivas que foram classificadas como positivas;
- FP (False Positive): número de ocorrências negativas que foram classificadas como positivas;
- TN (True Negative): número de ocorrências negativas que foram classificadas como negativas;
- FN (False Negative): número de ocorrências positivas que foram classificadas como negativas.

A Matriz de Confusão referente ao algoritmo Random Forest para as proficiências Língua Portuguesa e Matemática estão descritas nos Quadros 4 e 5.

Quadro 4. Matriz de Confusão para a Proficiência Língua Portuguesa.

Língua Portuguesa						
Verdadeiro	Level_2	Level_3	Level_1	Level_4	Level_5	Class Precision
Level_2	3325	541	509	4	1	75.91%
Level_3	537	1526	6	163	3	68.28%
Level_1	498	2	1564	0	0	75.78%
Level_4	3	107	0	355	97	63.17%
Level_5	1	0	0	51	192	78.69%
Class Recall	76.19%	70.13%	75.23%	61.95%	65.53%	

Fonte: Desenvolvido pelos Autores (2020).

Quadro 5. Matriz de Confusão para a Proficiência Matemática.

Matemática						
Verdadeiro	Level_2	Level_3	Level_1	Level_4	Level_5	Class Precision
Level_2	3253	588	556	7	0	73.86%
Level_3	471	1431	1	242	7	66.50%
Level_1	491	9	1607	0	0	76.27%
Level_4	3	71	0	293	126	59.43%
Level_5	0	5	0	56	151	71.23%
Class Recall	77.12%	68.01%	74.26%	49.00%	53.17%	

Fonte: Desenvolvido pelos Autores (2020).

Os resultados das matrizes de confusão indicaram que houve uma boa classificação em todos os níveis analisados. O próximo passo então, foi observar quais os principais fatores indicados pelo Random Forest que influenciaram no desempenho para Língua Portuguesa e Matemática. Ressaltando que o atributo alvo para execução os algoritmos foi o nível da escala de desempenho, então para Língua Portuguesa o atributo alvo foi “Escala_LP” e para Matemática o “Escala_MT”. Os resultados referentes aos atributos mais importantes, com base no Peso de Importância da Árvore, estão descritos no Quadro 6.

Quadro 6. Resultados referentes ao operador Peso de Importância da Árvore.

Peso de Importância da Árvore (Weight by Tree Importance)			
LP		Matemática	
EXP_FUNCAO	1,0	EXP_FUNCAO	1,0
Escala_MT	0,533	POSGRADUACAO	0,553
MELHOR_ENSINO_ALUNOS	0,532	ORG_TREINAMENTO	0,553
REPROVACAO_ESCOLAR	0,511	Escala_LP	0,535
POSGRADUACAO	0,508	NIVEL_SOCIOECONOMICO	0,528
ORG_TREINAMENTO	0,506	REPROVACAO_ESCOLAR	0,515
FALTAS_REUNIAO_PAIS	0,495	MELHOR_ENSINO_ALUNOS	0,496
NIVEL_SOCIOECONOMICO	0,456	FALTAS_REUNIAO_PAIS	0,484
FALTA_INDIVIDUAL_PAIS	0,440	MERENDA_REC_FINANCEIROS	0,479
FUNC_ESCOLA_REC_FINANCEIROS	0,437	FUNC_ESCOLA_REC_FINANCEIROS	0,430
MERENDA_REC_FINANCEIROS	0,434	FALTA_INDIVIDUAL_PAIS	0,428
CONSELHO_CLASSE	0,404	ABANDONO_ESCOLAR	0,381
ABANDONO_ESCOLAR	0,385	CONSELHO_CLASSE	0,381
FUNC_ESCOLA_FALTA_ADM	0,371	FUNC_ESCOLA_INEX_PROF	0,343
FUNC_ESCOLA_INEX_PROF	0,371	FUNC_ESCOLA_FALTA_ADM	0,339
CONSELHO_ESCOLA	0,306	MERENDA_QUANT_ALIMENTOS	0,334
MERENDA_QUANT_ALIMENTOS	0,294	FUNC_ESCOLA_FALTA_PED	0,262
FUNC_ESCOLA_FALTA_PED	0,287	CONSELHO_ESCOLA	0,260
FUNC_ESCOLA_FALTA_REC_PED	0,282	FUNC_ESCOLA_FALTA_REC_PED	0,259
MERENDA_QUALI_ALIMENTOS	0,247	MERENDA_QUALI_ALIMENTOS	0,243
FUNC_ESCOLA_FALTAS_PROF	0,207	FUNC_ESCOLA_ROTAT_PROF	0,193
FUNC_ESCOLA_ROTAT_PROF	0,174	EXER_CARGO_INTERF_EXTERNA	0,186
FUNC_ESCOLA_FALTAS_ALUNOS	0,163	FUNC_ESCOLA_FALTAS_ALUNOS	0,173
FUNC_ESCOLA_INTERRUP_ATIV	0,153	FUNC_ESCOLA_FALTAS_PROF	0,167
EXER_CARGO_INTERF_EXTERNA	0,146	FUNC_ESCOLA_INTERRUP_ATIV	0,159
FUNC_ESCOLA_INDISC_ALUNOS	0,135	FUNC_ESCOLA_INDISC_ALUNOS	0,145
EXER_CARGO_APOIO_SUPERIORES	0,0	EXER_CARGO_APOIO_SUPERIORES	0,0

Fonte: Desenvolvido pelos Autores (2020).

No Quadro 6, pode-se observar que os primeiros oito atributos são bem semelhantes para as duas proficiências, em se tratando de importância. O que fica evidente é o primeiro fator, “Experiência na Função” do Diretor, ou seja, para as duas proficiências analisadas o peso maior de Importância da Árvore sinaliza que o desempenho dos alunos pode ser diretamente influenciado por esse atributo. Os outros sete atributos com maior relevância estão diretamente relacionados com as atividades de gestão, e com isso, os resultados práticos referentes a como cada

atributo é operacionalizado no ambiente escolar, podem ser influenciados pela experiência do diretor. Esses sete atributos são:

- Nível de desempenho escolar conforme escalas (Escala_MT ou Escala_LP).
- Se existe no ambiente escolar, debate com os professores para propor medidas com o objetivo de melhorar o ensino e a aprendizagem dos alunos (MELHOR_ENSINO_ALUNOS).
- Se existe ações para reduzir a Reprovação Escolar (REPROVACAO_ESCOLAR).
- O nível de formação do Diretor em relação à sua Pós-Graduação (POSGRADUACAO).
- Se o Diretor organiza treinamentos e capacitações (ORG_TREINAMENTO).
- Se o Diretor convoca os pais/responsáveis para conversar, em reunião de pais, sobre o nível de faltas dos seus filhos, e finalmente
- O Nível Socioeconômico da escola (NIVEL_SOCIOECONOMICO).

A experiência em gestão pode ser adquirida pelo tempo de exercício na função, mas pode incluir também as percepções como os gestores observam as diversas demandas do ambiente escolar, e a flexibilidade na tomada de decisão que deve ser amparada por um modelo de gestão democrática, e isso pode fazer a diferença na condução das atividades rotineiras de uma escola (Souza & Ribeiro, 2017; Barros et al., 2019; Pereira & Gasque, 2019; Souza, 2020;).

Em relação ao nível de instrução do Diretor, a formação acadêmica se torna importante à medida que os profissionais adquirem mais conhecimentos, possuem espaços para promover debates com temas referentes às suas atividades profissionais, podendo inclusive, observar novas tendências da área de gestão escolar, tanto em práticas que auxiliem a equipe pedagógica, quanto nas atividades relacionadas à integração escola e comunidade (Vieira & Bussolotti, 2018; Marangoni et al., 2019; Morais et al., 2020; Ribeiro et al., 2019; Zambello et al., 2019).

No ambiente escolar, uma das tarefas que demandam maior tempo no cotidiano escolar, e que tem seus resultados amparados pelo nível de experiência dos Diretores, é a questão da condução das equipes, seja ela administrativa ou pedagógica. Tarefas como Organizar Treinamentos para capacitação das equipes se torna importante à medida que novas habilidades podem ser desenvolvidas para o auxílio de todos os envolvidos no ambiente escolar, seja o corpo docente, os alunos ou até mesmo a comunidade (Vieira & Bussolotti, 2018; Freitas, 2019).

A reprovação escolar é um atributo que demanda atenção pelos gestores escolares, pois as questões que podem levar à reprovação podem ter diversas origens, desde a desmotivação para os estudos, questões familiares, questões de saúde, e ainda aspectos relacionados com as atividades pedagógicas (Paula et al., 2018; Garcia et al., 2019;). Os gestores devem estar alinhados com todas essas demandas a fim de tentar evitar que os alunos não consigam aprovação nos cursos que estão estudando (Mohr & Naujorks, 2017). O nível de experiência dos Diretores pode ser primordial para o desenvolvimento de ações eficazes no combate à reprovação escolar. Importante observar que somada a outras questões, os alunos com reprovação escolar podem optar por abandonar definitivamente seus estudos (Góis & Rocha, 2019; Santos, 2019).

O apoio da família junto à escola é primordial para que o nível de relacionamento entre aluno-família-escola possa ser fortalecido. As dificuldades que os alunos possam ter, em relação à aprendizagem e ao nível de faltas, devem ser discutidas no ambiente escolar, reunindo para esse fim os Diretores, pais ou responsáveis pelos alunos e o corpo docente. As ações consideradas alternativas para solucionar questões como essas, após elaboradas, podem estar diretamente relacionadas com as atividades de gestão escolar, e a qualidade dos resultados alcançados pela implantação dessas ações, reflete o nível de experiência do Diretor de Escola (Oliveira et al., 2017; Santos, 2017; Soares et al., 2018; Dallazem & Coelho, 2019; Silva & Leal, 2020).

Foram encontradas publicações que tratam das questões envolvendo o nível socioeconômico dos alunos em relação aos seus desempenhos em Língua Portuguesa ou Matemática. Isso demonstra que, essa preocupação já ocupa um lugar no debate sobre os impactos da economia no desenvolvimento educacional dos indivíduos e da família.

A qualidade de acesso a renda vem se agravando ao longo do tempo, por exemplo, o Brasil não consegue evoluir no ranking referente ao Índice de Desenvolvimento Humano (IDH) que mede a desigualdade social no país. Em 2017 o Brasil ocupava o 78º lugar já em 2018 caiu para 79º, em um ranking que contempla 189 países e territórios, desenvolvido pelo Programa das Nações Unidas para o Desenvolvimento –PNUD (ONU, 2019).

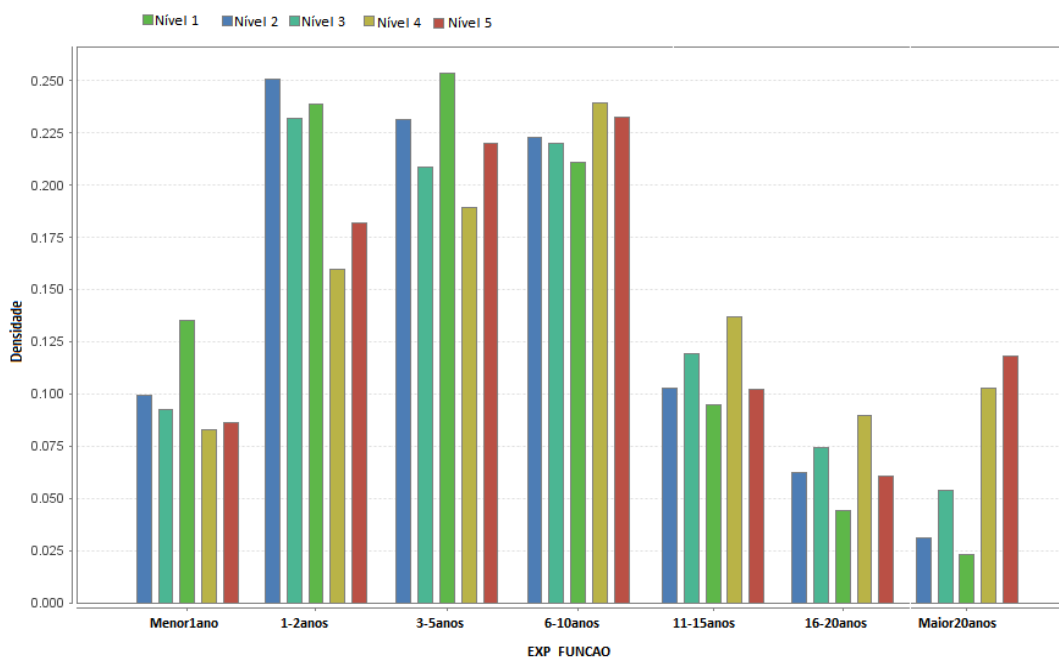
O nível socioeconômico pode ilustrar o poder aquisitivo da família, podendo justificar com isso, a qualidade do nível de acesso a recursos de ensino

aprendizagem, como exemplo recursos tecnológicos e Internet. O nível socioeconômico da família, inclusive, pode influenciar a autoestima dos alunos, principalmente entre aquelas de condição socioeconômica mais desfavorecida (Taberner et al., 2017). Já o desempenho escolar pode ser influenciado por questões relacionadas ao nível socioeconômico, em vários aspectos como relatam Matos & Rodrigues (2016), Almeida (2017), Silva & Silva (2018), Bassetto (2019) e Araujo et al. (2020).

Além desses fatores descritos, como influenciadores do desempenho escolar, os gestores precisam enfrentar condições diversas em suas escolas, envolvendo plano pedagógico, questões de infraestrutura, planejamento para a participação dos alunos nas avaliações em larga escala brasileiras, capacitação do corpo docente, projetos para a comunidade, sem levar em conta que todas essas atividades necessitam de recursos financeiros, constantemente em falta (Colares & Siqueira, 2017; Lefone, 2017; Núñez, 2019; Oliveira et al., 2019; Sá et al., 2020).

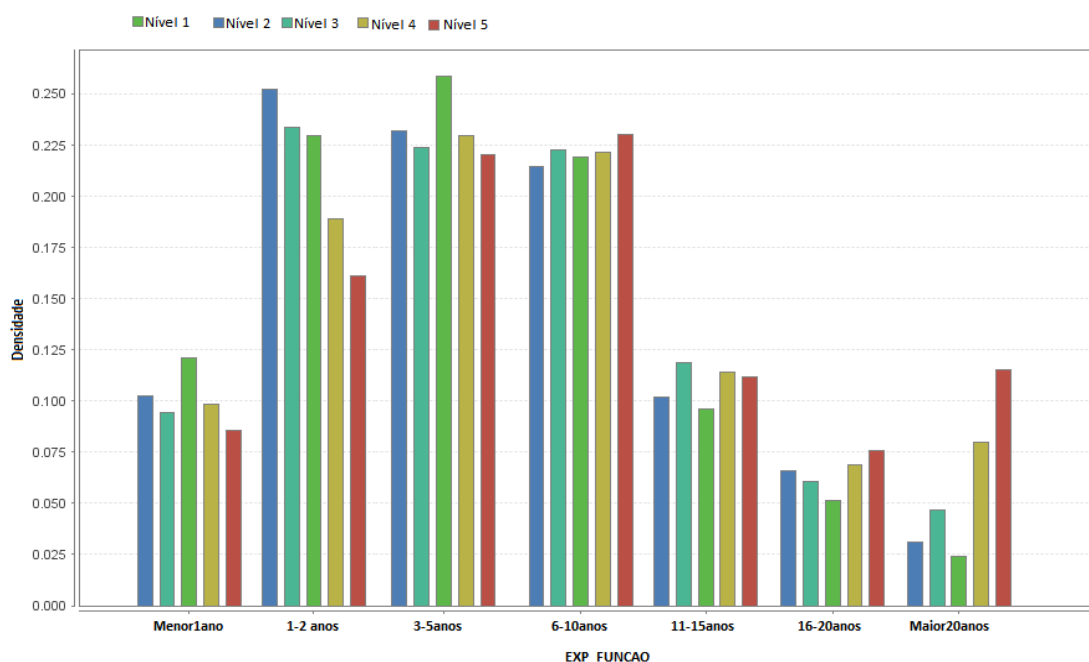
De forma a ilustrar o atributo mais importante, segundo o Quadro 6, este foi representado por gráfico gerado pelo algoritmo Naive Bayes, tanto para proficiência Língua Portuguesa (Figura 3) quanto para Matemática (Figura 4).

Figura 3. Resultados do atributo Tempo de Experiência na Função, para Língua Portuguesa.



Fonte: Desenvolvido pelos autores (2020)

Figura 4. Resultados do atributo Tempo de Experiência na Função, para Matemática.

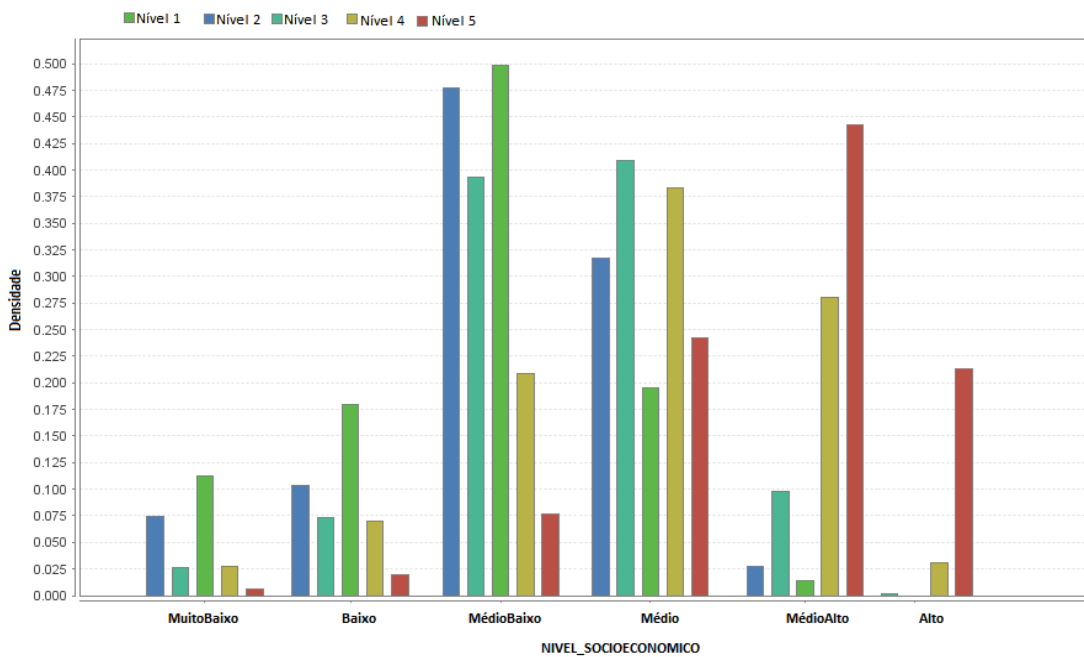


Fonte: Desenvolvido pelos autores (2020)

Pode-se observar nas Figuras 3 e 4 que a experiência na função pode contribuir para o desempenho escolar, pois os desempenhos apresentados pelas escolas, e que são compatíveis com os níveis 4 e 5 da escala de desempenho, vão se acentuando a partir do tempo de experiência “3-5anos”, até ficar em evidência quando o Diretor ultrapassar mais de vinte anos de experiência na função

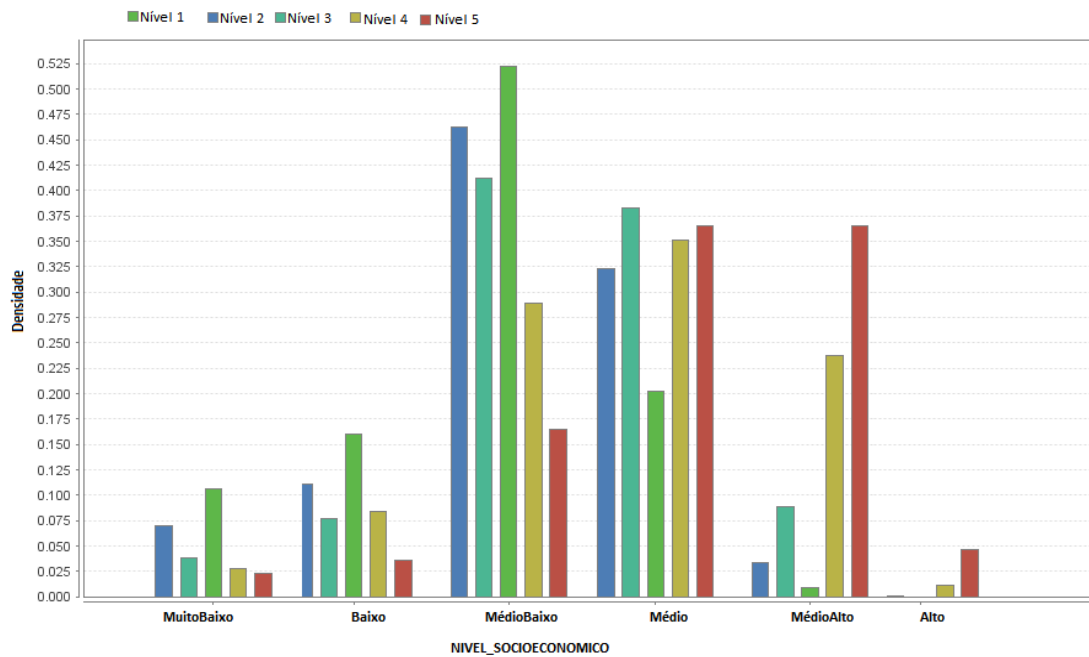
Outro atributo muito citado na bibliografia e que merece um destaque nesse estudo, por que também ficou entre os oito atributos com maior poder de influenciar o desempenho escolar, de acordo com o Quadro 6, é o Nível Socioeconômico. Então, de forma a representar graficamente esse atributo, as Figuras 5 e 6 ilustram os resultados para Língua Portuguesa e Matemática, respectivamente.

Figura 5. Resultados do atributo Nível Socioeconômico, para Língua Portuguesa.



Fonte: Desenvolvido pelos autores (2020)

Figura 6. Resultados do atributo Nível Socioeconômico, para Matemática.



Fonte: Desenvolvido pelos autores (2020)

De acordo com os resultados apresentados nas Figuras 5 e 6, e convergindo com a bibliografia, fica evidente que o Nível Socioeconômico é atributo importante e influenciador do desempenho escolar dos alunos. Observa-se que a partir do Nível Socioeconômico Médio, evoluindo para Médio Alto e Alto, o desempenho das escolas vai aumentando consistentemente, chegando a total evidência no Nível Socioeconômico Alto. De certo que alunos com maior poder aquisitivo podem ter melhor acesso a recursos de aprendizagem, incluindo informática e Internet.

Cada escola tem sua realidade, então um estudo mais minucioso dos atributos apresentados no Quadro 6, com maior peso de Importância, pode contribuir para o desenvolvimento de planos de ação para serem implantados no ambiente escolar, com intuito de melhorar positivamente a influência desses atributos, não só para o desempenho escolar dos alunos, mas também no modelo de gestão da escola.

4. Considerações Finais

De forma a observar quais atributos oriundos das atividades dos Diretores de Escola puderam contribuir para o desempenho escolar dos alunos do Ensino Médio, em avaliações de larga escala do SAEB, na edição de 2017, esse estudo analisou em conjunto com a bibliografia disponível, os resultados dessa edição com as respostas dos questionários aplicados aos Diretores de Escola. O desenvolvimento desse estudo foi amparado pela técnica de Mineração de Dados Educacionais e envolveu os algoritmos para o desenvolvimento de Árvore de Decisão, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes*.

Após a fase de seleção, pré-processamento, tratamento dos dados e da Mineração de Dados, os resultados indicaram que o algoritmo *Random Forest* apresentou o melhor desempenho tanto para a proficiência Língua Portuguesa quanto para Matemática. Em relação aos atributos que puderam influenciar o desempenho escolar, o mais importante foi o Tempo de Experiência na função, resultado esse, que também foi corroborado pelo gráfico gerado pelo algoritmo *Naive Bayes* e pela bibliografia. Uma lista de atributos, com base nos resultados do operador de Peso de Importância da Árvore foi criada e ilustrada no Quadro 6.

A contribuição desse artigo está principalmente na descoberta do atributo Tempo de Experiência na Função (EXP_FUNCTION) como fator preditor do desempenho escolar. Esse atributo pode inclusive estar relacionado com a qualidade dos resultados de ações promovidas pelos gestores, no ambiente escolar. Ações como organização de treinamentos e capacitação das equipes, combate à reprovação escolar, relacionamento com pais e responsáveis para promover melhorias na aprendizagem dos alunos e combater o índice de faltas e ainda se os Diretores promovem debates junto ao corpo docente para desenvolver práticas pedagógicas com intuito de melhorar a aprendizagem dos alunos, estão diretamente relacionadas com as atividades de gestão e podem ter seus resultados influenciados pela experiência desse Gestor Escolar.

Além disso as questões relacionadas ao Tempo de Experiência, Organização de Treinamentos e ações relacionadas as atividades pedagógicas podem contribuir para o desenvolvimento de melhorias nos planos de carreira, tanto da função de Diretor quanto do corpo docente.

Outro atributo importante, de acordo com o Quadro 6, para influenciar o desempenho dos alunos é o Nível Socioeconômico da escola. Esse atributo é citado em várias publicações e foi confirmado nesse estudo por meio dos resultados encontrados. De certo que escolas que possuem alunos com melhor poder aquisitivo podem apresentar um diferencial no nível de aprendizado, pois esses alunos têm melhores condições de acesso a recursos de tecnologia, internet, livros e inclusive questões envolvendo saúde e alimentação.

Infelizmente a escola não tem poderes para influenciar diretamente no nível socioeconômico dos alunos e suas famílias, mas os Diretores podem desenvolver planos de ação que visem contribuir com mais e melhor acesso a recursos de ensino aprendizagem, pelos alunos com menor poder aquisitivo, influenciando positivamente o nível de equidade entre os alunos da escola.

O estudo minucioso dos fatores preditores para o desempenho escolar, resultados dessa pesquisa, pode contribuir para um debate envolvendo políticas públicas, com intuito de promover melhorias no sistema escolar, para prover condições de gestão mais amplas e com mais recursos. Além disso, propostas para atualizações em planos de carreira, reforçando o tempo de experiência dos profissionais podem ser desenvolvidas, tanto para membros da gestão escolar quanto para o corpo docente.

Como sugestão de melhoria para a formação acadêmica dos gestores, é a inclusão nos curriculums, dos cursos de graduação direcionados para a educação e gestão escolar, de conteúdos mais atualizados e direcionados com as necessidades da função de Diretor, apenas como exemplos de conteúdos, aqueles relacionados com a Liderança e Desenvolvimento de Equipes, Gestão Financeira e Sistemas de Informação, poderiam contribuir muito com a rotina do Diretor de escola.

Como a maioria das pesquisas realizadas essa também encontrou limitações, uma delas está diretamente relacionada ao preenchimento dos questionários. Muitos Diretores de Escola não preencheram corretamente seus questionários, ou ainda, deixaram em branco. Outra limitação está relacionada com a complexidade escolar, ou seja, não foi possível associar os resultados da escola com o nível de complexidade de gestão escolar, com isso, escolas com poucas turmas e poucos alunos podem apresentar melhores desempenhos nas avaliações e vice-versa.

Como sugestões para novas pesquisas, seria tentar replicar a metodologia desse estudo em outros níveis de avaliação em larga escala, como exemplos, avaliação para alunos do Ensino Fundamental do 5º ou 9º anos, para alunos que realizaram avaliações do PISA e ainda alunos que realizaram os exames do ENEM. Ainda assim, a metodologia desse estudo poderia ser replicada nos resultados dos alunos do Ensino Médio para edição do SAEB 2019, assim que os resultados estiverem disponíveis.

Referências

- Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology*, 110(8), 1066–1083. <https://doi.org/10.1037/edu0000256>
- Almeida, L. C. (2017). As desigualdades e o trabalho das escolas: Problematizando a relação entre desempenho e localização socioespacial. *Revista Brasileira de Educação*, 22(69), 361–384. <https://doi.org/10.1590/s1413-24782017226919>
- Araujo, J. M. de, Martin, D. G., Ferreira, M. A. M., & Faria, E. R. de. (2020). Fatores determinantes do desempenho educacional no Sudeste Brasileiro. *Gestão e Sociedade*, 14(38), 3507–3525. <https://doi.org/10.21171/ges.v14i2.2942>
- Back, L. T., Polk, E., Keys, C. B., & McMahon, S. D. (2016). Classroom management, school staff relations, school climate, and academic achievement: Testing a model with urban high schools. *Learning Environments Research*, 19(3), 397–410. <https://doi.org/10.1007/s10984-016-9213-x>
- Barros, Á. G. de, Ramos, R. R., Marins, C. de S., & Oliveira, F. H. B. de. (2019). A Relevância da formação continuada para os Gestores Escolares da rede pública. *LinkSciencePlace - Interdisciplinary Scientific Journal*, 6(3), Article 3. <http://revista.srvroot.com/linkscienceplace/index.php/linkscienceplace/article/view/722>

- Bassetto, C. F. (2019). Family background and school performance: An approach with binary variables from SARESP results. *Revista Brasileira de Estudos de População*, 36. <https://doi.org/10.20947/s0102-3098a0077>
- Bezerra, L. N. M., & Silva, M. T. (2020, outubro 1). Educational Data Mining Applied to a Massive Course [Article]. *International Journal of Distance Education Technologies (IJDET)*. www.igi-global.com/article/educational-data-mining-applied-to-a-massive-course/263759
- BRASIL. (1988). Constituição Federal—Texto base e suas alterações. http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm
- BRASIL. (1996). Lei no 9.394, de 20 de dezembro de 1996, Estabelece as diretrizes e bases da educação nacional. http://www.planalto.gov.br/ccivil_03/leis/L9394.htm
- Brito, M. V. S. de. (2017). Determinantes do sucesso escolar: Uma análise multinível a partir dos dados do Pisa 2015 [Dissertação de Mestrado, Universidade de Brasília]. <http://repositorio.unb.br/handle/10482/24535>
- Calixto, K., Segundo, C., & Gusmão, R. P. de. (2017). Mineração de dados aplicada a educação: Um estudo comparativo acerca das características que influenciam a evasão escolar. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 28(1), 1447. <https://doi.org/10.5753/cbie.sbie.2017.1447>
- Cária, N. P., & Lambert-de-Andrade, N. (2016). Democratic Management at school: In search of participation and leadership. *Revista Eletrônica de Educação*, 10(3), 25–40. <https://doi.org/10.14244/198271991907>
- Colares, M. L. I. S., & Siqueira, A. O. dos S. (2017). Desafios da Gestão da Escola de Tempo Integral. *Cadernos de Pesquisa*, 24(3), 72–86. <https://doi.org/10.18764/2178-2229.v24n3p72-86>
- Colpani, R. (2018). Mineração de Dados Educacionais: Um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar. *Informática na educação: teoria & prática*, 21(3), Article 3. <https://doi.org/10.22456/1982-1654.87880>
- Dallazem, A., & Coelho, V. R. (2019). O Desempenho Escolar na voz dos Atores de Escolas Públicas Catarinenses. *Revista Teias*, 20(56), 398–417. <https://doi.org/10.12957/teias.2019.34205>
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Advances in Knowledge Discovery and Data Mining (U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy, Orgs.; p. 1–34). American Association for Artificial Intelligence. <http://dl.acm.org/citation.cfm?id=257938.257942>
- Fernandes, W. L., Godoi, P. F. C. de, Monteiro, L. R. L. de, & Costa, T. L. da. (2020). A Segurança Alimentar e Nutricional nas Escolas Municipais de Araguaína-TO sob a Influência de Políticas Públicas de Combate a Fome. *Facit Business and Technology Journal*, 2(16), Article 16. <https://jnt1.websiteseuro.com/index.php/JNT/article/view/609>
- Fonseca, S. O. da, Silva, A. da R., & Namen, A. A. (2018). Uma metodologia para a descoberta de conhecimento em bases de dados da Prova Brasil A methodology for knowledge discovering in Prova Brasil databases. *Educação Matemática Pesquisa : Revista do Programa de Estudos Pós-Graduados em Educação Matemática*, 20(2), Article 2. <https://doi.org/10.23925/1983-3156.2018v20i2p257-282>
- Freitas, E. D. S. de. (2019). Práticas de gestão do conhecimento como variáveis intervenientes nos resultados do IDEB [Dissertação de Mestrado, Universidade UNICESUMAR]. <http://rdu.unicesumar.edu.br/handle/123456789/5997>
- Garcia, P. S., Prearo, L. C., Romero, M. C., & Bassi, M. (2019). Ensino Fundamental no Grande ABC Paulista: Fracasso e desempenho escolar. *Educação e Cultura Contemporânea*, 16(44). <https://doi.org/10.5935/2238-1279.20190058>
- Góis, L. S., & Rocha, G. S. (2019). A Atuação do Gestor Escolar no Enfretamento da Evasão no Instituto Federal da Bahia (IFBA). *Revista Eletrônica Científica Ensino Interdisciplinar*, 5(14), 340–361.
- Goldschmidt, R., Passos, E., & Bezerra, E. (2015). Data mining: Conceitos, técnicas, algoritmos, orientações e aplicações. (2a). Elsevier.

- INEP. (2019). Censo Escolar. <http://portal.inep.gov.br/web/guest/censo-escolar>
- INEP. (2020a). INEP - IDEB. <http://portal.inep.gov.br/ideb>
- INEP. (2020b). SAEB - Microdados. <http://portal.inep.gov.br/web/guest/microdados>
- Júnior, O. de G. F., Rodrigues, W. R. M., Barbirato, J. C. C., & Costa, E. de B. (2019). Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: Um estudo de caso em escolas municipais de Maceió. *RENTE*, 17(1), 296–305. <https://doi.org/10.22456/1679-1916.95796>
- Lefone, F. R. (2017). O Protagonismo do Diretor de Escola em Interface com a Formação Continuada de Professores [Dissertação de Mestrado, Universidade Metodista de São Paulo]. <http://tede.metodista.br/jspui/handle/tede/1701>
- Marangoni, R. A., Araújo, F. S. de G., & Schwarz, R. R. (2019). Diretor de escola: Dificuldades e enfrentamentos. *Revista Triângulo*, 12(2), 14–31. <https://doi.org/10.18554/rt.v0i0.3297>
- Masci, C., De Witte, K., & Agasisti, T. (2018). The influence of school size, principal characteristics and school management practices on educational performance: An efficiency analysis of Italian students attending middle schools. *Socio-Economic Planning Sciences*, 61, 52–69. <https://doi.org/10.1016/j.seps.2016.09.009>
- Matos, D. A. S., & Rodrigues, E. C. (2016). Indicadores educacionais e contexto escolar: Uma análise das metas do Ideb. *Estudos em Avaliação Educacional*, 27(66), 662–688. <https://doi.org/10.18222/ae.v27i66.4012>
- MEC. (2014). PNE - Plano Nacional de Educação—Plano Nacional de Educação—Lei nº 13.005/2014. <http://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>
- Melo, L. L. de, & Miranda, N. A. de. (2020). Desafios da gestão na escola pública paulista: A percepção do diretor ingressante. *Revista Brasileira de Política e Administração da Educação - Periódico científico editado pela ANPAE*, 36(1), 130–152. <https://doi.org/10.21573/vol36n12020.96631>
- Mesquita, S. S. de A. (2018). Referenciais do “bom professor” de ensino médio: Exercício de articulação teórica. *Cad. Pesqui*, 48. <https://doi.org/10.1590/198053144820>
- Mohr, A. C., & Naujorks, M. I. (2017). Políticas Educacionais de Correção de Fluxo no RS: Efeitos da Racionalidade Neoliberal na Gestão Escolar. *Reflexão e Ação*, 25(1), 160–177. <https://doi.org/10.17058/rea.v25i1.7770>
- Morais, J. de S., Nascimento, F.-L. S. C. do, & Magalhães, N. R. S. (2020). Evocações narrativas de gestores escolares acerca das políticas de saberes e fazeres da gestão. *Educação, Ciência e Cultura*, 25(2), 221–237. <https://doi.org/10.18316/recc.v25i2.6430>
- Nascimento, V. P. do, & Chiusoli, C. L. (2019). O papel do gestor escolar: Estudo de caso sobre os desafios da educação pública. *Revista on line de Política e Gestão Educacional*, 23(2), 238–254. <https://doi.org/10.22633/rpge.v23i2.11997>
- Nichele, P. T., & Mello, M. A. da S. (2020). Gestão Escolar na Perspectiva da Educação Democrático-Participativa e a Função Social da Escola. *Revista Saberes Pedagógicos*, 4(3), 323–343. <https://doi.org/10.18616/rsp.v4i3.6213>
- Núñez, C. (2019). Escolas em foco: Uso de dados por gestores escolares. <https://doi.org/10.20500/RCE.V14I31.30118>
- Oliveira, A. C. P. D., Carvalho, C. P. D., Oliveira, A. C. P. D., & Carvalho, C. P. D. (2018). Gestão escolar, liderança do diretor e resultados educacionais no Brasil. *Revista Brasileira de Educação*, 23. <https://doi.org/10.1590/s1413-24782018230015>
- Oliveira, E. C. B. de, Lima, M. A. V. de, & Bryto, K. K. C. (2019). A Relevância da Gestão Estratégica de Pessoas no Ambiente Escolar. *Revista de Administração e Contabilidade - RAC*, 6(12), 137–150.

- Oliveira, I. L. dos S., Braga, A. P., & Prado, C. M. N. (2017). Participação da família no desenvolvimento da aprendizagem da criança. *Estação Científica (UNIFAP)*, 7(2), 33–44. <https://doi.org/10.18468/estcien.2017v7n2.p33-44>
- ONU. (2019, dezembro 9). Relatório de desenvolvimento humano do PNUD destaca altos índices de desigualdade no Brasil. ONU Brasil. <https://nacoesunidas486780792.wpcomstaging.com/relatorio-de-desenvolvimento-humano-do-pnud-destaca-altos-indices-de-desigualdade-no-brasil/>
- Paula, J. S. de, Franco, A. M. de P., & Silva, J. W. da. (2018). Fatores relacionados ao atraso escolar no estado de Minas Gerais. *Estudos em Avaliação Educacional*, 29(72), 886–917. <https://doi.org/10.18222/eaee.v29i72.4928>
- Pereira, I. A. (2020). Tipos de liderança dentro do ambiente educativo: Uma discussão sobre sua identidade organizacional. *Educationis*, 8(1), 8–22. <https://doi.org/10.6008/CBPC2318-3047.2020.001.0002>
- Pereira, P. R., & Gasque, K. C. G. D. (2019). Tomada de decisão do gestor escolar das escolas públicas de ensino médio no Distrito Federal e a interface com o letramento informacional. Em *Questão*, 25(3), 40–60. <https://doi.org/10.19132/1808-5245253.40-60>
- Provost, F., & Fawcett, T. (2016). *Data Science Para Negócios (1a Edição)*. Editora Alta Books.
- Puyalnithi, T., V., Dr. M. V., & Singh, A. (2016). Comparison of Performance of Various Data Classification Algorithms with Ensemble Methods Using RAPIDMINER. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6, 1–6.
- Ramos, L. S., Moreira, E. M. F., Silva, F. A. da, Nazário, E. A. M., Sobrinho, L. R. P., Brito, D. do N., Monteiro, K. P., Neves, O. T., Santos, R. P. dos, & Almeida, E. M. de. (2020). A humanização da merenda escolar na promoção da saúde e da educação pública: Uma breve revisão. *Revista Eletrônica Acervo Saúde*, 44, e3137–e3137. <https://doi.org/10.25248/reas.e3137.2020>
- Ribeiro, M. D., Síveres, L., & Brito, R. de O. (2019). A Formação de gestores escolares: A dimensão ética em questão. *Educação Por Escrito*, 10(2), e36685–e36685. <https://doi.org/10.15448/2179-8435.2019.2.36685>
- Riddle, S., & Apple, M. W. (2019). *Re-imagining Education for Democracy (1o ed)*. Routledge.
- Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs Data Mining and Knowledge Discovery*, 3(1), 12–27. <https://doi.org/10.1002/widm.1075>
- Sá, A. P. F. V. de, Lima, A. B. de, Gomes, M. J., & Pequeno, L. L. C. (2020). A (In)Disciplina no Contexto Escolar: Desafios e Perspectivas. *Revista Educação & Ensino*, 4(1), Article 1. <http://periodicos.uniateneu.edu.br/index.php/revista-educacao-e-ensino/article/view/57>
- Santos, C. T. F. dos, Rosas, L. V., Lopes, A. P., Acioli, A. N. S., & Lima, R. A. (2018). A Alimentação Escolar no Ensino Básico em Escolas Públicas de Tabatinga-AM. *South American Journal of Basic Education, Technical and Technological*, 5(2), Article 2. <https://periodicos.ufac.br/index.php/SAJEBTT/article/view/1593>
- Santos, D. M. dos. (2019). Evasão escolar na EJA: Estudo comparativo em uma escola da rede pública estadual de Pacatuba – CE / School evasion in EJA: a comparative study in a school of the Pacatuba state public network. *Brazilian Journal of Development*, 5(10), 18916–18940. <https://doi.org/10.34117/bjdv5n10-131>
- Santos, N. M. dos. (2017). *A Gestão Escolar sob a ótica dos pais: Um estudo de caso [Dissertação de Mestrado, Universidade Fernando Pessoa]*. <https://bdigital.ufp.pt/handle/10284/6073>
- Silva, R. M. de S. da, & Leal, D. A. (2020). A Família na vida escolar dos alunos: Uma análise do relacionamento das instituições Escola – Família no município de Fazenda Rio Grande / The Family in the school life of students: an analysis of the relationship of the School - Family institutions in the municipality of Fazenda Rio Grande. *Brazilian Journal of Development*, 6(8), 58476–58488. <https://doi.org/10.34117/bjdv6n8-309>

Silva, T. R., & Silva, J. B. L. (2018). A influência do nível socioeconômico na aprendizagem da leitura. *Revista Educação, Psicologia e Interfaces*, 2(2), 7–28. <https://doi.org/10.37444/issn-2594-5343.v2i2.93>

Soares, G. F., Farias, J. S., Soares, G. F., & Farias, J. S. (2018). Come educate with us: The encouragement of governments and schools to the coproduction of public education by students' relative. *Ensaio: Avaliação e Políticas Públicas em Educação*, 26(101), 1347–1371. <https://doi.org/10.1590/s0104-403620180026001299>

Souza, M. I. M. de. (2020). The role of the school principal: Challenges and possibilities of your professional acting, while facilitator of the teaching-learning process | *Research, Society and Development*. *Research, Society and Development*, 9, 1–15. <http://dx.doi.org/10.33448/rsd-v9i7.3900>

Souza, L. D. M., & Ribeiro, M. S. de S. (2017). O Perfil do Gestor Escolar Contemporâneo: Das permanências as incorporações para exercício da função. *Revista Espaço do Currículo*, 10(1). <https://doi.org/10.15687/rec.v10i1.33386>

Tabernerero, C., Serrano, A., & Mérida, R. (2017). Estudio comparativo de la autoestima en escolares de diferente nivel socioeconómico. *Psicología Educativa*, 23(1), 9–17. <https://doi.org/10.1016/j.pse.2017.02.001>

Vasconcelos, J. C., Lima, P. V. P. S., Rocha, L. A., Khan, A. S., Vasconcelos, J. C., Lima, P. V. P. S., Rocha, L. A., & Khan, A. S. (2020). School infrastructure and public investment in Education in Brazil: The importance of educational performance. *Ensaio: Avaliação e Políticas Públicas em Educação, AHEAD*. <https://doi.org/10.1590/s0104-40362020002802245>

Vieira, A. E. R., & Bussolotti, J. M. (2018). Gestão Escolar: um estudo de caso sobre Escolas Técnicas. *Interação - Revista de Ensino, Pesquisa e Extensão*, 20(1), 45–70. <https://doi.org/10.33836/interacao.v20i1.167>

Zanbello, B. L., Godoy, L. dos S., Amorim, E. de C. N., Dal Forno, L. F., & Macuch, R. da S. (2019). *Administração Escolar: Do Diretor ao Gestor*. UNIVERSIDADE CESUMAR. <http://rdu.unicesumar.edu.br/handle/123456789/3752>

6.6 ARTIGO 6. Use of Data Mining to identify the technological resources that contribute to school performance in large-scale evaluations of Brazilian high school.

Ivonaldo Vicente da Silva

Márcia Terra da Silva

Abstract.

The aim of this study was to observe, through Educational Data Mining, if the level of access to computer and library resources offered by schools could contribute to the performance of high school students, in the large scale evaluations of SAEB, in the edition 2017 in Brazil. After collecting data and carrying out the entire mining phase, the results, by Random Forest, indicated that the level of access to computers by students and teachers, the access to the inter-net by students and the offer of broadband internet by the school can contribute for better performance in SAEB assessments. These results can contribute to the development of projects that improve access to these resources throughout the school environment.

Keywords: School Performance · Educational Data Mining · Large Scale Assessments

1 Introduction

The aim of this study was to observe, through Educational Data Mining, whether the level of access to computer and library resources offered by schools could contribute to the performance of high school students, in large-scale assessments of the SAEB of 2017 edition in Brazil. The level of access to these types of resources can positively influence student performance, so it is important to note how schools are providing access to these resources. In Brazil, the cycle of basic education is composed of elementary education in 9 years and high school, normally offered in 3 years.

To observe the students' level of knowledge, a good alternative is large-scale assessments. In Brazil, SAEB, held every two years, is one of the main assessments for basic education, as it provides information on the quality of teaching, in addition to contributing to the formation of the Basic Education Development Index-IDEB. In addition to assessments for Portuguese language and mathematics proficiencies, students' sociodemographic information is also collected through forms, relating to the exercise of the function of Teachers and School Directors and about the school

structure. The results of SAEB assessments and forms are available on the government website [1].

Due to technological advances, it is important that young people have access to technology resources, such as computers, the Internet and laboratories. These resources are not always available in schools, compromising the students' level of learning [2]. Another important factor in relation to technological education, is that the job market needs qualified professionals and knowledge in technology becomes a pre-requisite.

2 Theoretical Background

To analyze large volumes of data, one of the alternatives is Educational Data Mining, because through machine learning algorithm it is possible to identify patterns of behavior that may contribute to the decision-making process. Data Mining is part of the process called Knowledge Discovery in Databases - KDD, and can be understood as a multi-step process to identify patterns in a large data set [3], [4].

For the development of this study, it was necessary to observe publications related to the theme and which used Data Mining. The studies of [5], [6] used Mining to observe students' performances, [7] has already observed school dropout, [8] the influence of family and teachers, and [9] school effectiveness.

Educational Data Mining has been gaining strength over time, as it can contribute to decision-making processes, providing essential information, through the prediction of patterns and behaviors.

3 Method

After the literature review phase, the next step involved data collection, directly on the government website. The set of data made available includes the results of students' performance, questionnaires addressed to Teachers and Directors and school representatives, the focus of this study [1]. The data of the schools that offered High School in 2017, the respective data dictionary and the performance scales, which were fundamental for the interpretation of these data, were duly separated. Based on the data dictionary, the attributes, Chart 1, were selected, which allowed to assess the quality of supply of technological resources and library, including the socioeconomic level of schools.

Chart 1. List of Factors present in the questionnaires of the Schools.

Attributes	Form questions
What are the conditions of use for:	
COMPUTER_STUDENTS	37. Computers for student use.
INTERNET_STUDENTS	38. Internet access for student use.
COMPUTER_TEACHERS	39. Computers for use by teachers
INTERNET_TEACHERS	40. Internet access for use by teachers.
INTERNET_BROADBAND	53. Internet with broadband connection.
LIBRARY	57. Library
READING_ROOM	58. Reading Room
COMPUTER_LAB	60. Computer Lab.
SCIENCE_LAB	61. Science Lab.
BORROW_BOOKS	68. Books can be handled and borrowed
SOCIOECONOMIC_LEVEL	School Socioeconomic Level
AVG_LP	Average Portuguese language proficiency
AVG_MT	Average Mathematical proficiency

The database, named “DataSetSchool”, was created, containing only the attributes described in Chart 1. Right after, the data treatment phase was performed, to remove the blank, null or erroneously filled records from the database. After that, the next step was to analyze the schools' performances.

Based on the results of the set of students from each school, the government calculates and assigns the school average, varying on a scale between 0 and 500 points, for the Portuguese Language and Mathematics proficiencies, thus, schools with high averages indicate that their students performed well in the assessments. The attributes “SCALE_LP” and “SCALE_MT” were created, to keep the classification referring to the scale level, compatible with the average performance of the school. As an example, Level 1 of the scale comprises values between 225 and 250 points, after that level, the scale progresses every 25 points successively, forming regular intervals. Performances below 225 points, Level Zero, were not analyzed, because according to the government, students with performances compatible with this level did not absorb enough content for their level of education.

To identify whether the school's average performance is above, equal to or below the national average, a simple arithmetic average of the averages assigned to schools was initially calculated. Once these tasks were carried out, the next step was the removal of records that could influence the mining phase. Values of isolated averages that could be considered as outliers, equal to or above 325 points and less

than 225, were removed from the database. That done, of the total of 19,643 schools offering secondary education, 14,492 were used for analysis in this study, corresponding to between Level 1 and 5 of the scales.

After these tasks, the data was ready for the Data Mining phase, and for that the RapidMiner software was used. At this stage, it becomes important to choose which algorithms will be used. Naive Bayes, Decision Tree, Random Forest and Gradient Boosted Tree were selected for this study due to the quality of performance in both classification and regression tasks [3], [10], [11].

The Naive Bayes algorithm, created by the mathematician Thomas Bayes, can be considered a probabilistic classifier based on the Bayes Theorem, which has the conditional probability as logic, that is, what is the probability of event X to occur given event Y. Usually used in classification and regression models, Decision Trees can be based on algorithms such as J48, C.45, ID3, CART, Random Forest, Gradient Boosted Tree, and can be supervised or not. A target attribute is required, in the supervised method, to create the trees. To assemble the trees, the algorithm is in charge of finding the best cut or pruning points based on the entropy calculation. After all calculations have been made, the main node is indicated based on the greatest gain of information. The formula for Entropy is:

$$Entropy(Y) = \sum_i p(c/Y) \log_2 p(c/Y) . \quad (2)$$

Where Y is the set of cases and $p(c/Y)$ the value of Y) is the proportion of class c.

To calculate the information gain, in RapidMiner, it is necessary to choose one of the criteria which are: Accuracy, Information_gain, Gain_ratio, Least_Square. In this way it can be understood that the most important attribute, based on Entropy and Information Gain, is at the top of the tree. The criterion chosen for this study was the Gain-ratio, as it presents a better gain of information in relation to the others.

Random Forest or Forest of Trees are considered a method of joint learning (Ensemble). The algorithm uses the Bootstrap Agregating method to create the forest of trees, which generates a set of data by sampling the bootstrap of the database used. Gradient Boosted Tree also uses the logic of learning together, and can be compared with the Random Forest. , but uses the optimization of the loss function as a criterion for the creation of the set of trees, so when creating a new tree, it tries to correct the errors of the previous tree, in the set, using the learning rate for this.

To execute the algorithms, the criteria used in RapidMiner were: Split Data 70/30, Cross Validation = 10 and for the Decision Tree, Random Forest and Gradient Boosted Tree algorithms, depth = 10 and total trees = 100 were used. The analysis of the models was carried out using the validation metrics, namely the Accuracy, Kappa Index, Recall, Precision and Weight by Tree Importance. Recalling that the target attributes used in the algorithms were "Scale_LP" and "Scale_MT". The results found in this mining phase are shown below.

3.1 Results and Discussion

After completing the Data Mining phase, the results are described in Table 1.

Table 1. Results of the Data Mining phase

	DT		RF		GBT		NB	
	LP	MT	LP	MT	LP	MT	LP	MT
Accuracy	62,03%	73,93%	74,66%	75,35%	74,61%	75,32%	70,53%	70,32%
Kappa	0,404	0,626	0,632	0,646	0,632	0,646	0,577	0,575
Recall	54,54%	65,46%	70,80%	67,56%	71,17%	67,88%	68,95%	65,47%
Precision	61,70%	70,18%	73,77%	73,08%	73,34%	72,72%	65,31%	60,57%
Correlation	65,30%	56,30%	68,80%	56,30%	68,90%	56,40%	68,10%	50,50%

DT= Decision Tree; RF=Random Forest; GBT=Gradient Boosted Tree; NB=Naive Bayes

According to the results in Table 1, it can be seen that the algorithm that showed the best accuracy performance was Random Forest, confirming the predictive power of this algorithm. Regarding the results of the operator Weight by Tree Importance, the 4 best evaluated attributes were COMP_STUDENTS (1.0), COMP_TEACHERS (0.899), INT_STUDENTS (0.859) and INT_BROADBAND (0.846) for Portuguese and COMP_STUDENTS (1.0), COMP_TEACHERS (0.851), INT_STUDENTS (0.789) and INT_BROADBAND (0.786) for Mathematics. Emphasizing that the operator Weight by Tree Importance presents its results normalized on a scale between 0 and 1. The closer to 1, the better the result.

The results found could reinforce the conclusions of [2], [12] that observed the importance of technology for school performance, in addition to these studies [13] [14] reinforce the importance of the school structure that allows quality of access to computers and internet for students. However, it is not enough to just provide access to technology, the teaching staff must be trained and qualified, so that they can contribute to the evolution of the students' learning level [15].

In relation to the average performance of schools, the simple arithmetic averages of the group of schools evaluated were 268.80 points for LP and 270.57 for MT. Of the total analyzed schools, 6,502 are performing at or above the average in LP and 6,050 in MT. With performances below average, they are 8,465 in LP and 8,917 in MT. Regarding the quality of access to resources and performance, of the total of 1,426 schools with performances equal to or greater than 300 points (Level 5), 91.93% indicated the COMP_STUDENTS attribute as good or regular, 96.21% the COMP_TEACHER, 88 , 28% INT_STUDENTS and 91.37% INT_BROADBAND. Thus emphasizing the importance of these resources for school performance.

4 Conclusion

After all the stages of this study were carried out, the results found by the Random Forest algorithm could illustrate that the attributes related to access to computers by students and teachers, level of internet access and, if the school has broadband internet were able to contribute to a better performance of students in the large scale evaluations of SAEB, in the 2017 edition. These results are compatible with the analyzed literature.

As a contribution of this study, in addition to proposing another study based on Random Forest in the area of education, the results of this study can contribute so that the government and education professionals can develop plans to improve access to these resources as well as develop methodologies for education. technology-based teaching to improve student learning.

As limitations was the fact that many forms were blank or incorrectly filled, so they were removed, reducing the volume of data analyzed. As a suggestion for further research, it would be interesting to replicate this model in other editions of the SAEB or in other types of large-scale assessments.

Acknowledgments. This work was carried out with the support of the Coordination of Improvement of Higher Level Personnel - Brazil (CAPES) - Financing Code 001.

References

1. INEP, “SAEB - Resultados e Microdados”. <http://portal.inep.gov.br> (2020).
2. Salam, S., Yang, M., Shaheen, A., Movahedipour, M., Zeng, J.: “ICT and students’ performance in Pakistan”. *Hum. Syst. Manag.* 36, 277–284 (2017). doi: 10.3233/HSM-17118
3. Provost, F., Fawcett, T.: *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. 1. ed. California-EUA: O’Reilly Media (2013).
4. Bezerra, L. N. M., Silva, M. T.: “Educational Data Mining Applied to a Massive Course”. *Int. J. Distance Educ. Technol - IJDET*. (2020). doi: doi:10.4018/IJDET.2020100102
5. Martínez-Abad, Gamazo, F. A., Rodríguez-Conde, M.-J.: “Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment”. *Stud. Educ. Eval.* 66, (2020). doi: 10.1016/j.stueduc.2020.100875
6. Depren, S. K., Askin, O. E., Öz, E.: “Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS”. (2017). doi: 10.12738/ESTP.2017.5.0634
7. Maheshwari, E., C. Roy, M. Pandey, e S. S. Rautray, “Prediction of Factors Associated with the Dropout Rates of Primary to High School Students in India Using Data Mining Tools”, in *Frontiers in Intelligent Computing: Theory and Applications*. (2020). doi: 10.1007/978-981-32-9186-7_26.
8. Giap, Y. C., Leonardi, N., Waseso, B., Rahim, R.: “Data Mining of Family, School, and Society Environments Influences to Student Performance”. *IOP Conf. Ser. Mater. Sci. Eng.* 420, (2018). doi: 10.1088/1757-899X/420/1/012090.
9. Martínez-Abad, F.: “Identification of Factors Associated With School Effective-ness With Data Mining Techniques: Testing a New Approach”. *Front. Psychol.* 10, (2019). doi: 10.3389/fpsyg.2019.02583.
10. Romero, C., Ventura, S.: “Data mining in education”, *WIREs Data Min. Knowl. Discov.* 3, 12–27 (2013). doi: 10.1002/widm.1075.
11. Puyalnithi, T., Viswanatham V, M., Singh, A.: “Comparison of Performance of Various Data Classification Algorithms with Ensemble Methods Using RAPIDMINER”. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 6, (2016).
12. Díaz, L. M. B., García, M. L. S., Cano, E. V.: “Effects on academic performance in secondary students according to the use of ICT”. *IJERI Int. J. Educ. Res.* (2019)
13. Assoumpta, U. I., Andala, D. H. O.: “Relationship between School Infrastructure and Students’ Academic Performance in Twelve Years Basic Education in Rwanda”. *J. Educ.* 3, (2020).
14. Silles, M., Hyland, M., Layte, R., Lyons, S., McCoy, S.: “Are classroom internet use and academic performance higher after government broadband subsidies to primary schools?” *Res. Memo. Univ. Hull Bus. Sch.* 93, (2014).
15. Syahid, A. A., Sunarya, D. T., Sudin, A., Sunaengsih, C., Karlina, D. A.: “ICT competences of primary school teachers”. *J. Phys. Conf. Ser.* 1318, (2019). doi: 10.1088/1742-6596/1318/1/012148.

7 DISCUSSÃO DOS RESULTADOS

O objetivo central desse estudo foi analisar, por meio da Mineração de Dados Educacionais, quais fatores presentes no ambiente educacional podem ser preditores para que os alunos do Ensino Médio obtenham um bom desempenho nas avaliações de larga escala do SAEB.

Os dados para compor a pesquisa foram os resultados das avaliações do SAEB, que são disponibilizados pelo governo no website do INEP. Além dos resultados dos desempenhos dos alunos também foram coletados os resultados dos formulários preenchidos pelos alunos, professores, diretores de escolas e representantes das escolas.

Como metodologia, inicialmente foi realizada a fase de pesquisa bibliográfica para observar publicações existentes sobre o tema desse estudo. Os resultados dessa fase foram essenciais para a construção da lista de fatores preditores do desempenho escolar e que já estavam publicados em artigos e trabalhos acadêmicos. Como segunda parte, os dados coletados no website do INEP foram devidamente tratados e preparados para as etapas envolvendo Estatística Descritiva e Mineração de Dados Educacionais. Os artigos que compõem essa tese foram desenvolvidos utilizando esses dados tratados, bem como os resultados das etapas da Estatística Descritiva e da Mineração de Dados.

Iniciando a fase de análise, a primeira observação está no fato de que os preenchimentos dos formulários do SAEB são realizados pelos próprios profissionais da educação, ou seja pelos professores, diretores e representantes de escola. Os formulários são auto avaliativos e por conta disso as respostas apresentadas podem não representar a realidade ou ser entendidas de diferentes formas. Pode ser que essa situação aconteça pelo fato de que, em avaliações em larga escala, os formulários apresentam as questões com um texto mais abrangente, observando as características da educação em território nacional, não sendo possível tratar das peculiaridades regionais.

Já em relação aos desempenhos apresentados pelos alunos, os resultados do Artigo 1 demonstraram que os alunos, em 2015, estavam concentrados nos níveis iniciais das escalas, sendo que não conseguiram classificação superior ao Nível 3, tanto em Língua Portuguesa quanto em Matemática. Quando foi observada a evolução desses desempenhos por estado, levando em conta a edição de 2017, no Artigo 3, também percebeu-se que, no conjunto dos resultados, os alunos não

apresentaram desempenho superior ao nível 3 das escalas de desempenhos. Inclusive, em relação aos resultados dos Artigos 1 e 3, notou-se uma evolução do percentual de alunos que apresentaram desempenhos compatíveis com o Nível Zero das escalas de desempenho entre as edições de 2015 e 2017.

As questões envolvendo os desempenhos dos alunos em avaliações de larga escala já foram observadas por Laros, Marciano e Andrade (2012) e Andrade, Franco e De Carvalho (2016) e os resultados apresentados por esses autores reforçam a necessidade de ações de melhorias no método de ensino-aprendizagem no ambiente escolar, ressaltando com isso os resultados encontrados nos Artigos 1 e 3 dessa tese.

Com base nesses resultados, se torna necessário o desenvolvimento de planos de ação envolvendo todos os profissionais da educação, família e comunidade, com intuito de propor melhorias em todo o ambiente escolar. Observação semelhante já foram apresentadas por Brito (2017), Calixto, Segundo e Gusmão (2017) e Soares e Rosa (2020).

Em relação aos fatores associados ao desempenho escolar, no Artigo 2 os resultados demonstraram que o nível de escolaridade dos pais ou responsáveis pelos alunos estão associados com os melhores desempenhos dos alunos nas avaliações do SAEB em 2015. Além disso, ficou evidente que quando a mãe e o pai possuem ensino superior, o nível de desempenho apresentado pelo aluno é melhor. Nesse sentido, foi também observado nos resultados desse artigo, que as ações relacionadas à maior participação dos pais, como: se incentivam seus filhos a estudar, se conversam sobre a rotina escolar e se os professores corrigem o dever de casa, também estão associados com o desempenho escolar.

Essas informações levam ao entendimento de que, quando os alunos possuem no ambiente familiar pais ou responsáveis que não abandonaram os estudos, e em alguns casos concluíram o nível superior, se sentem mais motivados para prosseguir com seus estudos e desenvolverem melhor em relação ao aprendizado. O envolvimento dos pais ou responsáveis na vida escolar dos filhos já foi observado por Lobo (2017) e Fantinato e Cia (2017), e os resultados encontrados por esses autores convergem para os resultados descritos nesse estudo.

Os professores exercem papel fundamental na vida escolar dos alunos, até mesmo por que estão em contato direto com eles. Os resultados apresentados no Artigo 4 indicam que o tempo de experiência na função, bem como as atualizações

em habilidades e competências relacionadas à função docente estão associados aos bons desempenhos dos alunos em avaliações em larga escala.

Nesse ponto se faz necessário observar que a função do docente é primordial em uma escola, professores qualificados e preparados para o exercício da função possuem mais condições de motivar seus alunos para o aprendizado. Estudos referentes ao exercício da função pelo professor em se tratando de aprimoramento de práticas pedagógicas foram realizados por Pettersson et al. (2016) e Jovarini, Leme e Correia-Zanini (2018). Os autores salientam que cabe ao professor observar como seus alunos estão se desenvolvendo, além de ficar atento aos sinais que os alunos apresentam em relação aos sintomas de estresse e desmotivação.

De forma similar, o Artigo 5, referente aos fatores dos diretores de escola, apresenta como resultados que o tempo de experiência do Diretor de Escola, no exercício da função, nível de pós-graduação, ações para controle de reprovações e nível socioeconômico da escola estavam associados com os melhores desempenhos apresentados pelos alunos nas avaliações em larga escala do SAEB, em 2017. Pode-se observar que o fator Tempo de Experiência no exercício da função surgiu como possível preditor para o desempenho escolar, tanto nos resultados do Artigo 4 (Professores) quanto no Artigo 5 (Diretores).

Algumas ações realizadas pelos Diretores de Escola podem influenciar na forma como o corpo docente exerce sua função, dessa forma a experiência que esse profissional tem no exercício da função se torna primordial para que as atividades escolares sejam conduzidas de forma a atender os objetivos previstos nos planos educacionais (BARROS et al., 2019; PEREIRA e GASQUE, 2019; SOUZA, 2020; SOUZA e RIBEIRO, 2017). A formação educacional dos Diretores também pode contribuir para que exerçam as atividades inerentes às suas funções com mais eficiência (MARANGONI, 2020; MORAIS, NASCIMENTO e MAGALHÃES, 2020; VIEIRA e BUSSOLOTI, 2018).

Além das atividades burocráticas relacionadas com as práticas de gestão escolar, os Diretores também exercem papel fundamental na condução de suas equipes, tanto pedagógicas quanto administrativas. O estilo de liderança exercido pelo Diretor pode influenciar diretamente na motivação de suas equipes. Além disso, cabe ao Diretor o planejamento dos treinamentos e capacitações necessárias para atualizações de conhecimentos e habilidades, de toda equipe.

De acordo com os resultados do Artigo 5, pode-se perceber que as atividades exercidas pelos Diretores de Escola podem proporcionar condições para que os professores e alunos possam ter um ambiente mais propício para o desenvolvimento de aprendizado.

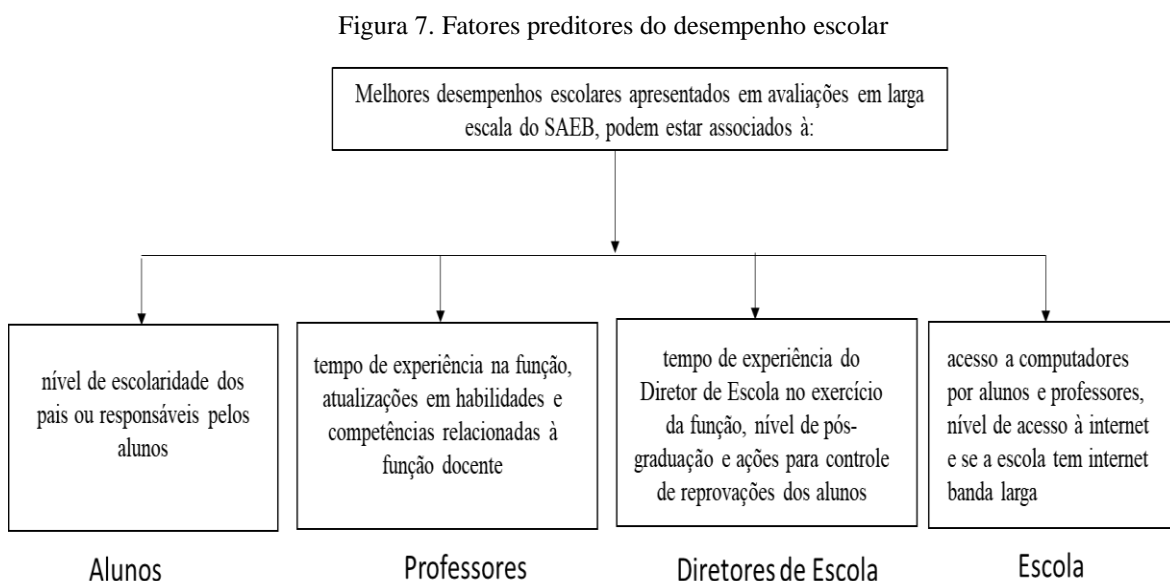
Em relação ao ambiente escolar, os resultados do Artigo 6 demonstraram que o acesso a recursos tecnológicos e internet, tanto por alunos quanto por professores, estão associados com os melhores desempenhos em avaliações de larga escala do SAEB, em 2017. De certo que a estrutura escolar se torna fundamental para o desenvolvimento do aprendizado, e nesse ponto o acesso a computadores e Internet podem fazer a diferença no desenvolvimento de conhecimentos em sala de aula (DÍAZ, GARCÍA e CANO, 2019; SALAM et al., 2017).

Para que as escolas possuam condições de ofertar recursos tecnológicos e acesso à Internet com qualidade, os gestores devem desenvolver planos educacionais que justifiquem a necessidade da existência desses recursos em suas escolas e ao governo cabe o provimento de recursos financeiros necessários para aquisição e manutenção desses equipamentos em toda rede de ensino.

Após realizadas essas análises com base nos resultados dos artigos desenvolvidos como estrutura dessa tese, os fatores que estão associados aos melhores desempenhos em avaliações de larga escala do SAEB são:

- Alunos – (nível de escolaridade dos pais ou responsáveis pelos alunos);
- Professores – (tempo de experiência na função, atualizações em habilidades e competências relacionadas à função docente);
- Diretores – (tempo de experiência do Diretor de Escola no exercício da função, nível de pós-graduação e ações para controle de reprovações dos alunos);
- Escolas – (acesso a computadores por alunos e professores, nível de acesso à internet e se a escola tem internet banda larga).

A Figura 7 apresenta graficamente os fatores encontrados.



Fonte: Elaborado pelo Autor (2020).

Após análises realizadas, os resultados apresentados por esse estudo atenderam plenamente aos objetivos específicos obedecendo a seguinte configuração: os Artigos 1,2 e 3 atenderam plenamente aos objetivos específicos A e B; o Artigo 4 atendeu plenamente aos objetivos específicos B e C; o Artigo 5 atendeu plenamente aos objetivos específicos B e C e, finalmente, o Artigo 6 atendeu plenamente aos objetivos específicos B e C.

8 CONSIDERAÇÕES FINAIS

O objetivo desse estudo foi analisar por meio da Mineração de Dados Educacionais, quais fatores presentes no ambiente educacional puderam ser preditores para que os alunos, egressos do Ensino Médio, obtivessem um bom desempenho nas avaliações de larga escala do SAEB. Foram analisados os resultados das edições do SAEB dos anos de 2015 e 2017. Não foi possível analisar os resultados da edição de 2019, por que até o momento do encerramento desse estudo o governo ainda não tinha disponibilizado os resultados dos desempenhos individuais dos alunos.

Como metodologia, após a coleta e tratamento de dados das edições do SAEB, foram utilizados, em primeiro momento, análises por meio das técnicas da Estatística Descritiva, e como fase complementar, envolvendo a Mineração de Dados Educacionais, foram utilizadas as técnicas baseadas em Árvores de decisão, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes*.

O estudo partiu de 4 proposições, sendo que algumas não foram totalmente confirmadas.

A P1, que previa que os alunos que são incentivados e orientados, por seus pais ou responsáveis, em relação ao compromisso com os estudos, durante o ciclo escolar, podem apresentar melhor nível de aprendizado e de desempenho escolar, foi plenamente confirmada pelos resultados da mineração encontrados no Artigo 2.

Já P2 foi parcialmente confirmada pois previa que, Professores comprometidos com o nível de aprendizagem dos alunos, têm mais condições de desenvolver práticas pedagógicas que proporcionem melhorias no ambiente de ensino e no desempenho escolar. Nos resultados encontrados pela Mineração de Dados, apresentados no Artigo 4, indicaram que o Tempo de Experiência no exercício da função se mostrou com maior peso na associação com o desempenho escolar, além do compromisso com as práticas pedagógicas,

Em relação a proposição P3, foi previsto que Diretores de escola que são comprometidos com as melhorias das práticas de gestão e com os processos de ensino-aprendizagem, conseguem desenvolver ambientes mais adequados para as práticas de ensino, também foi parcialmente confirmada, pois os resultados encontrados pela Mineração de Dados, Artigo 5, indicaram que o Tempo de experiência no exercício da função apresentou maior peso na associação com o

desempenho escolar, seguido de atributos relacionados com as práticas de gestão e nível formação em cursos de pós graduação.

E finalmente a proposição P4, que previa que o ambiente escolar adequado para as práticas de ensino e que possa proporcionar acesso a recursos tecnológicos e de Internet, apresentam melhores condições para o desenvolvimento do nível de conhecimento, e de aprendizado, dos alunos, foi plenamente confirmada pelos resultados da Mineração de Dados apresentados no Artigo 6.

Os resultados apresentados no Capítulo Resultados em Artigos, indicaram que os fatores que estão mais associados aos melhores desempenhos dos alunos em avaliações do SAEB são: Nível de escolaridade dos pais ou responsáveis pelos alunos, tempo de experiência no exercício da função, atualizações em habilidades e competências relacionadas à função docente, tempo de experiência do Diretor de Escola no exercício da função, nível de pós-graduação e ações para controle de reprovações dos alunos, acesso a computadores por alunos e professores, nível de acesso à internet e se a escola tem internet banda larga. Ressaltando que esses resultados convergem positivamente para a literatura analisada.

Em relação à análise do desempenho dos alunos, o alto índice de notas baixas observado reforça a necessidade de ações que promovam melhorias no modelo de ensino-aprendizagem, de forma que os alunos possam desenvolver um nível melhor de conhecimento e aprendizado, para que em avaliações futuras, menos alunos estejam classificados no Nível Zero da escala de desempenho e que a maioria dos alunos apresentem classificações compatíveis com os Níveis mais altos da escala.

Essa pesquisa contribuiu com informações importantes sobre quais são os fatores mais relevantes que estão associados ao desempenho escolar, inclusive por ordem de importância, para que, por meio de debates envolvendo os profissionais da educação, ações de melhorias possam ser desenvolvidas e implementadas no sistema educacional.

As limitações apresentadas por esse trabalho estão relacionadas com a qualidade de preenchimento dos formulários pelos alunos, professores, Diretores e Representantes de Escola. Muitos formulários foram descartados por que estavam preenchidos errados ou em branco. Vários registros da base de dados dos alunos não apresentavam notas para as proficiências, ou ainda faltavam informações em

outros campos que eram foco desse estudo, e por conta disso também foram descartados.

Como proposta para novos estudos, seria investigar fatores associados ao desempenho escolar, dos alunos que participaram de avaliações em larga escala, referentes aos cursos do Ensino Fundamental de 5º ou 9º anos do SAEB. Além disso, como outra sugestão, seria analisar os resultados das avaliações estaduais para os níveis da educação básica. E finalmente, já em nível internacional, observar os resultados dos alunos que participaram das avaliações do PISA.

REFERÊNCIAS

ALDRUP, K. et al. Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. **Journal of Educational Psychology**, v. 110, n. 8, p. 1066–1083, 2018.

ALVES, T.; SILVA, R. M. Estratificação das oportunidades educacionais no Brasil: contextos e desafios para a oferta de ensino em condições de qualidade para todos. **Educação & Sociedade**, v. 34, n. 124, p. 851–879, set. 2013.

AMÂNCIO-VIEIRA, S. F. et al. A Relação entre Custo Direto e Desempenho Escolar: Uma Análise Multivariada nas Escolas de Ensino Fundamental de Londrina/PR. **Educação em Revista**, v. 31, n. 1, p. 169–194, mar. 2015.

AMÂNCIO-VIEIRA, S. F.; PEREIRA, R. S. Custos diretos, nível socioeconômico, estrutura escolar e desempenho educacional: uma análise das escolas municipais de São José dos Pinhais/PR. **Anais do Congresso Brasileiro de Custos - ABC**, v. 0, 2016.

AMARAL, N. C. Os 10% do PIB como promotor da qualidade da educação: uma análise considerando os resultados do PISA e os valores aplicados por estudante em diversos países. **Rede Nacional de Ciência para a Educação (Rede CpE)**, 2015.

ANDRADE, J. M.; LAROS, J. A. Fatores associados ao desempenho escolar: estudo multinível com dados do SAEB/2001. **Psicologia: Teoria e Pesquisa**, v. 23, n. 1, p. 33–41, mar. 2007.

ANDRADE, M.; FRANCO, C.; DE CARVALHO, J. P. Gênero e Desempenho em Matemática ao final do Ensino Médio: Quais as relações? **Anais**, p. 1–16, 2016.

BACK, L. T. et al. Classroom management, school staff relations, school climate, and academic achievement: testing a model with urban high schools. **Learning Environments Research**, v. 19, n. 3, p. 397–410, 1 out. 2016.

BAKER, R. S. J. D; YACEF, K. The State of Educational Data Mining in 2009: A Review and Future Visions. **JEDM | Journal of Educational Data Mining**, v. 1, n. 1, p. 3–17, 1 out. 2009.

BAKER, R. S.; INVENTADO, P. S. Educational Data Mining and Learning Analytics. In: LARUSSON, J. A.; WHITE, B. (Eds.). . **Learning Analytics: From Research to Practice**. New York, NY: Springer, 2014. p. 61–75.

BARBOSA, S. G.; WILHELM, V. E. Evaluation of the performance of public schools through Data Envelopment Analysis/Avaliação do desempenho das escolas públicas por meio de Data Envelopment Analysis. **Acta Scientiarum. Technology**, 2009.

BARROS, Á. G. et al. A RELEVÂNCIA DA FORMAÇÃO CONTINUADA PARA OS GESTORES ESCOLARES DA REDE PÚBLICA. **LINKSCIENCEPLACE - Interdisciplinary Scientific Journal**, v. 6, n. 3, 6 nov. 2019.

BEZERRA, C. et al. Evasão Escolar: Aplicando Mineração de Dados para Identificar Variáveis Relevantes. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 27, n. 1, p. 1096, 7 nov. 2016.

BEZERRA, L. N. M.; SILVA, M. T. Educational Data Mining Applied to a Massive Course. **International Journal of Distance Education Technologies (IJDET)**, v. 18, n. 4, 1 out. 2020.

BEZERRA, L. N. MENDES. **Mineração de Dados Educacionais para a gestão de cursos massivos**. Tese de Doutorado—São Paulo: Universidade Paulista - UNIP, 2017.

BEZERRA, M. G.; KASSOUF, A. L. **Análise de fatores que afetam o desempenho escolar nas escolas das áreas urbanas e rurais do Brasil**. XLIV CONGRESSO DA SOBER - Questões agrárias, educação no campo e desenvolvimento. **Anais...** In: XLIV CONGRESSO DA SOBER. Fortaleza-CE: Sociedade Brasileira de Economia e Sociologia Rural, 2006 Disponível em: <<https://bdpi.usp.br/item/001536729>>. Acesso em: 28 set. 2019

BOUZADA, M. A. C.; RIBEIRO, L. O. M.; PEIXE, J. B. **Métodos Quantitativos Aplicados a Casos Reais - 1st Edition**. Rio de Janeiro: Elsevier Editora Ltda, 2013.

BRASIL. **Lei nº 9.394, de 20 de dezembro de 1996, Estabelece as diretrizes e bases da educação nacional**. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L9394.htm>.

BRASIL. **Lei nº 13005, de 25 de julho de 2014. Estabelece o Plano nacional de educação - PNE**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l13005.htm>. Acesso em: 28 abr. 2019.

BRASIL. **Tesouro divulga o relatório Aspectos Fiscais da Educação no Brasil**. Tesouro Nacional, , 2018. Disponível em: <<http://www.tesouro.fazenda.gov.br/-/tesouro-divulga-o-relatorio-aspectos-fiscais-da-educacao-no-brasil>>. Acesso em: 5 out. 2019

BRITO, M. V. S. **Determinantes do sucesso escolar: uma análise multinível a partir dos dados do Pisa 2015**. Dissertação de Mestrado—Brasília: Universidade de Brasília, 23 jun. 2017.

BROOKE, N.; SOARES, J. F. Pesquisa em eficácia escolar: origem e trajetória. **Est. Aval. Educ**, v. 22, p. 593–598, 2011.

CALIXTO, K.; SEGUNDO, C.; GUSMÃO, R. P. Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 28, n. 1, p. 1447, 27 out. 2017.

CAMPOS, H. R.; JORGE, S. D. C. Violência na escola: uma reflexão sobre o bullying e a prática educativa. **Em Aberto**, v. 23, n. 83, p. 22, 2010.

CÁRIA, N. P.; LAMBERT-DE-ANDRADE, N. Democratic Management at school: in search of participation and leadership. **Revista Eletrônica de Educação**, v. 10, n. 3, p. 25–40, 30 nov. 2016.

CARVALHO, L. A. V. D. **Datamining - A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Rio de Janeiro: Editora Ciência Moderna, 2005.

CASTILLO-VERGARA, M. et al. Does socioeconomic status influence student creativity? **Thinking Skills and Creativity**, v. 29, p. 142–152, 1 set. 2018.

CASTRO, J. A.; CARVALHO, C. H. A. Necessidades e possibilidades para o financiamento da educação brasileira no plano nacional de educação. **Educação & Sociedade**, v. 34, n. 124, p. 829–849, set. 2013.

CASTRO, L. N.; FERRARI, D. G. **Introdução a mineração de dados**. 1ª edição ed. [s.l.] SARAIVA, 2017.

CHIRINÉA, A. M.; BARREIRO, I. M. F. Qualidade da educação: eficiência, eficácia e produtividade escolar. **Revista on line de Política e Gestão Educacional**, v. 0, n. 7, 2009.

CILIATO, F. L. G.; SARTORI, J. Pluralidade cultural: os desafios aos professores em frente da diversidade cultural. **Revista Monografias Ambientais**, v. 14, n. 0, p. 65–78, 7 dez. 2015.

COLPANI, R. Mineração de Dados Educacionais: um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar. **Informática na educação: teoria & prática**, v. 21, n. 3, 2018.

CORREIA-ZANINI, M. R. G. **Um estudo prospectivo sobre o percurso escolar de crianças nos primeiros anos do Ensino Fundamental**. Tese de Doutorado—São Paulo: Universidade de São Paulo, 6 dez. 2013.

CUNHA, J. M. P. et al. Social segregation and academic achievement in state-run elementary schools in the municipality of Campinas, Brazil. **Geoforum**, Themed Issue: Land, Labor, Livestock and (Neo)Liberalism: Understanding the Geographies of Pastoralism and Ranching. v. 40, n. 5, p. 873–883, 1 set. 2009.

DALLAZEM, A.; COELHO, V. R. O Desempenho Escolar na voz dos atores de escolas públicas catarinenses. **Revista Teias**, v. 20, n. 56, p. 398–417, 26 mar. 2019.

DAMON, A. et al. What education policies and programmes affect learning and time in school in developing countries? A review of evaluations from 1990 to 2014. **Review of Education**, v. 7, n. 2, p. 295–387, 2019.

DÍAZ, L. M. B.; GARCÍA, M. L. S.; CANO, E. V. Effects on academic performance in secondary students according to the use of ICT. **IJERI: International Journal of Educational Research and Innovation**, n. 12, p. 90–108, 14 jun. 2019.

DOURADO, L. F.; OLIVEIRA, J. F.; SANTOS, C. A. A Qualidade da Educação: conceitos e definições. **Textos para Discussão**. Brasília: Inep/MEC, 2007.

DRABACH, N. P. **Processos de provimento, exigências e atribuições para os dirigentes escolares em contextos de reformas gerencialistas**. Tese de Doutorado—Campinas-São Paulo: UNICAMP, 2018.

DWYER, T. et al. Desvendando mitos: os computadores e o desempenho no sistema escolar. **Educação & Sociedade**, v. 28, n. 101, p. 1303–1328, dez. 2007.

FANTINATO, A. C.; CIA, F. Envolvimento parental, competência social e o desempenho acadêmico de escolares. **Psicologia Argumento**, v. 29, n. 67, 24 nov. 2017.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Advances in Knowledge Discovery and Data Mining. In: FAYYAD, U. M. et al. (Eds.). . Menlo Park, CA, USA: American Association for Artificial Intelligence, 1996. p. 1–34.

FELÍCIO, F.; FERNANDES, R. O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental no Estado de São Paulo. **Anais - ANPEC**, 2005.

FERNANDES, L. M. et al. Preditores do Desempenho Escolar ao final do Ensino Fundamental: Histórico de Reprovação, Habilidades Sociais e Apoio Social. **Trends in Psychology**, v. 26, n. 1, p. 215–228, mar. 2018.

FERNANDES, R. Índice de Desenvolvimento da Educação Básica (Ideb). **Textos para Discussão**, p. 29. Brasília: Inep/MEC, 2007.

FILHO, N. M.; NUÑEZ, D. F.; RIBEIRO, F. P. Comparando as escolas paulistas com melhor e pior desempenho no SARESP e na Prova Brasil. **São Paulo em Perspectiva**, v. 23, n. 1, p. 20, 2009.

FINI, M. INÊS. Currículo e Avaliação, articulação necessária em favor da aprendizagem dos alunos da rede pública de São Paulo. **Revista São Paulo em Perspectiva - Fundação Seade**, 2009.

FONSECA, S. O.; NAMEN, A. A. Mineração em Bases de Dados do INEP: Uma Análise Exploratória para Nortear Melhorias no Sistema Educacional Brasileiro. **Educação em Revista**, v. 32, n. 1, p. 133–157, mar. 2016.

FORD, T. G. Can the use of informal control mechanisms increase trust among teachers? An evaluation of the Accelerated Schools intervention. **Studies in Educational Evaluation**, v. 63, p. 59–71, 1 dez. 2019.

FRANCO, A. M. P.; MENEZES FILHO, N. Uma análise de rankings de escolas brasileiras com dados do SAEB. **Estudos Econômicos (São Paulo)**, v. 42, n. 2, p. 263–283, jun. 2012.

FREITAS, E. D. S. **Práticas de gestão do conhecimento como variáveis intervenientes nos resultados do IDEB**. Dissertação de Mestrado—Maringá-PR: Centro Universitário de Maringá - UNICESUMAR, 2019.

GARCIA, P. S. et al. A infraestrutura das escolas de ensino fundamental da Região do Grande ABC paulista. **Revista Ibero-Americana de Estudos em Educação**, v. 9, n. 3, p. 614–631, 2014.

GARCIA, P. S. et al. Desempenho escolar: uma análise do IDEB dos municípios da região do ABC. **Revista Eletrônica de Educação**, v. 10, n. 2, p. 95–114, 31 ago. 2016.

GATTI, B. Avaliação de sistemas educacionais no Brasil. **Sísifo. Revista de Ciências da Educação**, p. 7–18, 2009.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2010.

GIVISIEZ, G. H. N.; OLIVEIRA, E. L. Royalties do petróleo e educação: análise da eficiência da alocação. **Revista Brasileira de Pós-Graduação**, v. 8, n. 0, p. 1–13, 2011.

GÓIS, L. S.; ROCHA, G. S. A atuação do Gestor Escolar no enfrentamento da evasão no Instituto Federal da Bahia(IFBA). **Revista Eletrônica Científica Ensino Interdisciplinar**, v. 5, n. 14, p. 340–361, 7 out. 2019.

GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data mining: conceitos, técnicas, algoritmos, orientações e aplicações**. 2ª ed. Rio de Janeiro: Elsevier, 2015.

GONÇALVES, F. O.; FRANÇA, M. T. A. Transmissão intergeracional de desigualdade e qualidade educacional: avaliando o sistema educacional brasileiro a partir do SAEB 2003. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 16, n. 61, p. 639–662, dez. 2008.

GUSTAFSSON, J.-E.; NILSEN, T.; HANSEN, K. Y. School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. **Studies in Educational Evaluation**, Research on Equity in Education: Implications for School Evaluation. v. 57, p. 16–30, 1 jun. 2018.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)**. 3ª ed. [s.l.] Morgan Kaufmann, 2012.

HANUSHEK, E. A.; WOESSMANN, L. School Resources and Student Achievement: A Review of Cross-Country Economic Research. In: ROSÉN, M.; YANG HANSEN, K.; WOLFF, U. (Eds.). **Cognitive Abilities and Educational Outcomes: A Festschrift in Honour of Jan-Eric Gustafsson**. de Metodologia de Medição e Avaliação Educacional. Cham: Springer International Publishing, 2017. p. 149–171.

HOLFMANN, M.; KLINKENBERG, R. **RapidMiner: Data Mining Use Cases and Business Analytics Applications**. 1ª ed. London: Chapman and Hall/CRC, 2013.

INEP. **IDEB - SAEB**. Disponível em: <<http://portal.inep.gov.br/web/guest/educacao-basica/saeb>>. Acesso em: 29 mar. 2020a.

INEP. **Censo Escolar**. Disponível em: <<http://portal.inep.gov.br/web/guest/censo-escolar>>. Acesso em: 29 abr. 2020b.

INEP. **SAEB - Microdados**. Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 29 mar. 2020c.

INEP. **PNE - Monitoramento das Metas.** Disponível em: <http://simec.mec.gov.br/pde/grafico_pne.php>. Acesso em: 29 mar. 2020d.

INEP. **Indicadores Financeiros Educacionais.** Disponível em: <<http://inep.gov.br/indicadores-financeiros-educacionais>>. Acesso em: 29 mar. 2020e.

INEP. **Pisa em Foco.** Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/pisa-em-foco>>. Acesso em: 14 nov. 2020f.

INEP. **PISA.** Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa>>. Acesso em: 29 mar. 2020g.

INEP. **Dados Abertos - Educação Brasileira.** Disponível em: <<http://portal.inep.gov.br/web/guest/dados>>. Acesso em: 14 nov. 2020h.

INEP. **Indicadores Educacionais.** Disponível em: <<http://portal.inep.gov.br/indicadores-educacionais>>. Acesso em: 29 mar. 2020i.

INEP. **Sinopses Estatísticas - INEP.** Disponível em: <<http://portal.inep.gov.br/web/guest/sinopses-estatisticas>>. Acesso em: 14 nov. 2020j.

JOVARINI, N. V.; LEME, V. B. R.; CORREIA-ZANINI, M. R. G. Influência das Habilidades Sociais e Estressores sobre o Desempenho Escolar no 6º ano. **Paidéia (Ribeirão Preto)**, v. 28, 2018.

JÚNIOR, O. G. F. et al. Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de Maceió. **RENOTE**, v. 17, n. 1, p. 296–305, 28 jul. 2019.

LACRUZ, A. J. et al. Indicadores de qualidade na educação: análise discriminante dos desempenhos na Prova Brasil. **Revista Brasileira de Educação**, v. 24, 2019.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica.** São Paulo: Atlas, 2017.

LAROS, J. A.; MARCIANO, J. L.P.; ANDRADE, J. M. Fatores associados ao desempenho escolar em Português: um estudo multinível por regiões. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 20, n. 77, p. 623–646, dez. 2012.

LAROS, J. A.; MARCIANO, J. L. P.; ANDRADE, J. M. Fatores que afetam o desempenho na prova de Matemática do SAEB: um estudo multinível. **Avaliação Psicológica**, v. 9, p. 14, 2010.

LASATER, K.; BENGTON, E.; ALBILADI, W. S. Data use for equity?: How data practices incite deficit thinking in schools. **Studies in Educational Evaluation**, p. 100845, 15 fev. 2020.

LAUERMANN, F. Teacher motivation, responsibility, pedagogical knowledge and professionalism: a new era for research. **Pedagogical Knowledge and the Changing Nature of the Teaching Profession**, p. 171–191, 21 fev. 2017.

LOBO, G. D. Avaliação do Desempenho Escolar dos Estudantes da Região Nordeste que Realizaram o ENEM: Uma Análise com Modelos Hierárquicos. **Revista Espacios**, v. 38, p. 13, 2017.

LUZ, L. S. **Os Determinantes do Desempenho Escolar: A Estratificação Educacional e o Efeito Valor Adicionado**. . In: XV ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, ABEP. Caxambu-MG: 2006Disponível em: <<http://www.abep.org.br/publicacoes/index.php/anais/article/view/1693>>

MACHEBE, C. H.; EZEGBE, B. N.; ONUOHA, J. C. The Impact of Parental Level of Income on Students' Academic Performance in High School in Japan. **Universal Journal of Educational Research**, v. 5, p. 1614–1620, 2017.

MARANGONI, R. A. O Trabalho do Diretor de Escola: Análise a partir de uma perspectiva histórica. **Jornal de Políticas Educacionais**, v. 14, n. 0, 26 abr. 2020.

MARTURANO, E. M.; PIZATO, E. C. G. Preditores de Desempenho Escolar no 5º Ano do Ensino Fundamental. **PSICO**, v. 46, 2015.

MASCI, C.; DE WITTE, K.; AGASISTI, T. The influence of school size, principal characteristics and school management practices on educational performance: An efficiency analysis of Italian students attending middle schools. **Socio-Economic Planning Sciences**, Recent developments on the use of DEA in the public sector. v. 61, p. 52–69, 1 mar. 2018.

MEC. **PNE - Plano Nacional de Educação - Plano Nacional de Educação - Lei nº 13.005/2014**. Disponível em: <<http://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>>.

MESQUITA, S. S. A. Referenciais do “bom professor” de ensino médio: exercício de articulação teórica. **Caderno Pesquisa**, v. 48, 2018.

MINAYO, M. C. S. (OR). **Pesquisa social: teoria, método e criatividade**. 29. ed. Petrópolis/RJ: Vozes, 2010.

MOHR, A. C.; NAUJORKS, M. I. Políticas Educacionais de correção de fluxo no RS: Efeitos da racionalidade neoliberal na Gestão Escolar. **Reflexão e Ação**, v. 25, n. 1, p. 160–177, 2 jan. 2017.

MORAIS, J. S.; NASCIMENTO, F.L. S. C.; MAGALHÃES, N. R. S. Evocações narrativas de gestores escolares acerca das políticas de saberes e fazeres da gestão. **Educação, Ciência e Cultura**, v. 25, n. 2, p. 221–237, 5 ago. 2020.

MURPHY, S. School location and socioeconomic status and patterns of participation and achievement in senior secondary mathematics. **Mathematics Education Research Journal**, v. 31, n. 3, p. 219–235, 1 set. 2019.

NASCIMENTO, M. C.; GOMES, G. R. R. Formação continuada docente para a utilização das TIC no processo de ensino e aprendizagem. **Research, Society and Development**, v. 9, n. 2, p. e33921998–e33921998, 1 jan. 2020.

NASCIMENTO, V. P.; CHIUSOLI, C. L. O papel do gestor escolar: estudo de caso sobre os desafios da educação pública. **Revista on line de Política e Gestão Educacional**, v. 23, n. 2, p. 238–254, 6 mar. 2019.

NICHELE, P. T.; MELLO, M. A. S. Gestão Escolar na perspectiva da educação democrático-participativa e a função social da escola. **Revista Saberes Pedagógicos**, v. 4, n. 3, p. 323–343, 29 set. 2020.

OCDE. **PISA - PISA**. Disponível em: <<http://www.oecd.org/pisa/>>. Acesso em: 29 set. 2020.

OLIVEIRA, A. C. P.; WALDHELM, A. P. S. Liderança do diretor, clima escolar e desempenho dos alunos: qual a relação? **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 24, n. 93, p. 824–844, dez. 2016.

ONU. **Plataforma Agenda 2030**. Disponível em: <<http://www.agenda2030.com.br/>>. Acesso em: 28 set. 2019.

ORTIGÃO, M. I. R. Avaliação e Políticas Públicas: possibilidades e desafios para a Educação Matemática. **BOLEMA: Boletim de Educação Matemática**, 2008.

PAULA, J. S.; FRANCO, A. M. P.; SILVA, J. W. Fatores relacionados ao atraso escolar no estado de Minas Gerais. **Estudos em Avaliação Educacional**, v. 29, n. 72, p. 886–917, 20 dez. 2018.

PEÑA-AYALA, A. (ED.). **Learning Analytics: Fundaments, Applications, and Trends : A View of the Current State of the Art to Enhance e-Learning: 94**. 1. ed. New York, NY: Springer, 2017.

PEREIRA, I. A. Tipos de liderança dentro do ambiente educativo: uma discussão sobre sua identidade organizacional. **Educationis**, v. 8, n. 1, p. 8–22, 28 fev. 2020.

PEREIRA, M. J.; MORI, N. N. R. Diretrizes curriculares e o desempenho de alunos paranaenses da 8ª série do ensino fundamental na Prova Brasil. **RBPG. Revista Brasileira de Pós-Graduação**, v. 8, p. p121- 145. 25p., 2012.

PEREIRA, P. R.; GASQUE, K. C. G. D. Tomada de decisão do gestor escolar das escolas públicas de ensino médio no Distrito Federal e a interface com o letramento informacional. **Em Questão**, v. 25, n. 3, p. 40–60, 7 ago. 2019.

PHILIPPSSEN, M. R. **Avaliação externa do desempenho escolar e seus impactos na prática pedagógica e na gestão escolar: estudo de caso em uma escola pública de Porto Alegre (RS)**. Monografia—Rio Grande do Sul: Universidade Federal de Santa Maria, 1 dez. 2018.

PINTO, J. M. R. Remuneração adequada do professor: desafio à educação brasileira. **Retratos da Escola**, v. 3, n. 4, 6 maio 2012.

PROVOST, F.; FAWCETT, T. **Data Science Para Negócios**. 1ª Edição ed. Rio de Janeiro: Editora Alta Books, 2016.

QUEIROZ, M. V. A. B.; SAMPAIO, R. M. B.; SAMPAIO, L. M. B. Dynamic efficiency of primary education in Brazil: Socioeconomic and infrastructure influence on school performance. **Socio-Economic Planning Sciences**, v. 70, p. 100738, 1 jun. 2020.

RAZAK, N. A. et al. Successful implementation of information and communication technology integration in Malaysian public schools: An activity systems analysis approach. **Studies in Educational Evaluation**, v. 58, p. 17–29, 1 set. 2018.

REZENDE PINTO, J. M.; AMARAL, N. C.; CASTRO, J. A. O financiamento do Ensino Médio no Brasil: de uma escola boa para poucos à massificação barata da rede pública. **Educação & Sociedade**, v. 32, n. 116, p. 639–665, set. 2011.

RIBEIRO, M. D.; SÍVERES, L.; BRITO, R. O. A Formação de gestores escolares: a dimensão ética em questão. **Educação Por Escrito**, v. 10, n. 2, p. e36685–e36685, 2019.

RIDDLE, S.; APPLE, M. W. **Re-imagining Education for Democracy**. 1. ed. London: Routledge, 2019.

RODRIGUES, C. G.; RIOS-NETO, E. L. G.; PINTO, C. C. X. Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico, 1997 a 2005. **Revista Brasileira de Estudos de População**, v. 28, n. 1, p. 5–36, jun. 2011.

ROMERO, C.; VENTURA, S. Data mining in education. **WIREs Data Mining and Knowledge Discovery**, v. 3, n. 1, p. 12–27, 2013.

SALAM, S. et al. ICT and students' performance in Pakistan. **Human Systems Management**, v. 36, n. 4, p. 277–284, 1 jan. 2017.

SALES, F. et al. Evasão no Ensino Básico da Rede Pública Municipal de Juiz de Fora: uma Análise com Mineração de Dados. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, v. 30, n. 1, p. 1371, 11 nov. 2019.

SANTOS, A. F. Os Questionários Contextuais do SAEB 2013 e do PISA 2012: uma comparação dos elementos constituintes do Ensino. **Reuniões da ABAVE**, v. 0, n. 8, p. 453–486, 30 out. 2015.

SANTOS, R. S.; MONTANUCI, R.; PACHECO, E. J. Educação de qualidade e avaliação: Raízes históricas de uma correlação contemporânea. **Cadernos da FUCAMP**, v. 18, n. 35, 29 ago. 2019.

SCHAIK, P. VAN et al. Barriers and conditions for teachers' utilisation of academic knowledge. **International Journal of Educational Research**, v. 90, p. 50–63, 1 jan. 2018.

SCHILDKAMP, K. et al. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. **International Journal of Educational Research**, v. 103, p. 101602, 1 jan. 2020.

SCHNEIDER, G.; GOUVEIA, A. B. Qualidade da escola: uma proposta de índice para as condições materiais de escolas a partir de dados contextuais do Saeb. **Revista Brasileira de Pós-Graduação**, v. 8, n. 1, 31 dez. 2011.

SEBANC, A. M.; GUIMOND, A. B.; LUTGEN, J. Transactional Relationships Between Latinos' Friendship Quality and Academic Achievement During the Transition to Middle School. **The Journal of Early Adolescence**, v. 36, n. 1, p. 108–138, 1 jan. 2016.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. **Introdução á Mineração de Dados**. Rio de Janeiro: Elsevier, 2016.

SILVA, R. J. **Mineração de dados aplicada a análise de desempenho de alunos no 5º ano do ensino fundamental**. Monografia—Serra Talhada - PE: Universidade Federal Rural de Pernambuco, 2019.

SIMIELLI, L. E. R. Equidade e Oportunidades Educacionais: O Acesso a Professores no Brasil. **Arquivos Analíticos de Políticas Educativas**, v. 25, 2017.

SIMÕES, M. F.; FERRÃO, M. E. Competência percebida e desempenho escolar em Matemática. **Estudos em Avaliação Educacional**, v. 16, n. 32, p. 25–42, 2005.

SIMON, A.; CAZELLA, S. Mineração de Dados Educacionais nos Resultados do ENEM de 2015. **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**, v. 6, n. 1, p. 754, 27 out. 2017.

SOARES, C. S.; ROSA, F. S. Gastos públicos em educação e desempenho escolar: Análise qualitativa dos resultados do IDEB em Santa Maria/RS. **Revista Eletrônica de Estratégia & Negócios**, v. 13, n. 0, p. 38–62, 8 maio 2020.

SOARES, J. F.; ALVES, M. T. G. Desigualdades raciais no sistema brasileiro de educação básica. **Educação e Pesquisa**, v. 29, n. 1, p. 147–165, jun. 2003.

SOARES, T. M. et al. A expectativa do professor e o desempenho dos alunos. **Psicologia: Teoria e Pesquisa**, v. 26, n. 1, p. 157–170, mar. 2010.

SOUZA, L. D. M.; RIBEIRO, M. S. S. O Perfil do Gestor Escolar Contemporâneo: das permanências as incorporações para exercício da função. **Revista Espaço do Currículo**, v. 10, n. 1, 28 abr. 2017.

SOUZA, M. I. M. O fazer do gestor escolar: desafios e possibilidades de sua atuação profissional, enquanto facilitador do processo de ensino-aprendizagem. **Research, Society and Development**, v. 9, n. 7, p. e335973900–e335973900, 14 maio 2020.

SYMEONIDIS, V. Teacher competence frameworks in Hungary: A case study on the continuum of teacher learning. **European Journal of Education**, v. 54, n. 3, p. 400–412, 2019.

TARDA, K. B.; RODRIGUES, R. V. A Influência dos Gastos em Educação no Crescimento Econômico da Região Administrativa de Campinas. **Revista Estudo & Debate**, v. 22, n. 2, 23 dez. 2015.

TRAVITZKI, R.; FERRÃO, M. E.; COUTO, A. P. Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: Uma visão a partir da análise de dados do ENEM. **Education Policy Analysis Archives**, v. 24, p. 74, 11 jul. 2016.

VASCONCELOS, J. C. et al. School infrastructure and public investment in Education in Brazil: the importance of educational performance. **Ensaio: Avaliação e Políticas Públicas em Educação**, n. AHEAD, 2020.

VIDAL, E. M.; VIEIRA, S. L. Professores da educação básica: perfil e percepções sobre sucesso dos alunos. **Estudos em Avaliação Educacional**, v. 28, p. 64–101, 2017.

VIEIRA, A. E. R.; BUSSOLOTTI, J. M. Gestão Escolar: um estudo de caso sobre Escolas Técnicas. **Interação - Revista de Ensino, Pesquisa e Extensão**, v. 20, n. 1, p. 45–70, 2018.

VOINEA, M. Rethinking Teacher Training According to 21st Century Competences. **European Journal of Multidisciplinary Studies**, v. 4, n. 3, p. 20–26, 25 set. 2019.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3ª ed. [s.l.] Morgan Kaufmann, 2011.

ZAFANI, M. D.; OMOTE, S. Atribuições Da Família Na Percepção Do Professor. **Journal of Research in Special Educational Needs**, v. 16, n. S1, p. 221–224, 2016.

ZAKARIYA, Y. F.; BJØRKESTØL, K.; NILSEN, H. K. Teacher job satisfaction across 38 countries and economies: An alignment optimization approach to a cross-cultural mean comparison. **International Journal of Educational Research**, v. 101, p. 101573, 1 jan. 2020.

ZANBELLO, B. L. et al. **Administração Escolar: do Diretor ao Gestor**. Anais Eletrônico do XI EPCC - Encontro Internacional de Produção Científica. **Anais...** In: ENCONTRO INTERNACIONAL DE PRODUÇÃO CIENTÍFICA UNICESUMAR. Maringá-PR: 29 out. 2019 Disponível em: <<http://rdu.unicesumar.edu.br/handle/123456789/3752>>. Acesso em: 3 set. 2020

ANEXOS

Anexo 1 – Plano Nacional da Educação – Metas

Plano Nacional da Educação - Metas

Meta	Descrição
1	Universalizar, até 2016, a educação infantil na pré-escola para as crianças de 4 (quatro) a 5 (cinco) anos de idade e ampliar a oferta de educação infantil em creches, de forma a atender, no mínimo, 50% (cinquenta por cento) das crianças de até 3 (três) anos até o final da vigência deste PNE.
2	Universalizar o ensino fundamental de 9 (nove) anos para toda a população de 6 (seis) a 14 (quatorze) anos e garantir que pelo menos 95% (noventa e cinco por cento) dos alunos conclua essa etapa na idade recomendada, até o último ano de vigência deste PNE.
3	Universalizar, até 2016, o atendimento escolar para toda a população de 15 (quinze) a 17 (dezessete) anos e elevar, até o final do período de vigência deste PNE, a taxa líquida de matrículas no ensino médio para 85% (oitenta e cinco por cento).
4	Universalizar, para a população de 4 (quatro) a 17 (dezessete) anos com deficiência, transtornos globais do desenvolvimento e altas habilidades ou superdotação, o acesso à educação básica e ao atendimento educacional especializado, preferencialmente na rede regular de ensino, com a garantia de sistema educacional inclusivo, de salas de recursos multifuncionais, classes, escolas ou serviços especializados, públicos ou conveniados.
5	Alfabetizar todas as crianças, no máximo, até o final do 3º (terceiro) ano do ensino fundamental.
6	Oferecer educação em tempo integral em, no mínimo, 50% (cinquenta por cento) das escolas públicas, de forma a atender, pelo menos, 25% (vinte e cinco por cento) dos(as) alunos(as) da educação básica.
7	Fomentar a qualidade da educação básica em todas as etapas e modalidades, com melhoria do fluxo escolar e da aprendizagem, de modo a atingir as seguintes médias nacionais para o IDEB: 6,0 nos anos iniciais do ensino fundamental; 5,5 nos anos finais do ensino fundamental; 5,2 no ensino médio.
8	Elevar a escolaridade média da população de 18 (dezoito) a 29 (vinte e nove) anos, de modo a alcançar, no mínimo, 12 (doze) anos de estudo no último ano de vigência deste plano, para as populações do campo, da região de menor escolaridade no País e dos 25% (vinte e cinco por cento) mais pobres, e igualar a escolaridade média entre negros e não negros declarados à Fundação Instituto Brasileiro de Geografia e Estatística – IBGE.
9	Elevar a taxa de alfabetização da população com 15 (quinze) anos ou mais para 93,5% (noventa e três inteiros e cinco décimos por cento) até 2015 e, até o final da vigência deste PNE, erradicar o analfabetismo absoluto e reduzir em 50% (cinquenta por cento) a taxa de analfabetismo funcional.
10	Oferecer, no mínimo, 25% (vinte e cinco por cento) das matrículas de educação de jovens e adultos, nos ensinos fundamental e médio, na forma integrada à educação profissional.
11	Triplicar as matrículas da educação profissional técnica de nível médio, assegurando a qualidade da oferta e pelo menos 50% (cinquenta por cento) da expansão no segmento público.
12	Elevar a taxa bruta de matrícula na educação superior para 50% (cinquenta por cento) e a taxa líquida para 33% (trinta e três por cento) da população de 18 (dezoito) a 24 (vinte e quatro) anos, assegurada a qualidade da oferta e expansão para, pelo menos, 40% (quarenta por cento) das novas matrículas, no segmento público.
13	Elevar a qualidade da educação superior e ampliar a proporção de mestres e doutores do corpo docente em efetivo exercício no conjunto do sistema de educação superior para 75% (setenta e cinco por cento), sendo, do total, no mínimo, 35% (trinta e cinco por cento) doutores.
14	Elevar gradualmente o número de matrículas na pós-graduação stricto sensu, de modo a atingir a titulação anual de 60.000 (sessenta mil) mestres e 25.000 (vinte e cinco mil)

	doutores.
15	Garantir, em regime de colaboração entre a União, os Estados, o Distrito Federal e os Municípios, no prazo de 1 (um) ano de vigência deste PNE, política nacional de formação dos profissionais da educação de que tratam os incisos I, II e III do caput do art. 61 da Lei nº 9.394, de 20 de dezembro de 1996, assegurado que todos os professores e as professoras da educação básica possuam formação específica de nível superior, obtida em curso de licenciatura na área de conhecimento em que atuam.
16	Formar, em nível de pós-graduação, 50% (cinquenta por cento) dos professores da educação básica, até o último ano de vigência deste PNE, e garantir a todos(as) os(as) profissionais da educação básica formação continuada em sua área de atuação, considerando as necessidades, demandas e contextualizações dos sistemas de ensino.
17	Valorizar os(as) profissionais do magistério das redes públicas de educação básica, de forma a equiparar seu rendimento médio ao dos(as) demais profissionais com escolaridade equivalente, até o final do sexto ano de vigência deste PNE.
18	Assegurar, no prazo de 2 (dois) anos, a existência de planos de carreira para os(as) profissionais da educação básica e superior pública de todos os sistemas de ensino e, para o plano de carreira dos(as) profissionais da educação básica pública, tomar como referência o piso salarial nacional profissional, definido em lei federal, nos termos do inciso VIII do art. 206 da Constituição Federal.
19	Assegurar condições, no prazo de 2 (dois) anos, para a efetivação da gestão democrática da educação, associada a critérios técnicos de mérito e desempenho e à consulta pública à comunidade escolar, no âmbito das escolas públicas, prevendo recursos e apoio técnico da União para tanto.
20	Ampliar o investimento público em educação pública de forma a atingir, no mínimo, o patamar de 7% (sete por cento) do Produto Interno Bruto (PIB) do País no 5º (quinto) ano de vigência desta Lei e, no mínimo, o equivalente a 10% (dez por cento) do PIB ao final do decênio.

Anexo 2 – Escalas de Desempenhos para as proficiências Língua Portuguesa e Matemática



ESCALA DE PROFICIÊNCIA DE LÍNGUA PORTUGUESA 3ª SÉRIE DO ENSINO MÉDIO

Nível ¹	Descrição do Nível
<p>Nível 1 Desempenho maior ou igual a 225 e menor que 250</p>	<p>Os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Identificar elementos da narrativa em história em quadrinhos. • Reconhecer a finalidade de recurso gráfico em artigos. • Reconhecer a relação de causa e consequência em lendas. • Inferir o sentido de palavra em letras de música e reportagens.
<p>Nível 2 Desempenho maior ou igual a 250 e menor que 275</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Reconhecer a ideia comum entre textos de gêneros diferentes e a ironia em tirinhas. • Reconhecer relações de sentido estabelecidas por conjunções ou locuções conjuntivas em letras de música e crônicas. • Reconhecer o uso de expressões características da linguagem (científica, profissional etc.) e a relação entre pronome e seu referente em artigos e reportagens. • Inferir o efeito de sentido da linguagem verbal e não verbal em notícias e charges.
<p>Nível 3 Desempenho maior ou igual a 275 e menor que 300</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Localizar informação explícita em artigos de opinião. • Identificar a finalidade de relatórios científicos. • Reconhecer relações de sentido marcadas por conjunções, a relação de causa e consequência e a relação entre o pronome e seu referente em fragmentos de romances. • Reconhecer o tema de uma crônica. • Reconhecer variantes linguísticas em artigos. • Reconhecer o sentido e o efeito de sentido produzido pelo uso de recursos morfosintáticos em contos, artigos e crônicas. • Reconhecer opiniões divergentes sobre o mesmo tema em diferentes textos. • Inferir informação, o sentido e o efeito de sentido produzido por expressão em reportagens e tirinhas.
<p>Nível 4 Desempenho maior ou igual a 300 e menor que 325</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Localizar informações explícitas em infográficos, reportagens, crônicas e artigos. • Identificar o argumento em contos. • Identificar a finalidade e a informação principal em notícias. • Reconhecer a relação entre os pronomes e seus referentes em contos. • Reconhecer elementos da narrativa em contos. • Reconhecer variantes linguísticas em contos, notícias e reportagens. • Reconhecer o efeito de sentido produzido pelo uso de recursos morfosintáticos em poemas. • Reconhecer ideia comum e opiniões divergentes sobre o mesmo tema na comparação entre diferentes textos. • Reconhecer ironia e efeito de humor em crônicas e entrevistas. • Reconhecer a relação de causa e consequência em piadas e fragmentos de romance.



Nível ¹	Descrição do Nível
<p>Nível 4 Desempenho maior ou igual a 300 e menor que 325</p>	<ul style="list-style-type: none"> • Comparar poemas que abordem o mesmo tema. • Diferenciar fato de opinião em contos, artigos e reportagens. • Diferenciar tese de argumentos em artigos, entrevistas e crônicas. • Inferir informação, sentido de expressão e o efeito de sentido decorrente do uso de recursos morfosintáticos em crônicas. • Inferir o sentido decorrente do uso de recursos gráficos em poemas. • Inferir o efeito de sentido da linguagem verbal e não verbal e o efeito de humor em tirinhas.
<p>Nível 5 Desempenho maior ou igual a 325 e menor que 350</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Localizar informação explícita em resumos. • Identificar a informação principal em reportagens. • Identificar elementos da narrativa e a relação entre argumento e ideia central em crônicas. • Reconhecer a finalidade de propagandas. • Reconhecer variantes linguísticas e o efeito de sentido de recursos gráficos em crônicas e artigos. • Reconhecer a relação de causa e consequência e relações de sentido marcadas por conjunções em reportagens, artigos e ensaios. • Reconhecer o tema em poemas. • Diferenciar fato de opinião em resenhas. • Inferir o sentido de palavras e expressões em piadas e letras de música. • Inferir informação em artigos; inferir o sentido de expressão em fragmentos de romances.
<p>Nível 6 Desempenho maior ou igual a 350 e menor que 375</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Reconhecer efeitos estilísticos em poemas. • Reconhecer ironia e efeitos de sentido decorrentes da repetição de palavras em sinopses. • Reconhecer opiniões distintas sobre o mesmo tema, na comparação entre diferentes textos. • Reconhecer finalidade e traços de humor em reportagens. • Reconhecer o efeito de sentido do humor em tirinhas. • Reconhecer o tema em contos e fragmentos de romances. • Reconhecer relação de sentido marcada por conjunção em crônicas. • Inferir informação e tema em reportagens, poemas, histórias em quadrinhos e tirinhas. • Inferir o sentido e o efeito de sentido de palavras ou de expressão em poemas, crônicas e fragmentos de romances.



Nível ¹	Descrição do Nível
<p>Nível 7 Desempenho maior ou igual a 375 e menor que 400</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <ul style="list-style-type: none"> • Identificar a ideia central e o argumento em apresentações de livros, reportagens, editoriais e crônicas. • Identificar elementos da narrativa em crônicas, contos e fragmentos de romances. • Identificar ironia e tema em poemas e artigos. • Reconhecer relações de sentido marcadas por conjunção em artigos, reportagens e fragmentos de romances. • Reconhecer a relação de causa e consequência em reportagens e fragmentos de romances. • Reconhecer o efeito de sentido de recursos gráficos em artigos. • Reconhecer variantes linguísticas em letras de música e piadas. • Reconhecer a finalidade de reportagens, resenhas e artigos. • Inferir efeito de humor e ironia em tirinhas e charges.
<p>Nível 8 Desempenho maior ou igual a 400</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>Reconhecer o efeito de sentido resultante do uso de recursos morfossintáticos em artigos e letras de música.</p>

¹ A Prova Brasil não utilizou itens da 3ª série do Ensino Médio que avaliam as habilidades do Nível 0. Os estudantes da 3ª série com desempenho menor que 225 requerem atenção especial, pois ainda não demonstram habilidades muito elementares que deveriam apresentar nessa etapa escolar.



ESCALA DE PROFICIÊNCIA DE MATEMÁTICA 3ª SÉRIE DO ENSINO MÉDIO

Nível ¹	Descrição do Nível
<p style="text-align: center;">Nível 1 Desempenho maior ou igual a 225 e menor que 250</p>	<p>Os estudantes provavelmente são capazes de:</p> <p>TRATAMENTO DE INFORMAÇÕES</p> <p>Associar uma tabela de até duas entradas a informações apresentadas textualmente ou em um gráfico de barras ou de linhas.</p>
<p style="text-align: center;">Nível 2 Desempenho maior ou igual a 250 e menor que 275</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>ESPAÇO E FORMA</p> <p>Reconhecer as coordenadas de pontos representados em um plano cartesiano localizados no primeiro quadrante.</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Reconhecer os zeros de uma função dada graficamente.</p> <p>Determinar o valor de uma função afim, dada sua lei de formação.</p> <p>Determinar resultado utilizando o conceito de progressão aritmética.</p> <p>TRATAMENTO DE INFORMAÇÕES</p> <p>Associar um gráfico de setores a dados percentuais apresentados textualmente ou em uma tabela.</p>
<p style="text-align: center;">Nível 3 Desempenho maior ou igual a 275 e menor que 300</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Reconhecer o valor máximo de uma função quadrática representada graficamente.</p> <p>Reconhecer, em um gráfico, o intervalo no qual a função assume valor máximo.</p> <p>Determinar, por meio de proporcionalidade, o gráfico de setores que representa uma situação com dados fornecidos textualmente.</p> <p>Determinar o quarto valor em uma relação de proporcionalidade direta a partir de três valores fornecidos em uma situação do cotidiano.</p> <p>Determinar um valor reajustado de uma quantia a partir de seu valor inicial e do percentual de reajuste.</p> <p>Resolver problemas utilizando operações fundamentais com números naturais.</p>



Nível ¹	Descrição do Nível
<p>Nível 4 Desempenho maior ou igual a 300 e menor que 325</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>GRANDEZAS E MEDIDAS</p> <p>Resolver problemas envolvendo área de uma região composta por retângulos a partir de medidas fornecidas em texto e figura.</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Reconhecer o gráfico de função a partir de valores fornecidos em um texto.</p> <p>Determinar a lei de formação de uma função linear a partir de dados fornecidos em uma tabela.</p> <p>Determinar a solução de um sistema de duas equações lineares.</p> <p>Determinar um termo de progressão aritmética, dada sua forma geral.</p> <p>Determinar a probabilidade da ocorrência de um evento simples.</p> <p>Resolver problemas utilizando proporcionalidade direta ou inversa, cujos valores devem ser obtidos a partir de operações simples.</p> <p>Resolver problemas de contagem usando princípio multiplicativo.</p>
<p>Nível 5 Desempenho maior ou igual a 325 e menor que 350</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>GRANDEZAS E MEDIDAS</p> <p>Determinar medidas de segmentos por meio da semelhança entre dois polígonos.</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Determinar o valor de variável dependente ou independente de uma função exponencial dada.</p> <p>Determinar o percentual que representa um valor em relação a outro.</p> <p>Determinar o valor de uma expressão algébrica.</p> <p>Determinar a solução de um sistema de três equações sendo uma com uma incógnita, outra com duas e a terceira com três incógnitas.</p> <p>Resolver problema envolvendo divisão proporcional do lucro em relação a dois investimentos iniciais diferentes.</p> <p>Resolver problema envolvendo operações, além das fundamentais, com números naturais.</p> <p>Resolver problema envolvendo a relação linear entre duas variáveis para a determinação de uma delas.</p> <p>Resolver problema envolvendo probabilidade de união de eventos.</p> <p>Avaliar o comportamento de uma função representada graficamente, quanto ao seu crescimento.</p>



Nível ¹	Descrição do Nível
<p>Nível 6 Desempenho maior ou igual a 350 e menor que 375</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>ESPAÇO E FORMA</p> <p>Reconhecer as coordenadas de pontos representados em um plano cartesiano e localizados em quadrantes diferentes do primeiro.</p> <p>Associar um sólido geométrico simples a uma planificação usual dada.</p> <p>Resolver problemas envolvendo Teorema de Pitágoras, para calcular a medida da hipotenusa de um triângulo pitagórico, a partir de informações apresentadas textualmente e em uma figura.</p> <p>GRANDEZAS E MEDIDAS</p> <p>Determinar a razão de semelhança entre as imagens de um mesmo objeto em escalas diferentes.</p> <p>Determinar o volume de um paralelepípedo retângulo, dada sua representação espacial.</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Determinar os zeros de uma função quadrática, a partir de sua expressão algébrica.</p> <p>Resolver problemas de porcentagem envolvendo números racionais não inteiros.</p>
<p>Nível 7 Desempenho maior ou igual a 375 e menor que 400</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>ESPAÇO E FORMA</p> <p>Determinar a medida de um dos lados de um triângulo retângulo, por meio de razões trigonométricas, fornecendo ou não as fórmulas.</p> <p>Determinar, com o uso de do teorema de Pitágoras, a medida de um dos catetos de um triângulo retângulo não pitagórico.</p> <p>GRANDEZAS E MEDIDAS</p> <p>Determinar a área de um polígono não convexo composto por retângulos e triângulos, a partir de informações fornecidas na figura.</p> <p>Resolver problemas por meio de semelhança de triângulos sem apoio de figura.</p> <p>Resolver problemas envolvendo perímetros de triângulos equiláteros que compõem uma figura.</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Reconhecer gráfico de função a partir de informações sobre sua variação descritas em um texto.</p> <p>Reconhecer os zeros de uma função quadrática em sua forma fatorada.</p> <p>Reconhecer gráfico de função afim a partir de sua representação algébrica.</p> <p>Reconhecer a equação de uma reta a partir de dois de seus pontos.</p> <p>Reconhecer as raízes de um polinômio apresentado na sua forma fatorada.</p>



Nível ¹	Descrição do Nível
<p>Nível 7 Desempenho maior ou igual a 375 e menor que 400</p>	<p>Determinar os pontos de máximo ou de mínimo a partir do gráfico de uma função.</p> <p>Determinar o valor de uma expressão algébrica envolvendo módulo.</p> <p>Determinar o ponto de interseção de duas retas.</p> <p>Determinar a expressão algébrica que relaciona duas variáveis com valores dados em tabela ou gráfico.</p> <p>Determinar a maior raiz de um polinômio de 2º grau.</p> <p>Resolver problemas para obter valor de variável dependente ou independente de uma função exponencial dada.</p> <p>Resolver problemas que envolvam uma equação de 1º grau que requeira manipulação algébrica.</p> <p>Resolver problemas envolvendo um sistema linear, dadas duas equações a duas incógnitas.</p> <p>Resolver problemas usando permutação.</p> <p>Resolver problemas utilizando probabilidade, envolvendo eventos independentes.</p>
<p>Nível 8 Desempenho maior ou igual a 400 e menor que 425</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>ESPAÇO E FORMA</p> <p>Reconhecer a proporcionalidade dos elementos lineares de figuras semelhantes.</p> <p>Determinar uma das medidas de uma figura tridimensional, utilizando o Teorema de Pitágoras.</p> <p>Determinar a equação de uma circunferência, dados o centro e o raio.</p> <p>Determinar a quantidade de faces, vértices e arestas de um poliedro por meio da relação de Euler.</p> <p>Resolver problema envolvendo razões trigonométricas no triângulo retângulo, com apoio de figura.</p> <p>Associar um prisma a uma planificação usual dada.</p> <p>GRANDEZAS E MEDIDAS</p> <p>Determinar a área da superfície de uma pirâmide regular.</p> <p>Determinar o volume de um paralelepípedo, dadas suas dimensões em unidades diferentes.</p> <p>Determinar o volume de cilindros.</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Reconhecer o gráfico de uma função trigonométrica da forma $y = \text{sen}(x)$.</p> <p>Reconhecer um sistema de equações associado a uma matriz.</p> <p>Determinar a expressão algébrica associada a um dos trechos do gráfico de uma função definida por partes.</p>



Nível ¹	Descrição do Nível
<p>Nível 8 Desempenho maior ou igual a 400 e menor que 425</p>	<p>Determinar o valor máximo de uma função quadrática a partir de sua expressão algébrica e das expressões que determinam as coordenadas do vértice.</p> <p>Determinar a distância entre dois pontos no plano cartesiano.</p> <p>Resolver problema usando arranjo.</p> <p>Resolver problema envolvendo a resolução de uma equação do 2º grau sendo dados seus coeficientes.</p> <p>Interpretar o significado dos coeficientes da equação de uma reta, a partir de sua forma reduzida.</p>
<p>Nível 9 Desempenho maior ou igual a 425 e menor que 450</p>	<p>Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de:</p> <p>ESPAÇO E FORMA</p> <p>Reconhecer a equação que representa uma circunferência, dentre diversas equações dadas.</p> <p>Determinar o centro e o raio de uma circunferência a partir de sua equação geral.</p> <p>Resolver problemas envolvendo relações métricas em um triângulo retângulo que é parte de uma figura plana dada.</p> <p>GRANDEZAS E MEDIDAS</p> <p>Determinar o volume de pirâmides regulares.</p> <p>Resolver problema envolvendo áreas de círculos e polígonos.</p> <p>Resolver problema envolvendo semelhança de triângulos com apoio de figura na qual os dois triângulos apresentam ângulos opostos pelos vértices.</p> <p>Resolver problema envolvendo cálculo de volume de cilindro.</p> <p>NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES</p> <p>Reconhecer o gráfico de uma função exponencial do tipo $f(x)=10^{ax}$.</p> <p>Reconhecer o gráfico de uma função logarítmica dada a expressão algébrica da sua função inversa e seu gráfico.</p> <p>Determinar a expressão algébrica correspondente a uma função exponencial, a partir de dados fornecidos em texto ou gráfico.</p> <p>Determinar a inversa de uma função exponencial dada, representativa de uma situação do cotidiano.</p> <p>Determinar inclinação ou coeficiente angular de retas a partir de suas equações.</p> <p>Determinar um polinômio na forma fatorada, dadas as suas raízes.</p>

Nível ⁴	Descrição do Nível
Nível 10 Desempenho maior ou igual a 450	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: NÚMEROS E OPERAÇÕES; ÁLGEBRA E FUNÇÕES Determinar a solução de um sistema de três equações lineares, a três incógnitas, apresentado na forma matricial escalonada.

⁴ A Prova Brasil não utilizou itens da 3ª série que avaliam as habilidades do Nível 0. Os estudantes da 3ª série com desempenho menor que 225 requerem atenção especial, pois ainda não demonstram habilidades muito elementares que deveriam apresentar nessa etapa escolar.