

UNIVERSIDADE PAULISTA – UNIP
PROGRAMA DE MESTRADO EM ENGENHARIA DE PRODUÇÃO

**O USO DO *LEARNING ANALYTICS* PARA AUXILIAR A
GESTÃO DA EDUCAÇÃO A DISTÂNCIA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Paulista, para obtenção do título de Mestre em Engenharia de Produção.

ELISÂNGELA MÔNACO DE MORAES

SÃO PAULO
2016

UNIVERSIDADE PAULISTA – UNIP
PROGRAMA DE MESTRADO EM ENGENHARIA DE PRODUÇÃO

**O USO DO *LEARNING ANALYTICS* PARA AUXILIAR A
GESTÃO DA EDUCAÇÃO A DISTÂNCIA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Paulista, para obtenção do título de Mestre 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.

Orientadora: Prof.^a Dr.^a Marcia Terra da Silva

ELISÂNGELA MÔNACO DE MORAES

SÃO PAULO

2016

Moraes, Elisângela Mônaco de.

O uso de learning analytics para auxiliar a gestão da educação a distância. / Elisângela Mônaco de Moraes. - 2016.

80 f.: il. color.

Dissertação de Mestrado apresentado ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Paulista, São Paulo, 2016.

Área de concentração: Gestão de Sistemas de Operação.

Orientadora: Prof.^a Dra. Marcia Terra da Silva.

1. Learning analytics. 2. Distância. 3. Sistemas educacionais. 4. Integração de dados. I. Silva, Marcia Terra da (orientadora). II. Título.

ELISÂNGELA MÔNACO DE MORAES

**O USO DO *LEARNING ANALYTICS* PARA AUXILIAR A
GESTÃO DA EDUCAÇÃO A DISTÂNCIA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Paulista, para obtenção do título de Mestre em Engenharia de Produção.

Aprovado em:

BANCA EXAMINADORA

_____/_____/____

Prof.^a Dr.^a Marcia Terra da Silva (orientadora)

Universidade Paulista - UNIP

_____/_____/____

Prof. Dr. André Leme Fleury

Universidade de São Paulo - USP

_____/_____/____

Prof. Dr. Rodrigo Franco Gonçalves

Universidade Paulista – UNIP

DEDICATÓRIA

Dedico este trabalho às bases da minha vida:

Minhas filhas Luana e Larissa, minhas maiores motivações, que tiveram que suportar minha ausência por diversos momentos, para que este trabalho pudesse ser realizado;

Meu marido e companheiro de todas as horas, Messias Moraes, que sempre me apoiou;

Minha mãe Adélia agradeço a tudo o que fez e faz por mim, meu exemplo de vida;

Minhas irmãs Sandra e Adriana e meu cunhado Peter, pelo carinho, apoio e dedicação e ao meu pai Eluzai (*in memoriam*), de quem sempre me orgulhei e me espelhei.

AGRADECIMENTOS

Primeiramente, agradeço a Deus à oportunidade de crescimento.

Agradecimentos especiais à professora e orientadora Dr.^a Marcia Terra da Silva, aos ensinamentos, contribuições e paciência, sendo um exemplo de mestra e pesquisadora.

Ao Prof. Dr. Oduvaldo Vendrametto, Coordenador do Programa de Pós-Graduação em Engenharia de Produção (PPGEP), a todos os momentos de aprendizagem.

Ao Prof. Dr. Pedro Luiz de Oliveira Costa Neto, que acreditou em meu projeto.

Aos professores Dr. André Fleury Leme e Dr. Rodrigo Franco Gonçalves, às valiosas contribuições na banca.

Aos professores doutores do PPGEP: Irenilza de Alencar Nääs, José Benedito Sacomano e João Gilberto Mendes dos Reis, a todo o conhecimento compartilhado.

Aos colegas do programa e à secretária do PPGEP, Márcia Nunes, ao apoio nos momentos difíceis.

Ao professor Marcelo Souza, por acreditar e incentivar este trabalho.

À Universidade Paulista (Unip), ao suporte financeiro sob a forma de bolsa para o programa de mestrado.

A minha irmã Sandra, aos ensinamentos na língua inglesa.

Aos colegas da IES objeto deste estudo, ao apoio recebido.

RESUMO

A educação a distância é uma modalidade em expansão após a chegada da internet e do desenvolvimento das tecnologias de informação e comunicação. Na educação a distância a tecnologia é amplamente utilizada, os diferentes sistemas que compõem seu gerenciamento, como o sistema de gestão acadêmica e o ambiente virtual de aprendizagem, entre outros, geram grandes quantidades de dados. Surge então um desafio para as instituições de ensino superior: como consolidar os dados e deles extrair conhecimento. Esta pesquisa se propôs a implantar o *learning analytics* em uma instituição de ensino superior a fim de verificar se a utilização do mesmo auxilia na gestão da educação. A proposição é que a consolidação dos dados e a extração de conhecimento das diferentes bases de dados irá auxiliar os múltiplos usuários na gestão da educação. O modelo de Chatti et al. (2012) para implantação de *learning analytics* foi escolhido como referência para a aplicação em uma instituição de ensino superior de grande porte. Neste caso foram utilizadas as ferramentas de integração de dados, de banco de dados e analítica. Além das técnicas de visualização e de mineração de dados. Foi realizada uma aplicação analítica por meio da coleta de dados em três sistemas distintos e apresentados a perfis de usuários diferentes, cada qual acessava o conjunto de informações relativos ao seu perfil. Diante dos resultados, pôde-se perceber que, as informações obtidas são significantes indicadores que auxiliam os gestores na gestão da educação.

Palavras-chave: *Learning analytics*. Educação a distância. Sistemas educacionais. Integração de dados.

ABSTRACT

Distance education is a modality in expansion after the arrival of the internet and the development of information and communication technologies. In the distance education, technology is widely used; the different systems that comprise its management, as the academic management system and the virtual learning environment, among others, generate large amounts of data. Then emerge a challenge for institutions of higher education: how to consolidate the data and extract knowledge from them. This research aimed to deploy Learning Analytics in a higher education institution in order to verify that its use aids in the management of education. The proposition is that the consolidation of data and the extraction of knowledge from different databases will assist multiple users in the management of education. The model of Chatti et al. (2012) for the implementation of LA was chosen as a reference for application in a large higher education institution. In this case it was used data integration, database and analytics tools, besides the techniques of visualization and data mining. It was performed an analytical application through data collection in three different systems and presented to different user profiles, each accessed the set of information relating to your profile. Given the results, it was possible to realize that the information obtained is significant indicators that assist managers in the management of education.

Keywords: Learning analytics. Distance education. Educational systems. Data integration.

LISTA DE FIGURAS

Figura 1 – Cinco gerações de educação a distância.....	13
Figura 2 – Sequência de fases da pesquisa	20
Figura 3 – Fluxo dos dados	21
Figura 4 – A convergência do Analytics à diferentes escalas	25

LISTA DE TABELAS

Tabela 1 – Ranking das 10 maiores IES em número de aluno da EAD	15
Tabela 2 – Ranking dos 10 cursos que possuem maior número de alunos	15
Tabela 3 – Learning e academics analytics	24
Tabela 4 – Classificação dos casos de estudos de acordo com o método de análise.....	25

LISTA DE GRÁFICOS

Gráfico 1 – Número de matrículas em cursos de graduação, por modalidade de ensino – Brasil – 2003- 2014	13
Gráfico 2 – Evolução no número de alunos da EAD por categoria administrativa - Brasil.....	14
Gráfico 3 – Matriculados por grau acadêmico	14

LISTA DE ABREVIATURAS E SIGLAS

APMS - *Advances in Production Management Systems*

AVA – Ambiente virtual de aprendizagem

DBA – Administrador de banco de dados

DM - *Data mining*

EAD – Educação a distância

ETL - *Extract Transform Load*

HEI - *Higher education institutions*

HP - *Hewlett-Packard*

ICT - *Information and communication technologies*

IES – Instituição de ensino superior

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

IV - *Information visualization*

LA – *Learning analytics*

LMS - *Learning management system*

MEC – Ministério da Educação

POMS – *Production and Operations Management Society*

SNA - *Social network analysis*

TB – *Terabyte*

TI – Tecnologia da informação

TIC – Tecnologia de informação e comunicação

SUMÁRIO

1	CONSIDERAÇÕES INICIAIS	12
1.1	Introdução	12
1.2	Objetivos.....	16
1.2.1	Objetivo geral.....	16
1.2.2	Objetivos específicos.....	16
1.3	Justificativa	17
2	METODOLOGIA.....	18
2.1	Estrutura da dissertação	22
3	LEARNING ANALYTICS	24
3.1	O processo de implantação do LA (artigo 1).....	26
3.2	Modelos para a implantação do <i>learning analytics</i> : uma revisão da literatura (Artigo 1)	26
4	IMPLANTAÇÃO DO LEARNING ANALYTICS (ARTIGO 2).....	37
4.1	Implantação do <i>learning analytics</i> em uma Universidade.....	37
5	LEARNING ANALYTICS NA EAD: TÉCNICAS E FERRAMENTAS COMO BASE À GESTÃO DA EDUCAÇÃO (ARTIGO 3).....	46
5.1	Learning Analytics on Distant Education: Techniques and tools to base the Educational Management	46
6	CONSIDERAÇÕES FINAIS	64
6.1	Resultados.....	64
6.2	Contribuições.....	65
6.3	Limitações e recomendações para trabalhos futuros	65
	REFERÊNCIAS	67
	ANEXO A: Identificando o desempenho do aluno da EAD: relação entre as aprovações e as interações no AVA.....	69

1 CONSIDERAÇÕES INICIAIS

1.1 Introdução

A mudança de uma sociedade industrial para uma sociedade orientada às informações, conforme Castells (2000), criou novos estilos de vida, de consumo e também impôs mudanças na educação. Vivencia-se um mundo interligado pela rede e as tecnologias são as responsáveis pelas mudanças diretamente ligadas à educação. Estudos como os de Moore e Kearsley (2011) e Mendonça (2013) demonstram questões importantes sobre as mudanças no processo de ensino e aprendizagem na sociedade da informação diante das recentes tecnologias da informação e comunicação (TICs) principalmente na modalidade da educação a distância (EAD).

Para Yarmohammadian (2011), a educação a distância é um método em que o aluno não está presente fisicamente em ambientes educacionais e salas de aula. Esta definição está em consenso com as origens da EAD, quando se buscava ofertar conhecimento em larga escala independente da distância entre alunos e instituição de ensino.

Moore e Kearsley (2011) apontam que a EAD se fundamenta no aprendizado planejado que ocorre normalmente em um lugar diferente do local de ensino, o que para o autor exige técnicas especiais de criação, organização e administração, bem como de comunicação por meio de várias tecnologias.

Neste trabalho o conceito de EAD utilizado é: alunos e professores estão em ambientes diferentes em todo ou em parte do tempo enquanto aprendem e ensinam, e este processo de ensino e aprendizagem é mediado por recursos tecnológicos para transmitir informações e proporcionar meios de interação.

Para que a EAD chegasse a uma condição de popularizar a educação foi necessária a inserção de tecnologia no processo e isso foi possível com a evolução das tecnologias de informação e comunicação, responsáveis pelo amadurecimento dos modelos existentes. É possível divisar períodos e características de modelos utilizados, o que Moore e Kearsley (2011) definiram como gerações da educação a distância (Figura 1).

A última geração, a quinta, tem resultado em enorme interesse para a EAD, o que a faz ser uma modalidade em expansão com a chegada da internet e com o desenvolvimento das TICs. Ainda é vista como uma modalidade com potencial para viabilizar a demanda por ensino superior do Brasil, pois a EAD democratiza o acesso à educação superior para a

parcela da população que não tem oferta de cursos de graduação presencial em suas localidades.

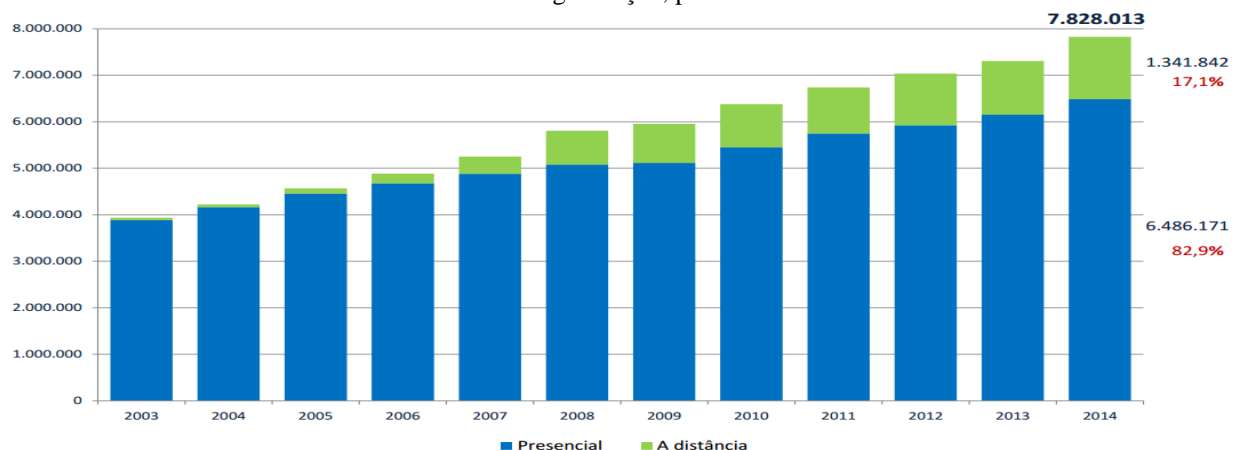
Figura 1 – Cinco gerações de educação a distância



Fonte: Moore e Kearsley (2011).

A expansão da modalidade fica clara nos censos de educação superior realizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP/MEC, 2014) (Gráfico 1).

Gráfico 1 – Número de matrículas em cursos de graduação, por modalidade de ensino – Brasil – 2003- 2014

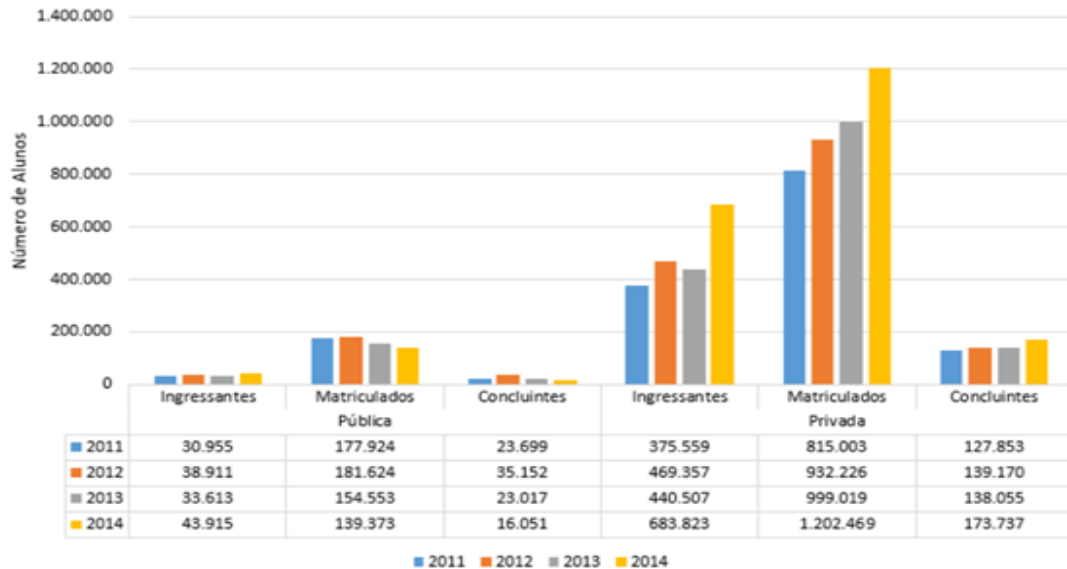


Fonte: Censo da Educação Superior 2014 - Notas Estatísticas (2014).

De acordo com o resumo técnico do censo da educação superior de 2014 publicado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP/MEC, 2014), com informações sobre a educação superior no Brasil, houve um crescimento no número de matrículas em cursos de EAD de 16,3% no período de 2013 a 2014, modalidade que conta com 17,1% das matrículas de graduação.

Em 2014, 1.341.842 alunos estavam matriculados em cursos de graduação a distância, com 1.202.469 matrículas na rede privada e 139.373 na rede pública, o que representa 89,6% na rede privada (Gráfico 2).

Gráfico 2 – Evolução no número de alunos da EAD por categoria administrativa - Brasil



Fonte: Autora - adaptado de Inep (2016).

Enquanto na rede pública a quantidade de alunos aumentou 1,5%, na rede particular o aumento foi de 5,6%. Chama a atenção a diferença entre ingressantes e concluintes o que sugere elevado número de evasão. Merece destaque as licenciaturas, das quais 540.693 matrículas são deste grau acadêmico, o que representa 36,9% do total de matriculados, número que representa um aumento de 6,7% entre 2013 e 2014 (Gráfico 3).

Gráfico 3 – Matriculados por grau acadêmico



Fonte: Autora - adaptado de Inep (2016).

As matrículas dos cursos a distância são predominantemente na rede privada e entre as 10 maiores IESs em EAD nenhuma é pública. A concentração também chama a atenção: o número de alunos das três maiores IES chegam a 47,47% do total de matriculados em 2014

(Tabela 1). Entre os 10 cursos com o maior número de alunos apenas dois não são das áreas de humanas/ciências sociais e só o curso de licenciatura atinge 27,5 % do total de matriculados (Tabela 2).

Tabela 1 – Ranking das 10 maiores IES em número de aluno da EAD

10 maiores IES em EAD

1º	UNIVERSIDADE NORTE DO PARANÁ - UNOPAR	310.855 alunos
2º	UNIVERSIDADE ANHANGUERA - UNIDERP	150.631 alunos
3º	CENTRO UNIVERSITÁRIO INTERNACIONAL - UNINTER	109.385 alunos
4º	UNIVERSIDADE PAULISTA - UNIP	107.990 alunos
5º	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI - UNIASSELVI	92.484 alunos
6º	UNIVERSIDADE ESTÁCIO DE SÁ - ESTÁCIO	68.766 alunos
7º	CENTRO UNIVERSITÁRIO - UNISEB	43.410 alunos
8º	CENTRO UNIVERSITÁRIO DE MARINGÁ - UNICESUMAR	39.038 alunos
9º	UNIVERSIDADE METROPOLITANA DE SANTOS - UNIMES	32.688 alunos
10º	UNIVERSIDADE DE UBERABA - UNIUBE	19.352 alunos

Fonte: Autora - adaptado de Inep (2016).

Tabela 2 – Ranking dos 10 cursos que possuem maior número de alunos

Posição dos Cursos na EAD

1º	PEDAGOGIA	331.191
2º	ADMINISTRAÇÃO	178.563
3º	SERVIÇO SOCIAL	101.272
4º	GESTÃO DE RECURSOS HUMANOS	88.292
5º	CIÊNCIAS CONTÁBEIS	85.620
6º	PROCESSOS GERENCIAIS	48.254
7º	LOGÍSTICA	36.358
8º	GESTÃO PÚBLICA	30.468
9º	HISTÓRIA	29.841
10º	GESTÃO AMBIENTAL	28.443

Fonte: Autora - adaptado de Inep (2016).

Este cenário da EAD no Brasil se torna mais complexo, por ser uma modalidade de ensino que utiliza os mais variados recursos das TICs para os processos de gestão acadêmica, administrativa e de desenvolvimento das atividades para o processo de ensino e

aprendizagem. Os processos da EAD precisam de organização e de novas estratégias de gestão para apresentar as informações de maneira adequada, e assim, que haja uma eficiente gestão desses processos.

Nesse panorama, Grajek (2014) refere que os utilizadores e as IESs têm agora novas necessidades e esperam que as TICs os ajudem cada qual no seu papel, professores, alunos e gestores a transformar o ensino e a aprendizagem, reter mais estudantes, melhorar o corpo docente e assim desenvolver vantagem competitiva.

De fato, muitas tecnologias têm surgido na educação provocando mudanças no modo de pensar e fazer educação. Na EAD a utilização de diferentes sistemas de gestão, como os de gestão acadêmica e de gestão de aprendizagem, entre outros, geram grande volume de dados. O aproveitamento destes dados permite um maior apoio no processo de tomada de decisão e esta é a ideia do *analytics* na educação que tem um grande potencial transformador pela sua potencialidade na utilização de forma organizada e agregada dos dados.

1.2 Objetivos

1.2.1 Objetivo geral

Esta dissertação de mestrado tem como objetivo geral realizar a implantação do *learning analytics* (LA), por meio de um modelo de referência encontrado na literatura para consolidar os dados, extrair conhecimento deles e verificar se a utilização do *learning analytics* na educação a distância em uma instituição de ensino superior auxilia na gestão da educação.

1.2.2 Objetivos específicos

Esta dissertação tem como objetivos específicos:

- Identificar modelos que viabilizem a implantação do LA e escolher o modelo mais adequado ao estudo;
- Identificar as ferramentas necessárias para a implantação do LA;
- Identificar as necessidades de recursos humanos e conhecimentos técnicos para a implantação do LA;
- Identificar as dificuldades encontradas na implantação do LA para a gestão da EAD.

1.3 Justificativa

Há um aumento da concorrência no setor do ensino superior que leva à adoção de práticas para garantir sucesso organizacional em todos os níveis, abordando questões como: retenção de alunos, admissão de novos alunos, captação de recursos e eficiência operacional. Barneveld, Arnold e Campbell (2012) fizeram com que as IESs entrassem na era do *big data*.

Na EAD a tecnologia é o meio de apoio no processo de ensino e aprendizagem. Os vários sistemas que envolvem estes cursos produzem uma enorme quantidade de dados. Este grande volume de dados, que tem um potencial informativo elevado para a gestão do ensino é perdido, seja porque não está devidamente preparado ou porque as fontes dos diferentes sistemas não se comunicam entre si. Surge então um desafio para as IESs que é como agregar e explorar os dados e deles gerar informações fundamentais.

Neste cenário surge o *Analytics* que se refere à capacidade de utilizar dados, análises e raciocínio sistemático para administrar um processo de tomada de decisão mais eficiente Barneveld, Arnold e Campbell (2012) e Campbell, Deblois e Oblinger (2007), e quando aplicado no setor da educação recebe o nome de *learning analytics* (LA).

Há desafios na implantação do LA nas IESs, como realizar a distribuição de aplicações analíticas que atendam, nas suas visualizações, os diferentes perfis de usuários com o intuito de auxiliar professores, alunos, coordenadores e gestores no suporte à tomada de decisão.

Os seguintes questionamentos são apontados na literatura:

- Como lidar com os dados de várias fontes: sistema de gestão acadêmica, sistema de gerenciamento de aprendizagem, sistema de inscrição Sharkey (2011); Ferreira e Andrade (2015) e Freitas et al. (2015).
- Como transformar as informações em benefícios para os vários usuários como os gestores, alunos e professores envolvidos no processo Shum (2012); Barneveld, Arnold e Campbell (2012) e Siemens e Long (2011).

Tendo em vista este contexto, uma grande IES que gera grande volume de dados de diversos sistemas utilizados na EAD, interessou-se em implantar e avaliar a eficácia da aplicação do LA e desta forma autorizou os acessos aos dados gerados. Portanto, esta pesquisa trata da implantação e verificação do benefício da utilização do LA em uma IES com foco na gestão do ensino.

2 METODOLOGIA

Justificada a importância da pesquisa e definidos os objetivos é necessário descrever a metodologia utilizada neste trabalho. De acordo com Martins, Mello e Turrioni (2013), os métodos qualitativos de pesquisas mais importantes são o estudo de caso e a pesquisa-ação. Para Thiollent (2005), uma pesquisa pode ser qualificada de pesquisa-ação quando houver realmente uma ação por parte das pessoas ou grupos implicados no problema sob observação, como é o caso deste trabalho.

O objetivo proposto pôde ser alcançado em uma IES que oferece cursos na modalidade a distância e possui diversos sistemas que não são integrados para agregar e explorar os dados e gerar informações fundamentais com eles.

Desta forma, como critério, foi definido analisar cursos já iniciados e em operação. A pesquisadora faz parte da equipe que realiza a implantação do LA na IES. Assim, foram identificados o problema de pesquisa e o objeto de estudo.

O presente trabalho teve início com uma pesquisa bibliográfica e descreveu, na primeira fase, a formulação do problema e a justificativa de um estudo que identificasse as causas e propusesse soluções. O resultado é apresentado no capítulo 1 – Considerações iniciais, seguindo as orientações para pesquisa-ação de Martins, Mello e Turrioni (2013).

A aplicação da pesquisa foi em uma IES de grande porte, com 27 cursos de graduação e grande número de alunos matriculados na modalidade a distância, que utiliza vários os sistemas no processo da EAD.

A equipe para esta pesquisa contou com os seguintes profissionais:

- Cientista de dados com formação em matemática e sólido conhecimento em linguagem R, banco de dados de alto desempenho, técnicas estatísticas e técnicas analíticas;
- DBA – *Database administrator* (Administrador de banco de dados), profissional responsável pelo gerenciamento, instalação, configuração, atualização e monitoramento dos bancos de dados da IES;
- Gerente de sistemas - responsável pelo principal sistema da IES, o de gestão acadêmica, com conhecimento de seu funcionamento e do banco de dados;
- Gerente de desenvolvimento - responsável pelo desenvolvimento dos sistemas próprios da IES, este profissional conhece o funcionamento do sistema e do banco de dados;

- Gerente de infraestrutura – responsável pelo gerenciamento, instalação, configuração, atualização e monitoramento dos *hardwares* e sistemas envolvidos nesta pesquisa;
- A pesquisadora conduziu o estudo e participou do processo de implantação do LA, além de observar os obstáculos e como foram contornados durante a implantação.

Para a utilização e verificação dos benefícios da utilização do *learning analytics*, os seguintes clientes / usuários da IES fizeram parte desta pesquisa:

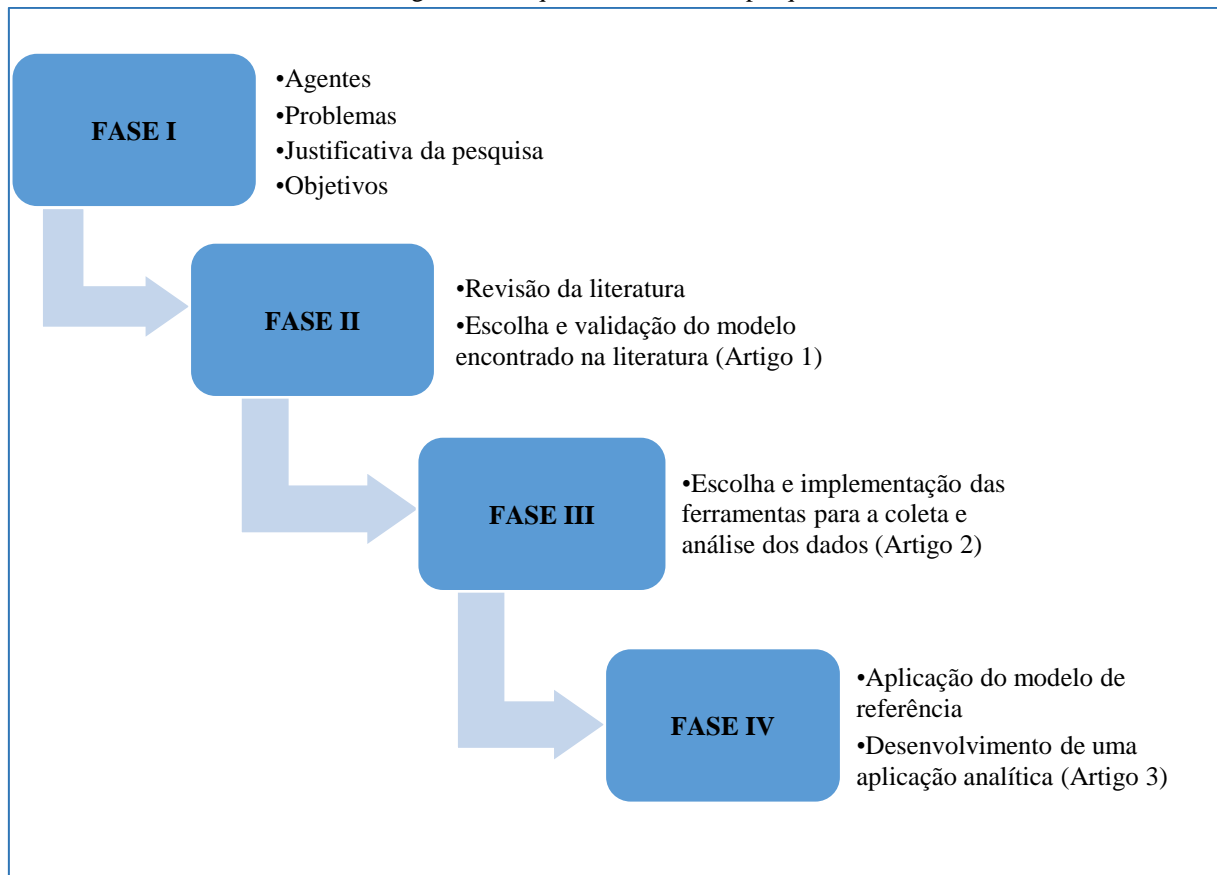
- Gerentes das áreas de tecnologia da informação (TI) e de marketing;
- Diretores administrativos e pedagógicos da IES;
- Coordenadores de curso da IES;
- Polos de apoio presencial;

Constantemente a equipe se reunia para analisar o processo e os gerentes, que também atuam como clientes, eram constantemente questionados para a validação dos dados.

Assim, o presente trabalho tem a sequência de fases da pesquisa conforme apresentado de forma simplificada na (Figura 2) e abaixo descrita.

- Fase I – Reconhecimento do campo de pesquisa - Nesta fase foram identificados os agentes e os problemas, a justificativa e a definição dos objetivos da pesquisa. Esta fase é apresentada no capítulo 1 – Considerações iniciais.
- Fase II – Escolha de um modelo para implantação do LA - Foi realizada uma revisão da literatura por meio de pesquisa bibliográfica sobre *learning analytics*. Com base na revisão bibliográfica escolheu-se o modelo de referência de *learning analytics* de Chatti et al. (2012), o que possibilitou classificar a literatura e facilitar o seu entendimento.
- Fase III – Escolha e implementação das ferramentas para a coleta e análise dos dados – Com a escolha do modelo de referência a pesquisadora fez a validação do modelo que foi implantado e testado na IES. A implantação foi acompanhada e analisada para esta pesquisa.

Figura 2 – Sequência de fases da pesquisa



Fonte: Autora (2016).

Nesta fase, utilizando o modelo de referência validado na fase II, foram escolhidas as ferramentas e quais sistemas da IES seriam utilizados para a coleta e análise dos dados. As ferramentas foram apresentadas pelo cientista de dados e a seleção foi feita pela pesquisadora por estarem ligadas ao propósito da pesquisa (Figura 3) e são apresentadas abaixo:

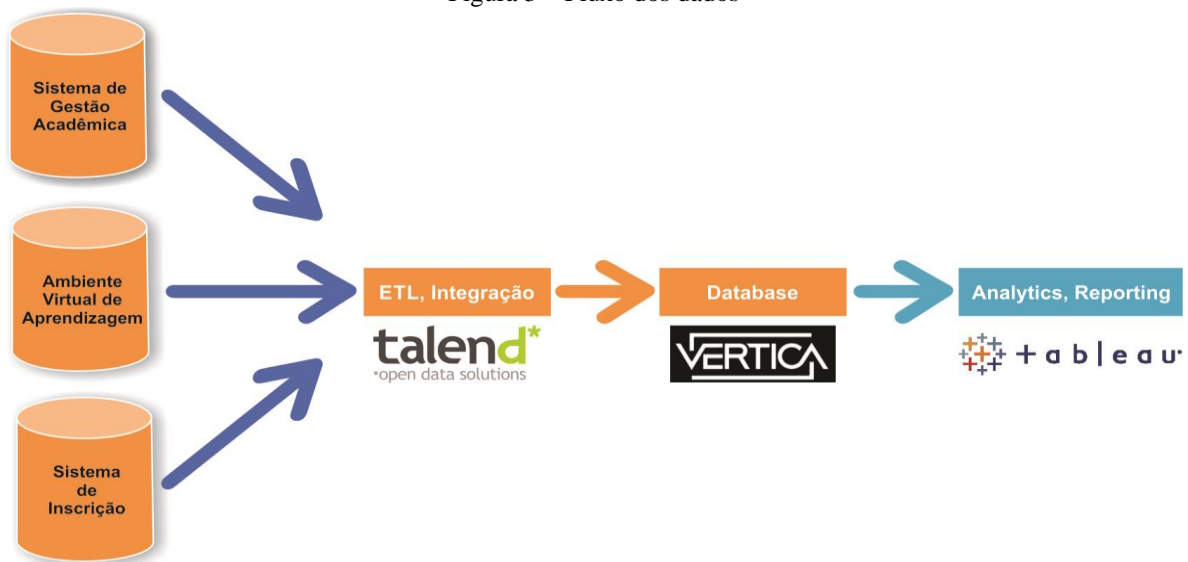
- *Talend Open Studio* – É uma solução em código aberto para a integração de dados e de extração, transformação e carga (ETL - *Extract Transform Load*), que é um processo para extrair dados de um banco de dados e inseri-los em outro banco, sendo necessário que os dados sejam processados e modificados;
- Banco de dados *HP Vertica* – Banco de dados para agilizar a manipulação de grandes volumes de dados, com utilização gratuita até 1TB de dados, o que, para esta pesquisa, está bem acima da necessidade da análise;
- *Tableau* – Ferramenta de análise de dados em versão pública e também gratuita para professores e de fácil utilização.

A IES em pauta possui vários sistemas que envolvem a EAD, cujos principais são: sistema de inscrição, sistema de gestão acadêmica, ambiente virtual de aprendizagem, sistema

de biblioteca, sistema de controle de fluxo de conteúdo, sistema de atendimento ao aluno e sistema de provas. A escolha dos sistemas para a coleta e análise de dados foi realizado pela pesquisadora, com base nos objetivos da pesquisa (Figura 3), e foram eles:

- Sistema de gestão acadêmica – Responsável pelo controle dos processos administrativos e da gestão acadêmica da IES;
- Ambiente virtual de aprendizagem – Seu principal objetivo é dar suporte ao processo de ensino *on-line*;
- Sistema de inscrição – Sistema responsável pela admissão dos alunos na IES.

Figura 3 – Fluxo dos dados



Fonte: Autora (2016).

- Fase V – Utilização do modelo de referência e desenvolvimento de uma aplicação analítica – nesta fase foi realizado o desenvolvimento de uma aplicação analítica para atender aos objetivos dos coordenadores e dos gestores, conforme abaixo.
- Coordenadores:
 - Distribuição dos alunos matriculados por polo, por região, faixa etária, por sexo, por turma, por disciplina;
 - Distribuição dos candidatos por quantidade de inscritos e quantidade de aprovados;
 - Árvore de decisão para previsão de evasão.
- Gestores de polos:
 - Quantidade de alunos matriculados por curso, por faixa etária, por situação acadêmica;

- Diretores da IES:
 - Quantidade de alunos matriculados por curso, por polo, por região, por faixa etária, por sexo;
 - Quantidade de polos de apoio distribuídos por região;
 - Quantidade de candidatos inscritos e aprovados no sistema de vestibular;
 - Árvore de decisão para previsão de evasão.

No desenvolvimento da aplicação com o modelo de referência foram utilizados os dados dos alunos dos sistemas de gestão acadêmica, do sistema de inscrição e do ambiente virtual de aprendizagem.

2.1 Estrutura da dissertação

Esta dissertação apresenta-se em seis capítulos. O primeiro capítulo mostra os problemas encontrados, a justificativa da pesquisa e os objetivos. Neste segundo capítulo é descrita a metodologia utilizada e a sequência de fases do estudo.

O capítulo 3 apresenta a utilização do LA por meio de uma revisão sistemática na literatura para encontrar modelos de referência para implantação. Apresenta-se o artigo 1 (Fase II), Modelos para implementar *learning analytics*: uma revisão da literatura, um estudo sobre o *learning analytics* e os modelos de referência encontrados na literatura para auxiliar a implantação do *learning analytics*.

O capítulo 4 apresenta o artigo 2 (Fase III) – Implantação do *learning analytics* em uma universidade, que contempla a aplicação e validação do modelo de referência encontrado na revisão da literatura na fase II, as escolhas das ferramentas utilizadas na implantação do *learning analytics* e a seleção dos sistemas utilizados.

O capítulo 5 apresenta o artigo 3 (Fase IV) – *Learning analytics* na EAD: técnicas e ferramentas como base à gestão da educação -, que contempla a aplicação do modelo de referência e o desenvolvimento com as ferramentas de integração, banco de dados e analíticas escolhidas na Fase III.

O capítulo 6 da dissertação - Considerações finais - apresenta as discussões e conclusões da pesquisa. O modelo de referência encontrado na literatura serviu de base para a implantação do *learning analytics* na IES. Neste capítulo também são apresentadas as limitações da pesquisa e recomendações para trabalhos futuros.

No final apresenta-se a lista de referências, que relaciona os autores citados nos capítulos 1, 2, 3 e 6 desta dissertação. Os autores citados e a bibliografia utilizada dos capítulos 3, 4 e 5 estão referenciados no final de cada artigo apresentado.

3 LEARNING ANALYTICS

Analytics refere-se à capacidade de utilizar dados, análises e raciocínio sistemático para administrar um processo de tomada de decisão mais eficiente Barneveld, Arnold e Campbell (2012) e Campbell, Deblois e Oblinger (2007). A definição de Bichsel (2012) é mais completa: *analytics* é a utilização de dados, análise estatística e modelos explicativos e preditivos para obter *insights* e agir sobre questões complexas.

O *analytics*, quando aplicado na educação, para alguns autores como Siemens e Long (2011). e Barneveld, Arnold e Campbell (2012) distinguem-se de duas maneiras, conforme pode ser observado na Tabela 3.

Academic analytics - reflete a análise de dados com o foco em nível institucional; é a aplicação de *business intelligence* (inteligência organizacional) na educação Siemens e Long (2011);

Learning analytics - uma das definições mais adotadas na literatura para *learning analytics* é "a medição, coleta, análise e comunicação de dados sobre alunos e seus contextos, para fins de compreensão e aprendizagem otimizando os ambientes em que ocorre" Siemens e Long (2011, p.32).

Tabela 3 – Learning e academics analytics

Tipo de <i>analytics</i>	Nível ou objeto de análise	Quem beneficia
<i>Learning analytics</i>	Curso: redes sociais, desenvolvimento conceitual, análise do discurso, currículo inteligente	Estudantes, faculdade
	Departamento: predição, padrões de sucesso; insucesso	Estudantes, faculdade
<i>Academic analytics</i>	Institucional: perfis dos estudantes, desempenho escolar, fluxo de conhecimento	Administradores, financiadores, <i>marketing</i>
	Regional (estado/região): comparação entre sistemas	Administradores, financiadores
	Nacional e internacional	Governos nacionais, autoridades em educação

Fonte: Siemens e Long (2011).

Por outro lado, para Shum (2012), não há esta divisão. Quando o *analytics* é aplicado na educação surge o termo *learning analytics*, com vasta atuação desde a atividade individual

do aluno até políticas nacionais/internacionais, conforme descrito pelo autor ao esquematizar o *learning analytics* em níveis macro, meso e micro (Figura 4) no qual cada nível dá acesso a um conjunto de diferentes dados e contextos.

Figura 4 – A convergência do Analytics à diferentes escalas



Fonte: Shum (2012).

O estudo de Papamitsiou e Economides (2014) mapeou na literatura as técnicas mais utilizadas no *learning analytics* (Tabela 4).

Tabela 4 – Classificação dos casos de estudos de acordo com o método de análise

Método de análise	Autores e Anos (trabalho ref.)
Classificação	Baker et al., 2008; Barla et al., 2010; Chen & Chen, 2009; Dejaeger et al., 2012; Dekker et al., 2009; Jeong & Biswas, 2008; Guruler et al. 2010; Guo, 2010; Huang & Fang, 2013; Khribi et al., 2009; Kizilcec et al., 2013; Klačnja-Milićević et al., 2011; Li et al., 2011; Lin et al., 2013; Lykourantzou et al., 2009a; Lykourantzou et al., 2009b; Moridis & Economides, 2009; Pardos et al., 2013; Romero et al., 2008; Thai-Nghe et al., 2011
Clusterização	Abdous, He & Yen, 2012; Chen & Chen, 2009; Khribi et al., 2009; Kizilcec et al., 2013; Klačnja-Milićević et al., 2011; Lykourantzou et al., 2009b; Romero et al., 2009
Regressão	Abdous, He & Yen, 2012; Macfadyen & Dawson, 2010; Romero-Zaldivar et al., 2012
Mineração de Texto	He, 2013; Leong et al., 2012; Lin, Hsieh & Chuang, 2009
Mineração de regra de associação	Merceron & Yacef, 2008; Romero et al., 2009
Análise de rede social	Fournier et al., 2011; Macfadyen & Dawson, 2010
Modelos de visualização	Ali et al., 2012; Pardos et al., 2013; Shih, Koedinger & Scheines, 2008
Estatísticas	Clow & Makriyannis, 2011; Fournier et al., 2011; Santos et al., 2012 Statistics Giesbers et al., 2013; Guo, 2010

FONTE: Papamitsiou E Economides (2014)

Neste estudo foi utilizado o conceito para o *learning analytics* de Shum (2012) e técnicas de classificação e de visualização. Em busca de modelos para auxiliar na implantação do *learning analytics* foi realizada uma revisão da literatura apresentada no artigo 1.

3.1 O processo de implantação do LA (artigo 1)

Como visto na metodologia, foi feita uma pesquisa bibliográfica para conhecer e escolher o modelo de LA mais adequado. Esta etapa foi útil como preparação para a implantação, para escolher a equipe e para definir os passos que deveriam ser dados na sequência.

Os resultados desta pesquisa geraram um artigo aprovado e apresentado no POMS 27th Annual Conference, 2016 e que está transcrito no subitem 3.2.

3.2 Modelos para a implantação do *learning analytics*: uma revisão da literatura (Artigo 1)

Models to implement Learning Analytics: A literature review

Elisângela Mônaco de Moraes

Paulista University-UNIP, Graduate Program in Production Engineering, Dr. Bacelar St. 1212,
São Paulo, Brazil

epmonaco@gmail.com

Marcia Terra da Silva

Paulista University-UNIP, Graduate Program in Production Engineering, Dr. Bacelar St. 1212,
São Paulo, Brazil

marcia.terra@uol.com.br

Marcelo Costa Souza

Paulista University-UNIP, Torres de Oliveira Ave. 330, São Paulo, Brazil

msouza@unip.br

Abstract

Managing Distance Education courses produces a great amount of data organized in different isolated systems. Large and preexisting databases can be an obstacle to gathering useful information for decision-making. Based on literature review, this paper presents models to implement Learning Analytics and the main challenges the managers must overcome.

Keywords: Learning analytics, Distance educations, Educational systems.

INTRODUCTION

The shift from an industrial society to a society oriented by information (Castells, 2000) formed new lifestyles, consumption and learning habits. Education on the information society looks for different ways the pursuit for knowledge, (Junior, 2004) demonstrated the search for knowledge by performing a comparison between traditional education and education in the information society as can be seen in Table 1

Table 1 - Comparison between traditional education and education in the information society (Junior, 2004)

Traditional education	Education in the Information Society
Focus on teaching	Focus on learning
Focused on the teacher	Focused on the learner
Emphasis on the transmission of knowledge	Emphasis on developing skills and attitudes (learning to learn)
It teaches how to do it	Student learn by doing
Works logical-mathematical and linguistic intelligence	Works multiple intelligences
Model production line	Personalized learning model
Awards submission	Awards participation
Little use of technology	Employs new media and new technologies
Little motivation	High motivation
Teacher's time nearly fully invested in the presentation of lessons	Considerable preparation of classes
Low utilization rate (classes and teacher)	High utilization rate (classes and teacher)

Thus, increasingly teaching and learning method has been supported by the Information and Communication Technologies (ICTs). Education has been using the data for the approach to teaching for years, the difference is that we now have the opportunity to efficiently capture large amounts of data; this is due to the evolution of the computer systems capability, networks and methodologies to analyze these data which provides previously unavailable insights.

In distance education, where students can build knowledge anywhere and anytime always mediated by ICT (Shin et al. 2011), the use of different systems such as academic management systems, learning management systems, among others, generate large amounts of data. The employment of these data collected from different sources, with different standards and users with different access levels allow greater support in the decision- making process that is the idea of Learning Analytics (LA).

Higher Education Institutions (HEI) are increasingly required to measure, demonstrate and improve student performance, but also have implemented predictive modeling for strategic decision-making at the institution level. (Mattingly et al. 2012)

In order to evaluate this new area, this work suggests a systemic literature review, covering the study of applications, the reference models found in the literature to implement the Learning Analytics and its forms. The work seeks to understand what types of data are analyzed, what techniques are applied, which are the actors involved and what are the goals for conducting the analysis.

LEARNING ANALYTICS

This area is founded and integrates the research and methodologies that are related to data mining, social network analysis, visualization data, machine learning, learning sciences, psychology, semantics, artificial intelligence, e-learning, and educational theory and practice

(Dawson and Siemens, 2014). As happens on any new area of research, many definitions, as you can see below, were introduced to Learning Analytics.

One of the most adopted definition in the literature for Learning Analytics is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long and Siemens 2011).

The Learning Analytics is used for teaching and learning processes and it refers to the interpretation of the data produced by the students in order to assess academic progress, to predict future performance and identify potential problems (Johnson et al., 2011). In educational institutions, the use of LA focuses on relevant data to students and teachers, using for this purpose, analytical techniques aiming to improve the learning outcomes of students through a better learning guidance, resources and curriculum interventions (Elias, 2011; Van Barneveld et al 2012).

Learning Analytics is the third wave of development of instructional technology, as study by (Fiadhi, 2014); the first wave began in 1991 with the appearing of Learning Management System (LMS) and the second wave integrating to the LMS a broader educational enterprise involving students in social networks, Web 2.0.

Using Learning Analytics, according to (Siemens et al 2011), improves the effectiveness of learning and its benefits include:

- Customization of the learning process and content;
- Provide students with information about their performance and of their colleagues and suggesting activities that address identified knowledge gaps;
- Provide the teacher with information of students whom need additional help, which teaching practices are having more effects that are positive.

Learning Analytics exists in various organizational levels with micro, meso and macro levels where each level gives access to a different set of data and contexts. For instance, the analysis of a classroom may include social network analysis in order to assess levels of individual engagements, whereas the institutional level analysis may be concerned with improving the operational efficiency of the university or comparing performance with other universities (Shum, 2012).

METHODOLOGY

The methodology adopted, on the research results for the Learning Analytics, qualify the article as a systematic literature review (Mancini and Sampaio, 2007). So that it is possible the completion of the literature review, the research was divided into distinct phases: 1) research questions; 2) searching the literature for data collection; 3) review and evaluation of the results of research; 4) synthesis of results.

Research Question

On the systematic mapping proposed in this paper the following research, question will be addressed:

RQ1 - Which are the models found in the literature to implement the Learning Analytics?

Literature searching

To collect established studies, the search was conducted in search engines of Science Direct, Web of Knowledge and Google Scholar. To perform the searching, it was used the Boolean expression: TS = ("learning analytics" AND "model"), therefore will be selected only the items where on its terms contains the word "learning analytics " and "model".

To facilitate the articles selection, were defined inclusion and exclusion principles (table 2).

Table 2 - Inclusion/ exclusion criteria

Include	Exclude
Articles available for download	Duplicate articles
Conference papers or periodicals	Small items
Articles with four or more pages	Articles not released for download
Data from 2011 to 2016	
Most cited articles	

The inquiry on Google scholar brought numerous results, so for this searching were included only the 80 most cited work. This searching procedure, after removing duplicate files, produced 152 results on the three bases that will be analyzed, with the majority of the articles published in 2013 (figure 1), 60 were found in Web of Knowledge , 23 in Science Direct and 69 on Google Scholar.

Then, it was evaluated the quality of work found according to the criteria of inclusion and exclusion defined, and based on the research questions, a new classification has been made.

Throughout the process of articles evaluation, it was identified that among the 152 articles, 04 of them were considered more central to our review, answering the RQ1.

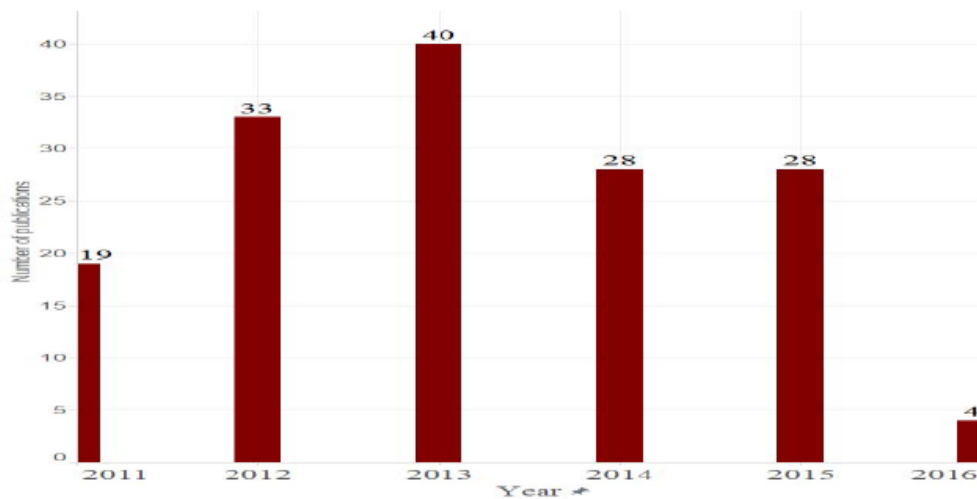


Figure 1 - Articles published per year.

RESULTS

In this section, it will be presented the findings based on analyzes carried out in published papers and the responses on the research questions.

RQ1. Which are the models found in the literature to implement the Learning Analytics?

To facilitate the understanding and thus classify the literature on Learning Analytics, table 3 divided the key studies with the models adopted for the implementation of Learning Analytics.

Table 3 - articles that describe models for the implementation of LA.

Learning Analytics - Model	Author and year (Paper ref.)
Model based on 4 dimensions: data and environment (What?), interested party (Who?), Goals (Why?) and methods (How?)	Chatti et al. 2012
This article presents a model for Learning Analytics focused on the student, the process aims with the creation of reports, services support to students and tutors, retention of students.	Freitas et al. 2014
Using of four types of technology resources : sharing , data gathering , information processing and knowledge application	Lias and Elias, 2011.
In this model of Learning Analytics includes seven components: collection, storage, data cleaning, integration, analysis, representation and visualization, and action.	Siemens, 2013

The reference model proposed by (Chatti et al 2012), divides the application of Learning Analytics in four dimensions as shown in (Figure 2), as follows:

- **What?** What kind of data is used in the analysis? - These data can be from the academic management systems, learning environments, library systems, etc.
- **Who?** Who is the target of analysis? - It refers to who will be presented the analysis; it can be to teachers, students, coordinators, managers and other actors who are part of the process.
- **Why?** What is the objective to be achieved by analyzing the data collected? - Some of the possible goals of LA include monitoring, analysis, forecast, intervention, mentoring / guidance, assessment, feedback, adaptation, personalization, recommendation and reflection.
- **How?** What techniques will be used to perform the analysis of the data collected? - Some of the most applied techniques for detecting important patterns are statistics, information visualization (IV), data mining (DM), and social network analysis (SNA).

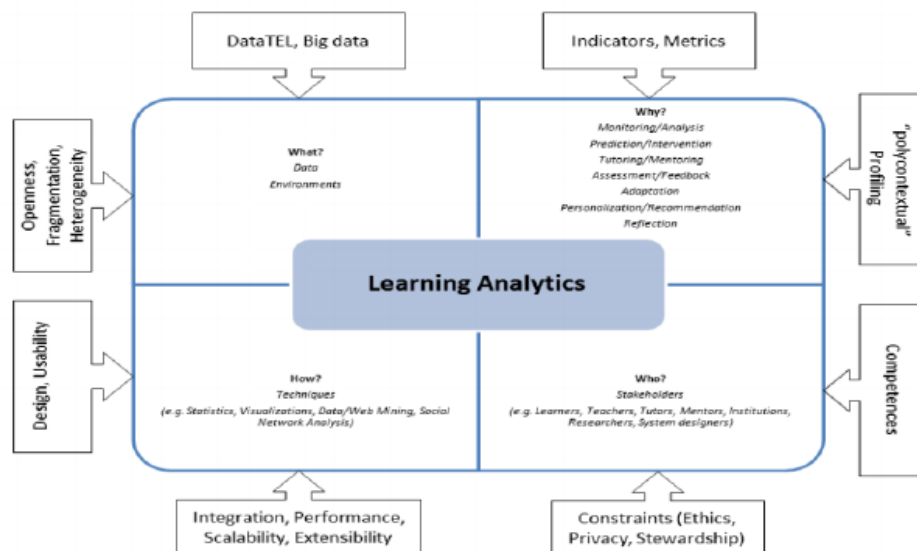


Figure 2 - Learning Analytics Reference Model (Chatti et al. 2012).

In the work of (Freitas et al. 2014), the proposed model (figure 3) is focused on the student, being an essential point that the student is seen as an element of analysis. For this purpose, a set of principles was developed shown below:

- **Develop a learning analytics strategy** - Have a single strategy analysis for the entire university.

- **Commit to create an infrastructure for big data integration** - Most Important Part of the Learning analytics is the access to information in real time. This information is derived from data integration of different systems and also from the processes of analysis.
- **Learner-centred service ethos—the unit of one** - Experience centered on the individual-student with emphasis on adaptive development and not only predictive.
- **Dynamic look at the students' learning journey** - Systems and processes must have an approach to be followed creating a student lifecycle.
- **Adaptively model user behavior** - The intention is to apply an adaptive model for the whole University, taking into account the mission of the University
- **Linking learning analytics within a wider dynamic context** - The link of all data sources to a student will create a network analysis. To treat such large volumes of data generated by various sources, it is necessary the use of different agile methodologies
- **Qualitatively driven crowd-sourced hypotheses formation** - Qualitative informations are used, boosting changing ideas, to generate multiple hypotheses and gain greater commitment and understanding of the interested parties
- **Rigorous view of ethics and adherence to highest standards of ethical procedures** – This principle refers to ethics in the protection of data such as: anonymity of data, identity protection and safe storage
- **External as well as internal review and cross-validation processes** - On this principle, different approaches of external evaluations should be used crossing internal data and be applied in a continuous manner, for the reason that, the contexts change as the conclusions.

Learning Analytics Model Principles

An effective analytics model ...

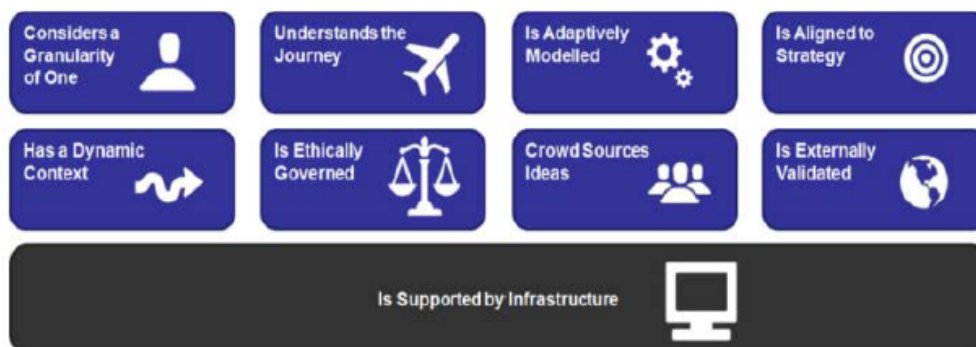


Figure 3 - The learning analytics model principles (Freitas et al. 2014)

In the model (Lias and Elias, 2011), four types of technology resources focus on the Learning Analytics: computers, people, theory and organizations. These features, along with the three phases of the cycle proposed by LA (Dron and Anderson, 2009) complete the model (figure 4).

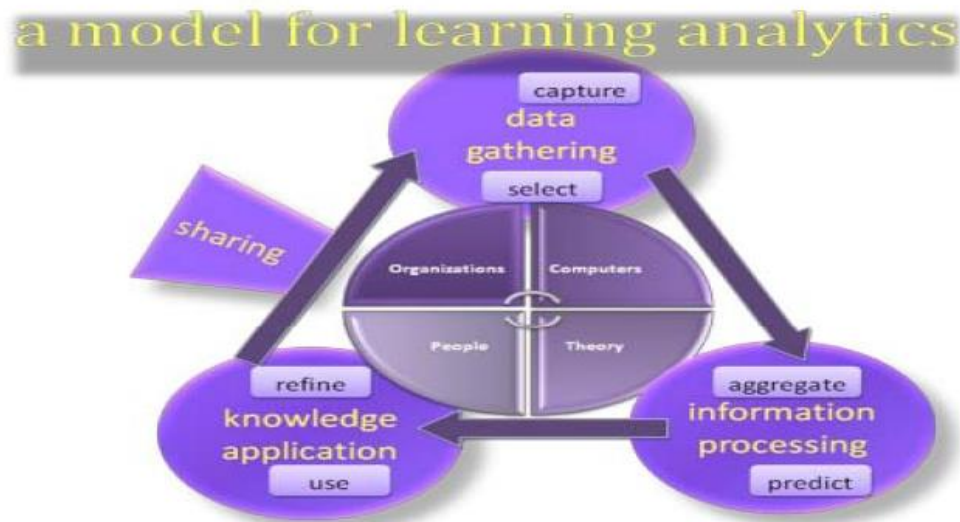


Figure 4 - Learning analytics continuous improvement cycle (Lias and Elias, 2011.)

- **Computers** - The challenge of this resource, with regard to hardware and software of data collection, is the integration of various data sources. Therefore, solutions able to gather and correlate are needed, for instance, Virtual Learning Environment data with the academic management system data.
- **Theory** - The theory for Learning Analytics includes analytical knowledge and good practices already used in other areas.
- **People** - Perhaps the most important resource, as, even in various aspects of the system, is necessary the knowledge and skills of human beings.
- **Organizations** - The social or relational resource, must be used, being essential for the analysis to know who will be involved in the project, how the decisions and actions will be supported and finally how communications and interactions will with others.)

For (Siemens, 2013), in the Learning analytics proposed model, a systematic approach ensures that all support resources are systematized so that, interventions or the creation of predictive models will only be possible with the full support of the whole education institution. This model has seven components: collection, storage, data cleaning, integration, analysis, representation and visualization, and action (figure 5).

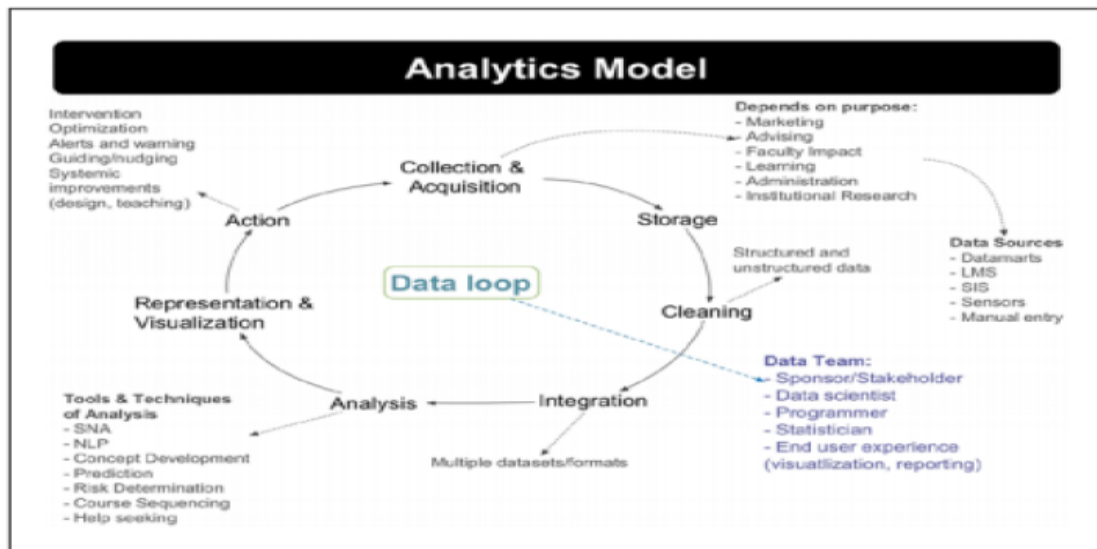


Figure 5 - Learning analytics model (Siemens, 2013)

CONCLUSION

The aim of this study, was, through the systematic mapping carried out, verify which are the models found in the literature to implement the Learning Analytics, in order to understand its concept and possible directions.

Given the results presented, it is possible to verify that are few models of Learning Analytics and no guidelines were found to indicate whether the use of the models in an implementation of Learning Analytics, brought indicative of success.

On the models found, it is possible to realize that the most significant challenges in the Learning analytics deployment are; the introduction of the tool throughout the Institution of Higher Education and for people to get involved in the project.

We believe that as the area evolves, other models will be proposed.

Bibliography

- Castells, M., & Gerhardt, K. B. (2000). *A sociedade em rede* (Vol. 3). São Paulo: Paz e terra.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318-331.
- Dawson, S., & Siemens, G. (2014). Analytics to literacies: The development of a learning analytics framework for multiliteracies assessment. *The International Review of Research in Open and Distributed Learning*, 15(4).
- Dron, J. and Anderson, T. (2009). On the design of collective applications, *Proceedings of the 2009 International Conference on Computational Science and Engineering*, Volume 04, pp. 368-374.
- Fiaidhi, J. (2014). The Next Step for Learning Analytics. *IT Professional*, (5), 4-8.

- Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., ... & Amab, S. (2014). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*.
- Johnson, L., R. Smith, H. Willis, A. Levine, Haywood, K. (2011). Learning Analytics. *The 2011 Horizon Report*. Austin, Texas: The New Media Consortium.
- Junior, V. G. de B. (2004) Modelo de inclusão digital para construção do conhecimento em qualidade de vida e atividade física, v. 121.
- Lias, T. E., & Elias, T. (2011). Learning Analytics: The Definitions, the Processes, and the Potential.
- Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society*, 15(3), 149-163.
- Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 4(3), 236-247.
- Sampaio, R. F., & Mancini, M. C. (2007). Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Braz. J. Phys. Ther.(Impr.)*, 11(1), 83-89.
- Siemens, G. (2013). Learning analytics the emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., ... & Baker, R. S. J. D. (2011). Open Learning Analytics: an integrated & modularized platform. *Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*.
- Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE review*, 46(5), 30.
- Shin, D. H., Shin, Y. J., Choo, H., & Beom, K. (2011). Smartphones as smart pedagogical tools: Implications for smartphones as u-learning devices. *Computers in Human Behavior*, 27(6), 2207-2214.
- Shum, S. B. (2012). Policy Brief - Learning Analytics. Moscow, Russian Federation: UNESCO.
- Van Bameveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE learning initiative*, 1, 1-11.

4 IMPLANTAÇÃO DO LEARNING ANALYTICS (ARTIGO 2)

Como visto na metodologia foi realizada a aplicação e validado o modelo de referência encontrado na revisão da literatura no capítulo 3. Neste capítulo apresentam-se as escolhas das ferramentas utilizadas na implantação do *learning analytics* e a seleção dos sistemas utilizados, que foram realizados na Fase II.

Os resultados desta fase geraram um artigo que foi submetido ao APMS 2016 *International Conference Advances in Production Management Systems* e está transcrito na íntegra no subitem 4.1.

4.1 Implantação do *learning analytics* em uma Universidade

Learning Analytics deployment at an University

Elisângela Mônaco de Moraes¹; Marcia Terra da Silva¹

¹Paulista University-UNIP, Graduate Program in Production Engineering,
Dr. Bacelar St. 1212, São Paulo, Brazil

Elisangela Mônaco de Moraes - emonaco@unip.br; Marcia Terra da
Silva - marcia.terra@uol.com.br

Abstract. This article presents the implementation of Learning Analytics in a University with the aim of testing the extraction of knowledge from databases of two systems used in distance education, aiming to provide support for the management of the distance education as a way to understand students' difficulties. Therefore, it was validated the applicability of literature reference model for the implementation the Learning Analytics. The data were extracted from the system of academic management and from the entrance exam to universities and were chosen the integration data, database and analytical tools. On the presented results, the model has been successfully applied, guiding the implementation of Learning Analytics at the University. The chosen tools facilitated the deployment and, brought benefits to university.

Keywords: Learning Analytics. Tools visualization. Distance Education

1 Introduction

The Distance Education is a type of education in expansion, which meets the new demands of a society oriented to information, as in new lifestyles and of consumption. We live in an interconnected world over the network, where more and more people study at home, as they can access from there, the information available online.

The distance education in Brazil is seen as a potential model to enable the demand for higher education in Brazil. EAD democratizes the entry to higher education, for the portion of the population that did not have undergraduate course offering in their localities. The expansion of this modality is clearly in higher education census conducted by INEP.

According to the technical summary of 2014 higher education census, published by the National Institute of Educational Studies Anísio Teixeira [1], which provides information on higher education in Brazil, the distance mode continues growing, with 1.34 million enrollment of higher education, representing 17.1% of total enrollment. Noteworthy the degree, where there were a total of 1,466,635 enrollments, with 540,693, or 36.9% of the distance mode. This figure shows an increase of 6.7% between 2013 and 2014.

This increase in the number of students in the distance education mode, and the characteristic of this type of education which is measured by Information and Communication Technology, results in an increasing volume of data produced, introducing the need to manipulation of various types of data in a fast way. Thus universities, seek ways to manage these data in order to transform them into information, and with that information, take faster decisions so they can be effective in their actions in the various organizational levels.

In this sense the objective of this study is to verify whether it is possible to extract knowledge from the two systems databases that are used in distance education, the main system that is the Academic Management and the Vestibular system.

Therefore it will be validated the applicability of the reference model of Chatti et al. [2] chosen in the literature review of Moraes et al [3] so that it is possible the extraction of knowledge from these two databases. Thus, provide support for the management of the Distance Education, in order to improve the targeting strategies for the achievement of goals, for example the evasion control.

2 Learning Analytics

Different definitions are assigned to the term learning analytics, being one of the most widely adopted in the literature: "the measurement, collection, analysis and reporting of data of students and its contexts, for the purpose of understanding and learning, optimizing the environment it occurs in"[4]

Another definition for Learning Analytics, is that if applied in several organizational levels, where each level gives access to a different set of data and contexts, this process provides valuable information [5].

3 Methodology

This study used data obtained from the Information Systems used in Distance Education, specifically the Academic Management System and Vestibular system. The courses that will be analyzed in this research are technological undergraduate courses, both information and communication area, as follows: Course of Technology Analysis and Systems Development, whose mnemonic is DS, this course is offered in the distance model at the University since 2014 and now has 1,889 students; and the Course of Technology in Information Technology Management, whose mnemonic is IT, this course is offered in the distance model at the University since 2009 and today has 2,122 students.

So that it is possible the realization of this work, the research was divided into distinct phases: research questions; collection and production of data; search results and conclusion.

3.1 Application process of Learning Analytics

The reference model, proposed by Chatti et al.[2], divides the application of learning analytics in 04 dimensions: what? Why? Who? How?, thus facilitating and classifying the literature.

Hence, using the model for the application of Learning Analytics, the following research questions will be addressed:

RQ1 - How and what are the data to be analyzed?

RQ2 - What are the aims of the analysis carried out and who will be presented?

3.2 Collection and production data

RQ1 - How and what are the data to be analyzed?

Using the reference model Chatti et al. [2] were chosen tools and which systems used for this analysis. Thus, to achieve the goal of dimension How?, proposed in the model, the tools will be used the tools:

- Talend Open Studio – An open source solution for data integration and extraction, transform and load (ETL)
- Database HP Vertica - Database to streamline the handling of high volumes of data, its use is free up to 1TB of data, which, for this research is well above the need of analysis.
- Tableau - Data analysis tool, leader in the market, has a public version and is also free for teachers, easy to use.

The choice of systems for the collection and analysis of data, thus meeting the dimension What? (Fig.1), was carried out by the researcher, based on the analysis aim, namely:

- Academic Management System –Responsible for controlling the administrative processes and the academic management of the university
- Vestibular System - system responsible for the student, admission to the University.

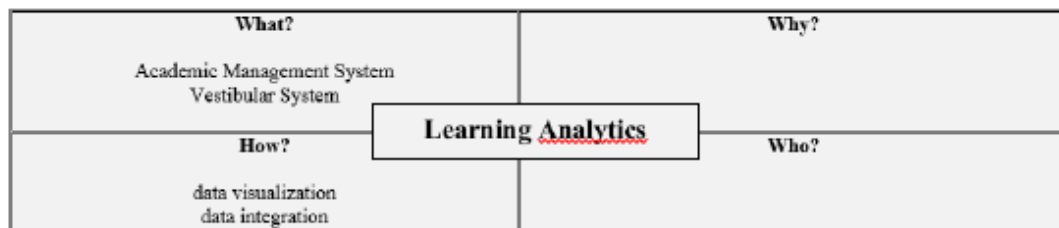


Fig. 1. Source: Adapted from Chatti, and Dyckoff Thus

The tools were chosen by the large volume of data, around 100 million records only in the Academic Management System, the variety of data and the speed required in the presentation of information. Each system uses a transactional database, but using a columnar database connected to a display tool, we obtain some benefits as: reduced

time for consultations; ability to analyze faster the University data and thus consolidating the information into a single tool for use in the various levels of the organization (Fig.2).



Fig. 2. Source: Author - data flow

RQ2 – What are the aims of the analysis carried out and who will be presented?

Using the reference model Chatti et al. [2], dimensions, Why? and Who? were filled with questions asked by the course coordinators, namely: How are the students distributed supported poles and what is the profile? What are the subjects that failed most? What is the number of students who pass the entrance exam? The managers of the University would like to know way online: How many people enrolled in the University entrance exam, by year, month, and day? What is the course that has the largest number of students enrolled? What is the profile of the student who seeks the IES, to take the entrance exam? This way, the model was completed in all dimensions, as figure 3, completing with the following information the dimensions, Why? and Who?

What? Academic Management System Vestibular System	Why? How <u>are the students</u> distributed supported poles and what is the profile? What are the subjects that failed most? What is the number of students who pass the entrance <u>exam</u> ? How many people enrolled in the University entrance exam, by year, month, and day? What is the course that has the largest number of students enrolled? What is the profile of the student who seeks the IES, to take the entrance exam?
How? data visualization data integration	<div data-bbox="504 1397 833 1451" data-label="Section-Header"> <p>Learning Analytics</p> </div> <div data-bbox="948 1417 1355 1512"> Who? Coordinators Course managers </div>

Fig. 3. Source: Adapted from Chatti, and Dyckoff Thus

Now with the model completed in its four dimensions, the results will be presented to the Coordinators Course and managers of IES.

4 Results

In this section, it will be presented the results of analyzes, based on the questions raised, first by the coordinators and then by managers.

For the data to be presented on online form to those involved, data integration tool was connected to transactional databases, data were extracted, transformed and loaded into columnar database and visualization tool connected on this database, so that it were developed the visualizations. Information is updated hourly and the involved people received access credentials.

For the coordinators the information's were presented, first answering the question: How are the students distributed at the support poles and what is the profile? (Fig.4)

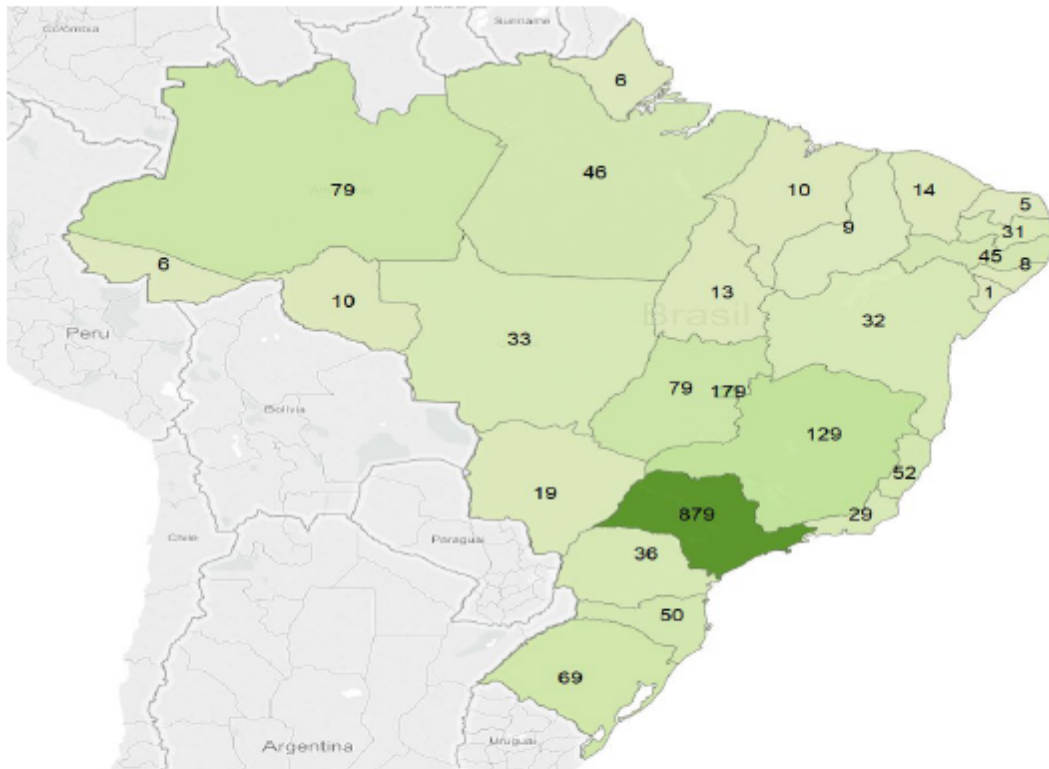


Fig. 4. Geographical distribution of students. (Source: Author)

The information extracted for the coordinators of the two courses, show the highest concentration of students in the Southeast, the number of students per pole, the percentage of students by sex, being the male far superior, to DS course 88,06% are male and 11,94 are female, and for IT course are 13.62% female and 86.38% male. The number of enrolled students, distributed by age group for both courses, has the highest concentration in the range of 30 to 34 years.

In the visualization, the coordinators go through the information, interacting in a drill-down fashion, thus, down in the details as follows: in the student enrollment numbers information, you can click to go in the geographical distribution by state, clicking again appear the information by city and in the city, the presential support poles, in the poles appear the students and clicking in the student is possible to view the failed, approved, registered, exempt and locked disciplines

For questions: What are the disciplines that most failed? What is the number of students who pass the entrance exam? The visualizations were presented as Figure 5 and 6.

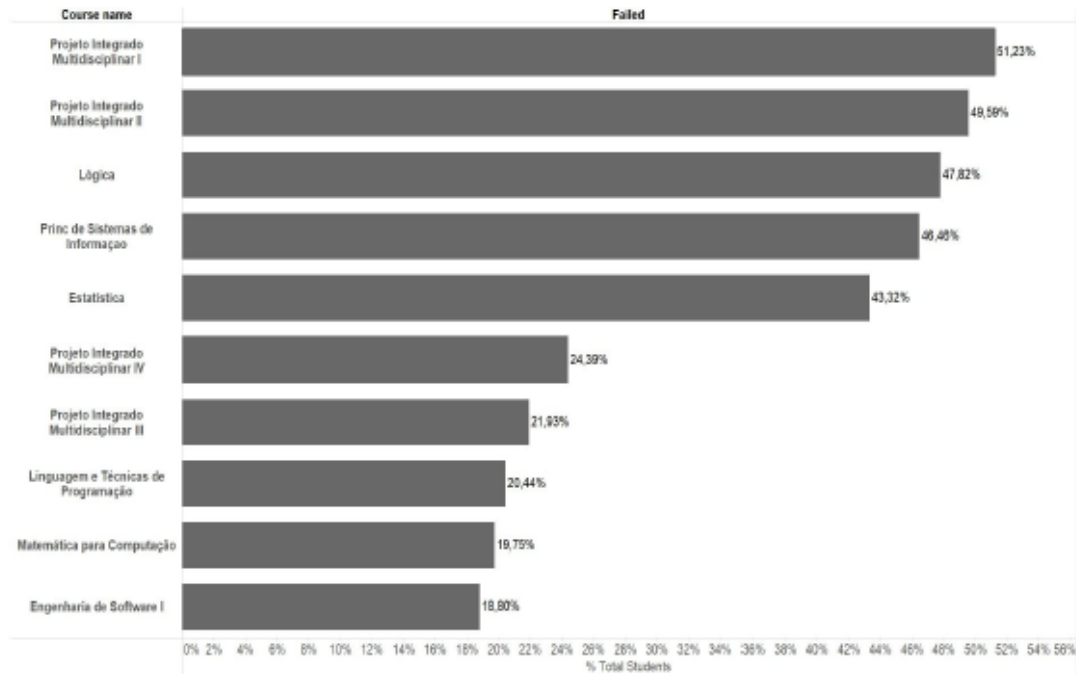


Fig. 5. Failed in the disciplines and exam (Source Author)

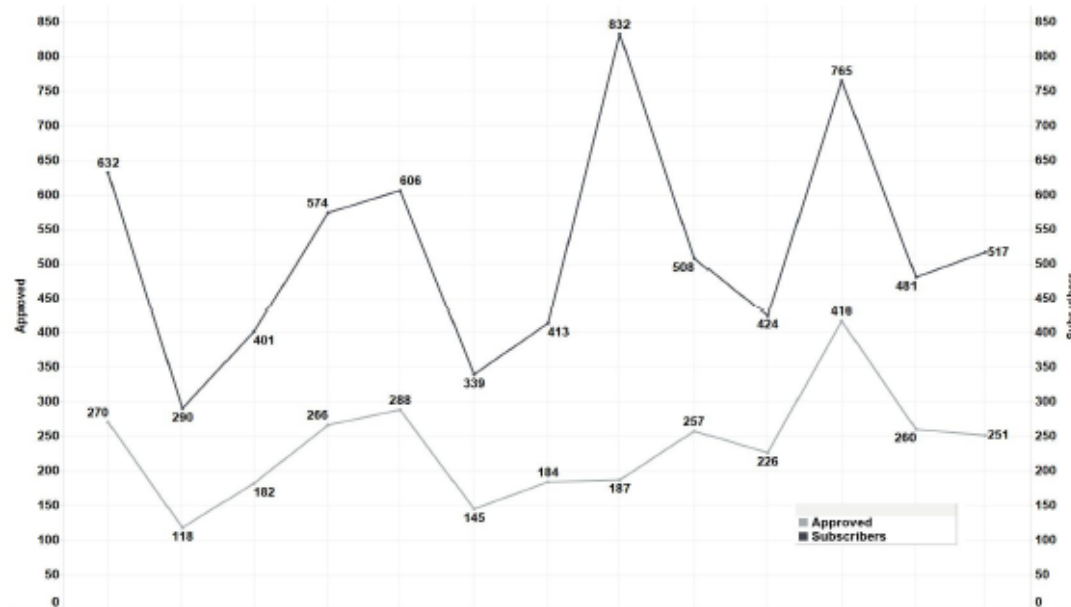


Fig. 6. Number of students who pass the entrance exam (Source Author)

The extracted data generated information to managers and the aim analysis were of objectives were presented.

The questions: How many enrolled in the University entrance exam by year, month, and day? What is the course that has the largest number of students enrolled? (figure 8) What is the profile of the student who seeks the IES, to take the entrance exam? (figure 9).

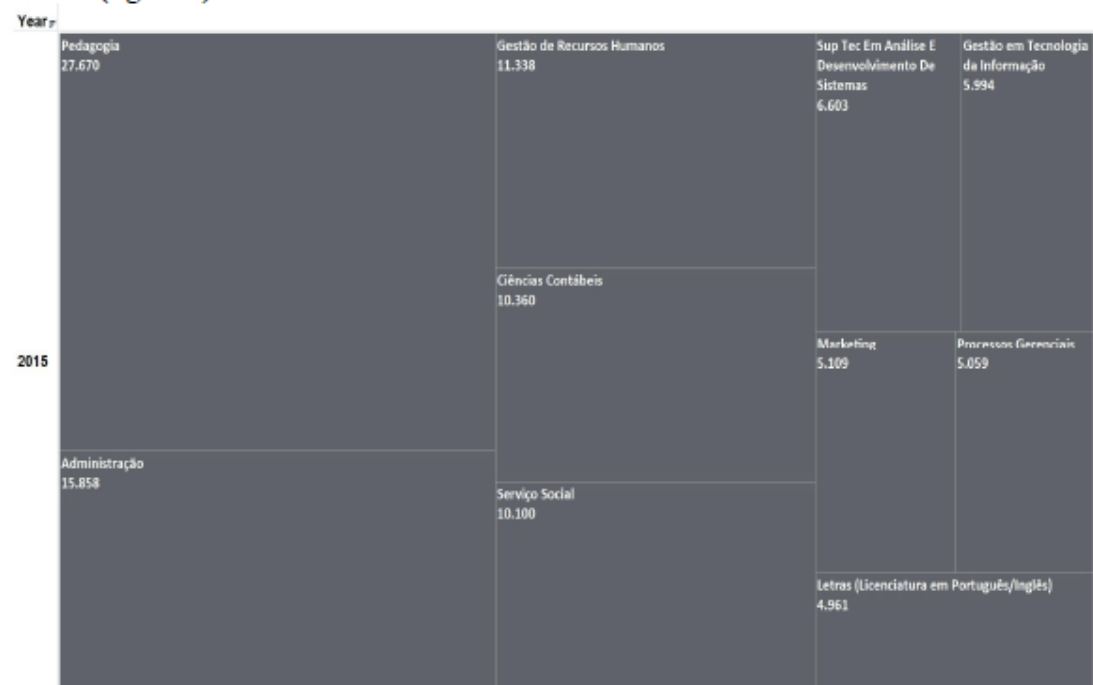


Fig. 7. Course that has the largest number of students enrolled

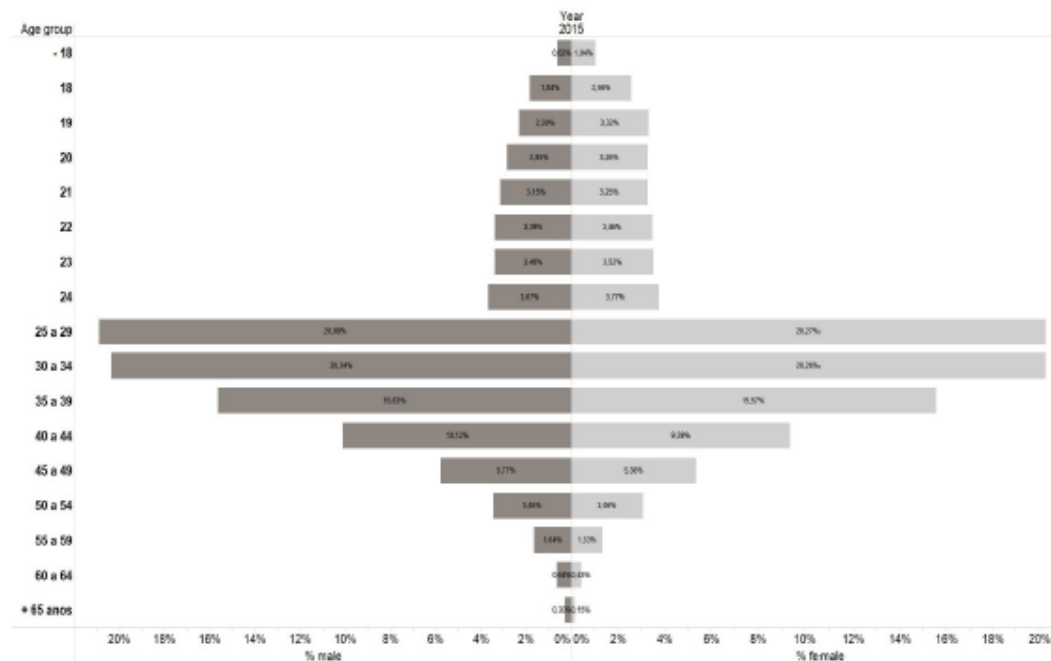


Fig. 8. Profile of the student

It were answered, so that you can check that the registered numbers in the college entrance is on the rise since 2011, the interaction with the tool enables the display of information per year, per month, per week, per day, per hour and even by minutes. The course with more enrolled people is Pedagogy, the age of the candidates are mostly between the groups: from 25 to 29 and 30 to 34.

5 Conclusion

The aim of this study was to verify the applicability of the reference model of Chatti et al [2], chosen in Moraes et al literature review, so that it is possible to extract knowledge from two databases.

On the presented results, the model has been successfully applied, guiding the implementation of Learning Analytics at the University. The chosen tools facilitated the deployment and brought benefits to the University such as:

- Reduced time to check the number of entries in the vestibular system, because the data were processed at dawn and presented in spreadsheets, which with the deployment the information's are accessed online.
- The access to information from different systems, concentrated in a single tool.

By gaining the ability to analyze students from the extraction of knowledge, using the integration, database and visualization tools, the university gained a clearer profile of their student population.

Managers and Coordinators of courses involved in this study evaluated that the information's presented are useful and relevant.

Since the focus of this study was the implementation, from a model of learning analytics, and thus extracting knowledge from data, in the next studies are suggested the correlation of data from other systems. Additionally it is necessary expand to more departments of the higher education institution.

References

1. INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Brasília: Inep, 2014. Disponível em: <http://portal.inep.gov.br/basica-levantamentos-acessar>. Acesso em: 5 fev. 2016.
2. Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318-331.
3. Moraes, E.M., Silva, M.T., Souza, M.C. (2016). Models to Implement Learning Analytics: A Literature Review. *Pomsmeeting*, 065-0931.
4. Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE review*, 46(5), 30.
5. Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *In Paper presented at the 2nd international conference on learning analytics & knowledge, Vancouver, British Columbia*

5 LEARNING ANALYTICS NA EAD: TÉCNICAS E FERRAMENTAS COMO BASE À GESTÃO DA EDUCAÇÃO (ARTIGO 3)

Uma versão preliminar deste artigo foi apresentada e publicada nos anais do XXXV Encontro Nacional de Engenharia de Produção: Perspectivas Globais para a Engenharia de Produção, realizado em Fortaleza – CE, em 14/ 10/ 2015 (Anexo A). No entanto, nesta pesquisa inicial foi identificado que somente o uso de estatística descritiva não apresentava o resultado esperado. Além disso, a análise dos resultados não é simples e identificou-se que os gestores precisavam de uma ferramenta mais automatizada para alcançarem a informação necessária para a tomada de decisão. Um refinamento da análise foi feito com a aplicação do modelo de referência e o desenvolvimento de uma aplicação utilizando as ferramentas de integração, de banco de dados e analíticas. Os resultados geraram o artigo 2 desta dissertação, que foi submetido à revista *Computers & Education* e está transcrito na íntegra no subitem 5.1

5.1 Learning Analytics on Distant Education: Techniques and tools to base the Educational Management

Abstract

In this work using the reference model for the implementation of the Learning Analytics, it was developed an analytical application using for this; integration, high-performance database and data analysis tools. Based on the model dimensions, the application was implemented and divided as follows: For the dimension “What?” it was used the data from the following systems: Academic Management, student admission and Learning Management. For the dimension "Why?" it was chosen as objective the following: (1) monitoring and analysis – it was performed the monitored of the activities of the students candidates, and the data visualization was made available (2) Forecast - a predictive model was built on withdrawal of the student and (3) Reflection - through data comparison among courses or in any Higher Education Institution. For the dimension "Who?" the application it was used by the following clients: (1) Course Coordinators; (2) Marketing Manager; (3) Administrative Officers; (4) Learning Center Managers and (5) Pedagogical Directors? For the dimension “How?”, the techniques used were: (1) statistical techniques - with the use of simple statistical operations such as media and trend line; (2) data mining techniques – it was used the classification process and the method was the decision tree and (3) data

visualization- panels were presented graphically showing different performance indicators. The implementation of an automated tool to achieve information for decision-making was successfully performed. The presented results show the effectiveness of the tool and brought benefits such as agility in decision-making, information accessible by online means and data consolidation in a single tool.

Keywords: architectures for educational technology system, Educational systems, learning analytics

1. Introduction

Information and communication technologies are increasingly present in all areas, today's society is driven by data, collecting and storing large volumes coming from different sources. With this increase in the number of data, the use of a resource, such as analytics, facilitates the process, thus, this technology has been adjusting itself to the various sectors of Industry and Services.

Due to increased competition in the higher education sector, Higher Education Institutions (HE) are seeking to adopt practices to ensure organizational success at all levels, addressing issues such as student retention, ingress of student, fundraising and operational efficiency (Van Barneveld et al., 2012). In addition, this competitive climate causes the HE to enter also on this era of "big data volumes."

The characteristic of the Distance Education is being mediated by Information and Communication Technology (ICT), and the rise on the number of students result in data volumes increase, in this context Universities seek ways to manage these data in order to transform them into information, and with this information, make faster decisions so they can be effective in their actions.

In this scenario emerge the Analytics, that refers to the ability to use data, analysis and systematic reasoning to administer a more efficient decision making process, (Van Barneveld et al., 2012), (Campbell et al., 2007), and when applied in the education sector receives the name of Learning Analytics (LA).

In this sense the objective of this study is, on using the reference model for the Learning Analytics, develop an analytical application using for this, integration tools, high-performance database and of data analysis.

2. Learning Analytics and models for implementation

Learning Analytics is defined as “the measurement, collection, analysis and communication of data on students and their contexts for understanding and learning purposes, optimizing the environment in which it occurs (Long and Siemens, 2011). The Learning Analytics (LA) is used for teaching and learning process and refers to the interpretation of the data produced by the students in order to, evaluate the academic progress, predict future performance and identify potential problems (Johnson et al., 2011). For (Shum, 2012), when the analytics is applied in education, comes the term Learning Analytics, with extensive interaction that can range from student's individual activity to national / international policies.

In order to make a clear understanding about the Learning Analytics, table 1 separated key studies with the models adopted for the implementation of Learning Analytics

Learning Analytics - Model	Author and year (Paper ref.)
Model based on four dimensions: data and environment (what?), interested part (who?), goals (why?) and methods (how?)	(Chatti et al., 2012)
This article presents a model for Learning Analytics focused on the student; the process aims, with the creation of reports, services to support students and tutors, student retention.	(Freitas et al., 2015)
Use of four types of technology resources: sharing, data gathering, information processing and knowledge application	(Lias and Elias, 2011)
In this model of Learning Analytics includes seven components: collection, storage, data cleaning, integration, analysis, representation and visualization, and action.	(Siemens, 2013)

Table 5-articles that describe models for the implementation of LA - Fonte Moraes et al., 2016

In this work, to facilitate the implementation of learning analytics, will be adopted the model of (Chatti et al., 2012), this model was chosen for ease of understanding and implementation. Table 2 shows how the dimensions guided this work and thus, how the dimensions will be used in this work.

WHAT? Academic Management System Vestibular System Learning Management system	WHY? Monitoring and Analysis Prediction Reflection
HOW? Statistical techniques Data mining techniques Information visualization	WHO? Course Coordinators Marketing Manager Managing Directors Polo Managers Pedagogical Directors

Table 6- Adapted from Learning Analytics Reference Model (Chatti et al., 2012)

3. Method

3.1. Project

The aim of this study is, using the reference model for the Learning Analytics, to develop an analytical application using, for this purpose, integration tools, high-performance database and of data analysis. In this study will be used the tools: Talend, for the integration of data sources with the Vertica database tool and the Tableau tool for data analysis (Figure 1). Based on the dimension “How?” of the reference model, which relates to the techniques that will be used in the implementation of Learning Analytics, it was decided to use statistical techniques, data mining and information visualization techniques.

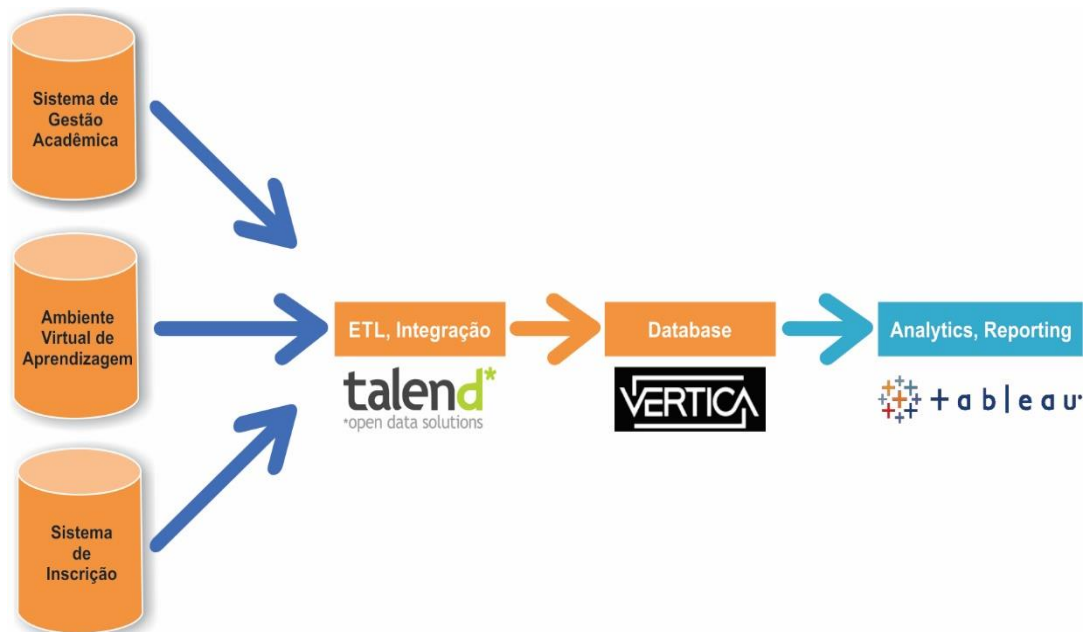


Figure 1- Flow of data - Source: prepared by the author

3.2. The participants

In this work, took part on the deployment team, 5 participants of the Higher Education Institution and the researcher, distinguishing the roles as: (1) Data Scientist; with degree in mathematics and strong knowledge in: R language, high-performance database, statistical techniques and analytical techniques; (2) DBA - Database Administrator; professional responsible for managing, installing, configuring, updating and monitoring the database of the HE; (3) Systems Manager - responsible for the Academic Management System of the HE, being this the main HEI system, consequently this professional has to have knowledge of the operating system as well as of the database; (4) Development Manager - responsible for the development of HEIs own systems, this professional has to have knowledge of the operating system and database; (5) Infrastructure Manager - responsible for the management, installation, configuration, updating and monitoring of the hardware and of the systems involved on this research and (6) the researcher; who conducted the study and participated in the process of LA implementation.

People whom took part on the validation and use of the application: (1) Managers of IT departments; (2) Marketing Managers; (3) Managing Directors of the Higher Education Institution; (4) Pedagogical Directors of the Higher Education Institution; (5) Course Coordinators of the Higher Education Institution; (6) Managers of the Learning Centers for classroom support.

3.3. Materials, procedures and analysis

The reference model divides the implementation of Learning Analytics in 4 dimensions. For the dimension "What?" that refers to which data will be analyzed, the data used were from the Management Academic system, student admission system and from the Learning Management system. For the dimension "Why?", which refers to the purpose of the analysis, were chosen as goal: (1) monitoring and analysis – having as objective the monitoring of the students' activities, the candidates and to generate reports; (2) Forecast - aiming to build a predictive model on the abandonment of the student, based on their current and past achievements; (3) Reflection - aiming to promote reflection through data comparison between courses or in the whole Higher Education Institution, and thus being able to reflect on the effectiveness of actions. For the dimension "Who?" that refers to whom will be presented the result of the analysis; the clients were: (1) Course Coordinators; (2) Marketing Manager; (3) Managing Directors; (4) Learning Center Managers and (5) Pedagogical

Directors. For the dimension "How?" that refers to which techniques will be used, were chosen: (1) statistical techniques - simple statistical operations were used as media and standard deviation, (2) data mining techniques - we used the process of classification, the method used was the decision tree; (3) information visualization - were presented graphically panels showing different performance indicators.

For the process of automating application, the following sequence was embraced:

Step 1: Prepare a consultation that brings the data related to the object of study - This step is done so that there is no need to transfer all the tables of one database to another;

Step 2: Once elaborated the consultation, a database extraction process is applied generating a CSV temporary file. Particular feature of SQL Server and Vertica, there is the need to convert the file exported by SQL Server since it exports in all file formats but in UTF -8, which is the standard Vertica format;

Step 3: Once converted to UTF8 format, all these steps are performed by Talend, the import is made to the columnar database;

Step 4: The tableau is connected to the Vertica database.

Subsequently, the task following the steps 1 to 3, was scheduled on Talend as figure 2

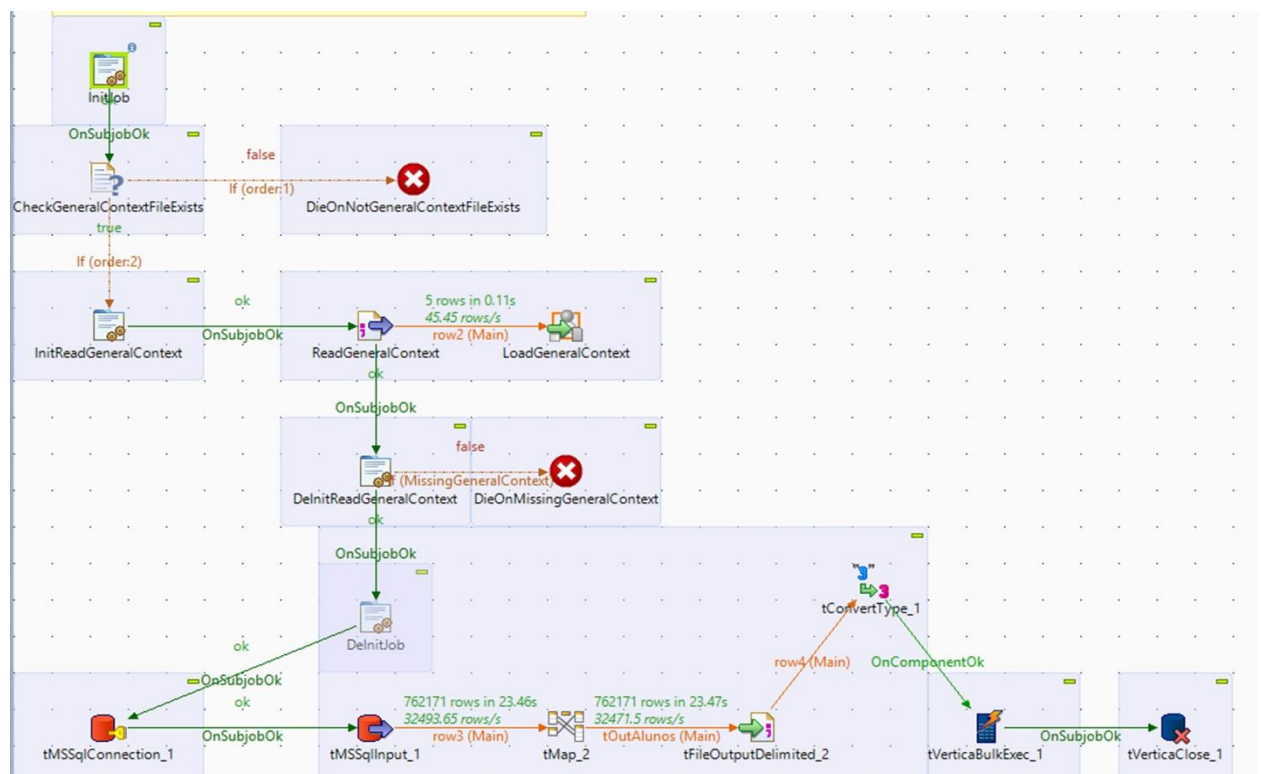


Figure 2- Drawing of the task to be executed - Source Author

The tasks were scheduled at different times; some are performed hourly and others every 12 hours.

To build the prediction model for the withdrawal of the students, a course on the HE was chosen where the population was 2321 students in 2015 semester 1, the following variables were used: the student's situation, series, year, semester, total of approvals, total of failures, total of outstanding charges and total of charges paid, it was used the tree of decision where the outcome is the situation of the student; for this technique was used the R language tool to develop data analysis and analysis statistics, language widely used by statistician and data scientists (Fox and Andersen, 2005).

4. Results and discussion

This session will be presented the results and discussions from the analytical application implantation in the HE. The HE did not have any tool to access the information on a consolidated basis, there were multiple worksheets generated and distributed by email; to create these spreadsheets hours were taken and It was not performed on a daily basis. With the implementation of analytical application some information is updated hourly and others every 12 hours, all automatically.

For that a successful deployment happened, team members made the validations on the data, it is crucial that people who knows the system processes are part of the team, in this way, before the application were actually used, the managers whom were part of the team validated the data.

Thereby, the application was presented to the various customers of the HE, where each one had access only to the information that was relevant to them; we will divide them into studies to make it clear your understanding.

For the study that was called coordinator, the result of the analysis was presented to the course Coordinators restrictively, where each coordinator visualizes his or her course information. For administrative and pedagogical Directors, the access is not restricted, in this way is possible to visualize the information of all courses. One of the objective on this analysis was to promote the reflection; it was presented data visualization about the course such as: the number of enrollments per year - so the coordinators were able to perceive how is the evolution of the course over the years, and was made a simple media calculation of the

students (figure 3); course student profile - such as age, gender and place of study, these information are useful for directing the actions that are carried out in the course and for marketing actions for prospection of new students (Figure 4) and (figure 5); numbers of registered and approved in the admission systems – it can be compared the quantity of registered people and how many of these pass the selection process, a simple media calculation was performed and also a trend line of registered and approved (Figure 6).

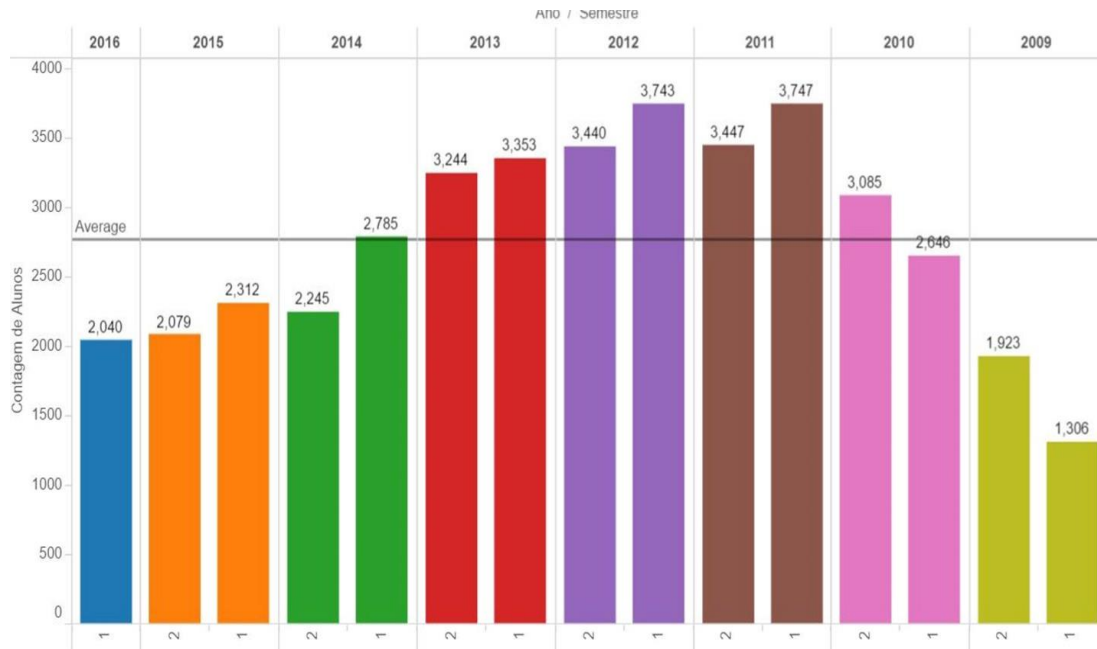


Figure 3- The evolution of the number of registered over the years

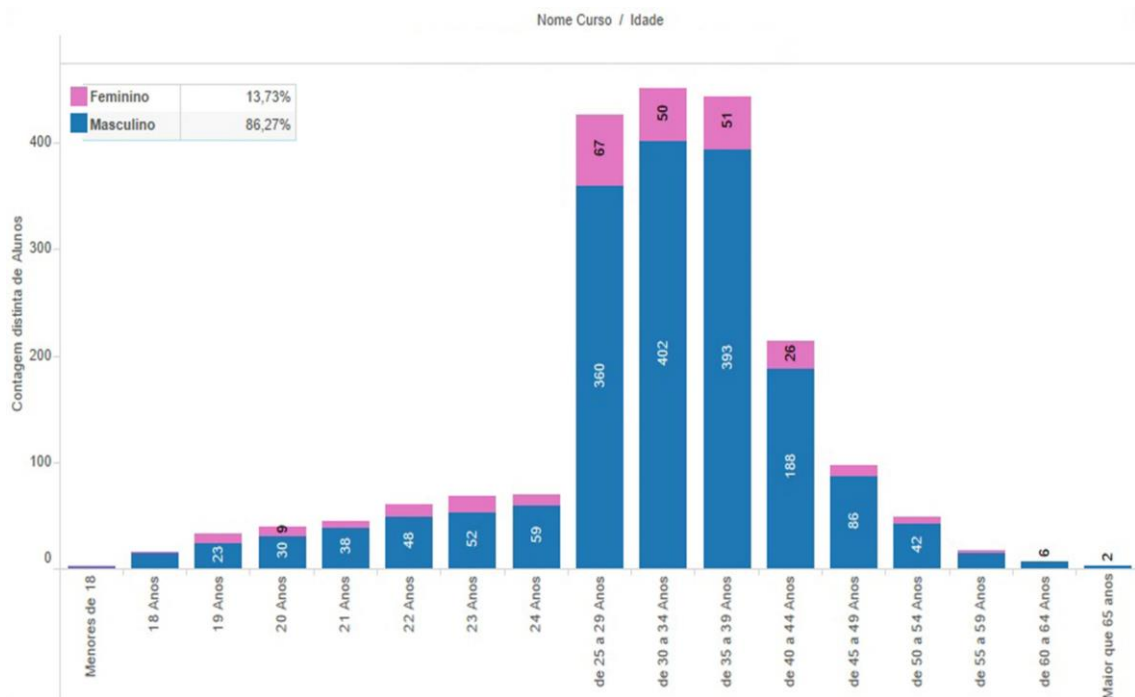


Figure 4 - Distribution of students by age and gender

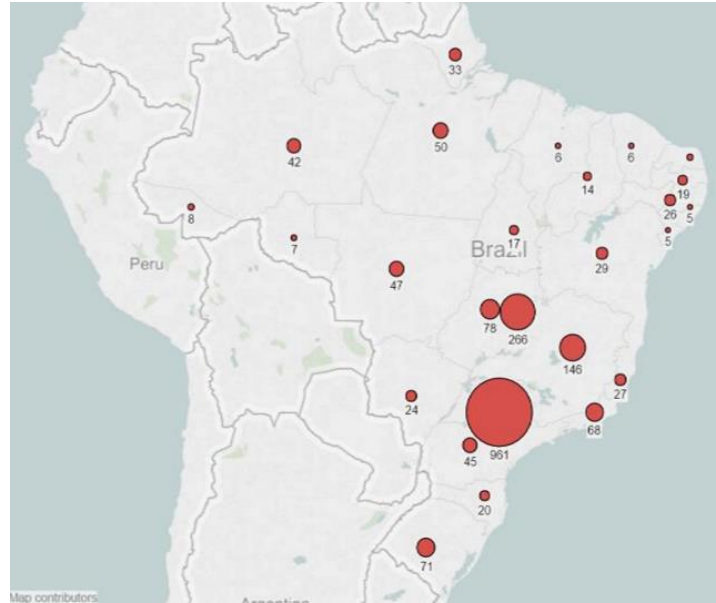


Figure 5 - Distribution of students by place of study

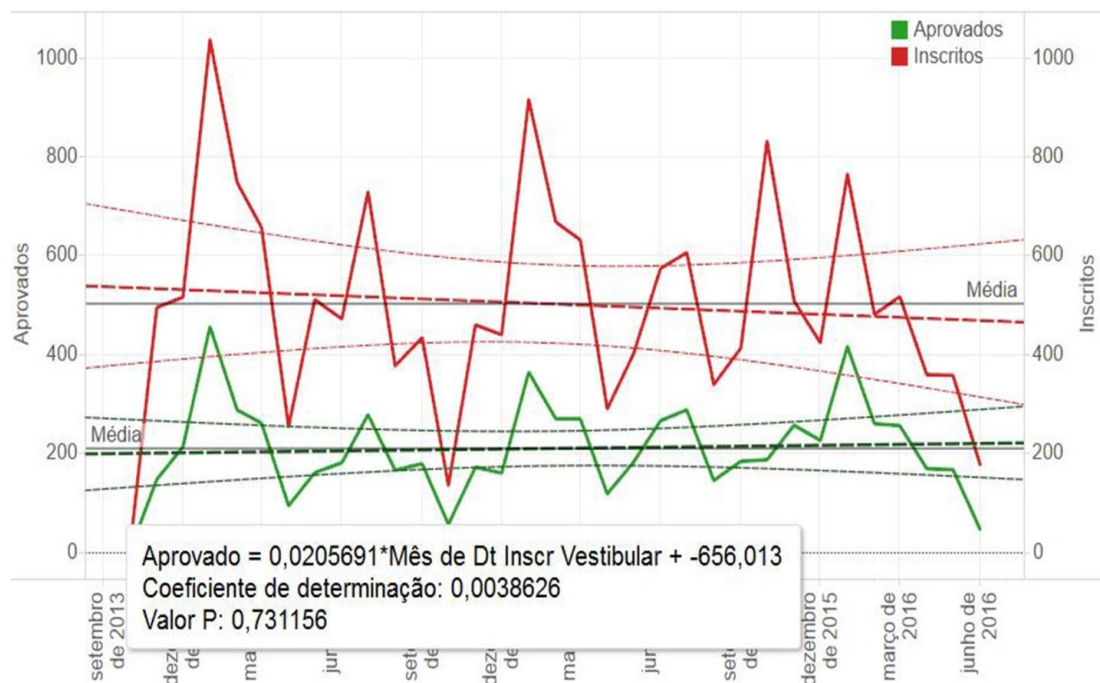


Figure 6- Monitoring registered and approved people in the selection process

In the study of the coordinator, the manager responsible for the Academic management system needed to inform to the data scientist and to the researcher which variables determined the status of enrollment of the student, the differences between year of admission and curriculum of admission of the student, on which tables contained the personal information of the students; these are just some examples of the importance of monitoring by people who knows the systems. Important to mention that the same procedure was necessary

with the manager responsible for the registration system and in this case, it was also necessary the qualification of the data, since some information were not in the correct standards.

Another objective on this study of the coordinator, was the monitoring of the questionnaires conducted in the LMS by the students, and so the reports were presented (Figure 7).

Still in the study of the coordinator, the prediction model was applied to the year 2015 on a course of the HE, firstly it was placed the entire population of the course independent of the semester, the result was not satisfactory, then it was separated the variables by semester, and then to the application of the decision tree, the population was divided into 70% for the tool training and 30% for the validation, are the results presented as figure 8.

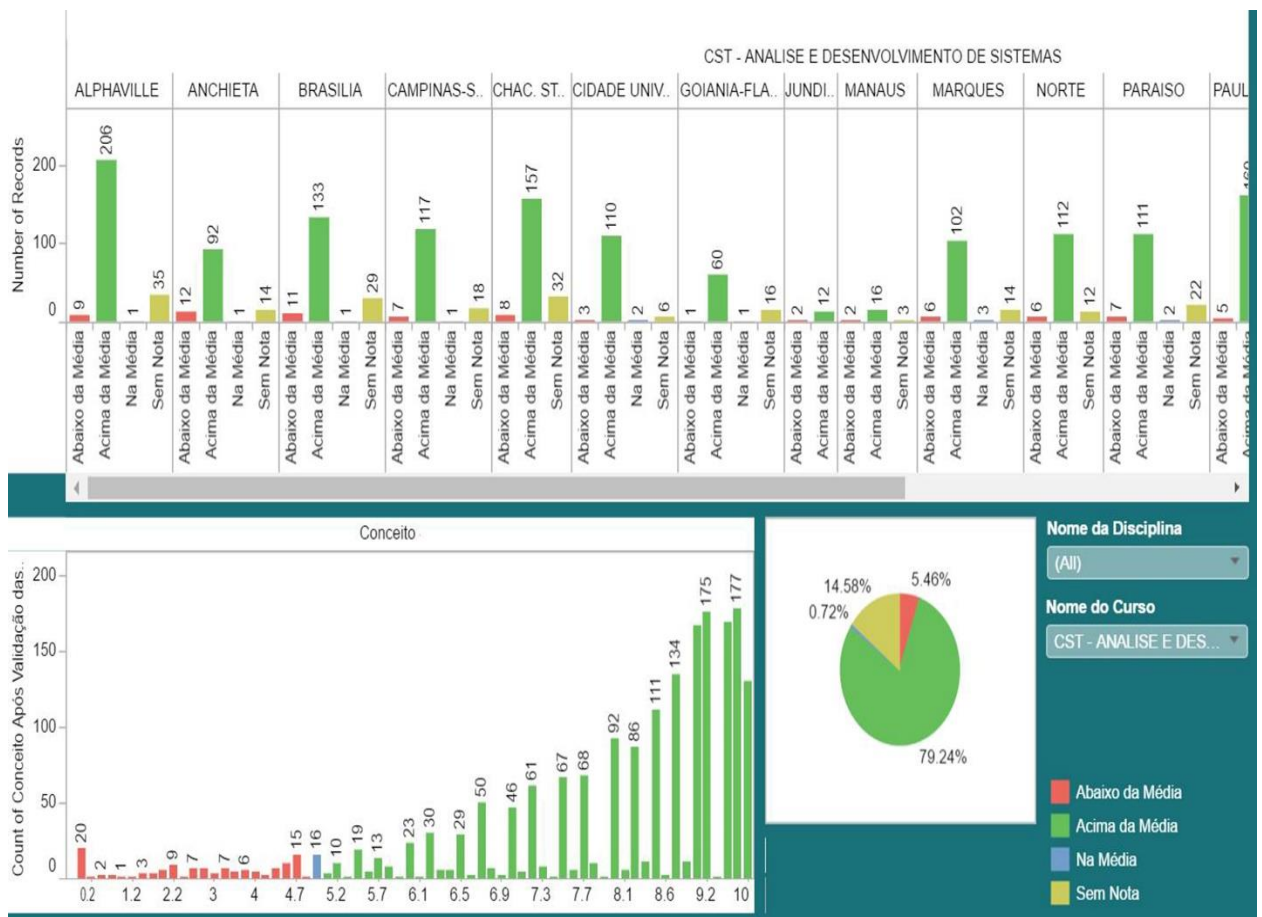


Figure 7- Realization of the exercises in the LMS

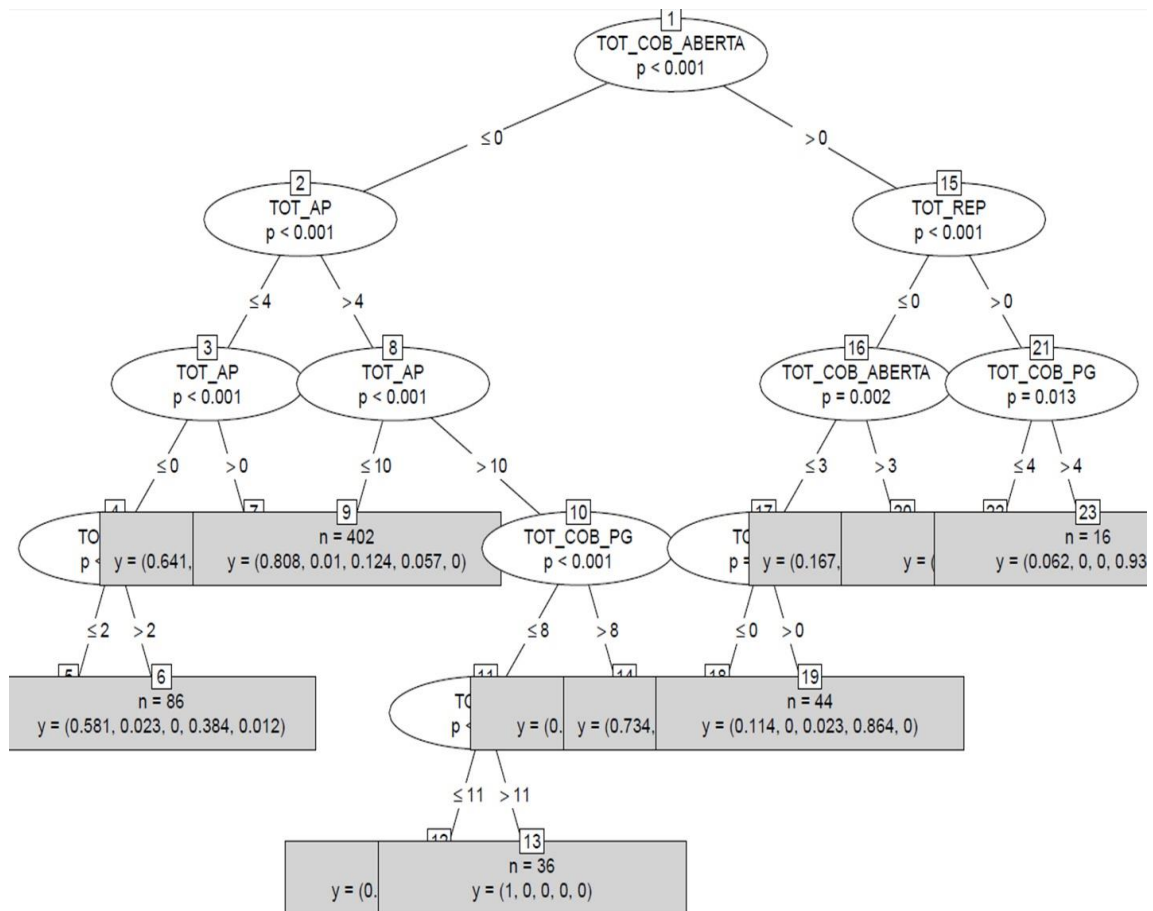


Figure 8 - Tree of decision for the prediction of the student permanence

The possible outcomes to this tree decision were the situations: active, canceled, completed, evaded and transferred; then following the tree decision, if the student is in total collection charge ≤ 0 and the total of approvals > 4 and ≤ 10 we have 80 % chance that the student has the outcome situation as active, if in another situation the student is in total collections charge > 0 and ≤ 3 , and total of approvals > 0 we have 86 % chance that the student has the outcome situation as evaded.

Accordingly, the tree is presented below:

Conditional inference tree with 12 terminal nodes

Response: SIT_ALUNO

Inputs: TOT_AP, TOT_REP, TOT_COB_ABERTA, TOT_COB_PG

Number of observations: 1584

- 1) TOT_COB_ABERTA ≤ 0 ; criterion = 1, statistic = 485.526
- 2) TOT_AP ≤ 4 ; criterion = 1, statistic = 82.455
- 3) TOT_AP ≤ 0 ; criterion = 1, statistic = 48.555
- 4) TOT_REP ≤ 2 ; criterion = 1, statistic = 73.172
- 5)* weights = 356
- 4) TOT_REP > 2
- 6)* weights = 86
- 3) TOT_AP > 0
- 7)* weights = 198
- 2) TOT_AP > 4
- 8) TOT_AP ≤ 10 ; criterion = 1, statistic = 46.386
- 9)* weights = 402
- 8) TOT_AP > 10
- 10) TOT_COB_PG ≤ 8 ; criterion = 1, statistic = 33.864
- 11) TOT_AP ≤ 11 ; criterion = 1, statistic = 27.563
- 12)* weights = 170
- 11) TOT_AP > 11
- 13)* weights = 36
- 10) TOT_COB_PG > 8
- 14)* weights = 9
- 1) TOT_COB_ABERTA > 0
- 15) TOT_REP ≤ 0 ; criterion = 1, statistic = 82.882
- 16) TOT_COB_ABERTA ≤ 3 ; criterion = 0.998, statistic = 20.542
- 17) TOT_AP ≤ 0 ; criterion = 0.994, statistic = 17.559
- 18)* weights = 64
- 17) TOT_AP > 0
- 19)* weights = 44
- 16) TOT_COB_ABERTA > 3
- 20)* weights = 66
- 15) TOT_REP > 0
- 21) TOT_COB_PG ≤ 4 ; criterion = 0.987, statistic = 11.525
- 22)* weights = 137
- 21) TOT_COB_PG > 4
- 23)* weights = 16

For the study called Vestibular, the result of the analysis was presented to the Managing Directors and Marketing Managers, the HE used to control the number of registered and approved on the examinations, through spreadsheets, where the execution were taken hours. In this analysis the aim was reflection, monitoring and analysis and were used the statistical techniques and information visualization.

Therefore, several comparison studies were conducted, such as: yearly comparison of the total enrolled in the college entrance examination (vestibular), comparisons of the registrations by geographic distribution and monthly comparison of registrations. These previews showed to the Directors of the HEI how is the progress and concentration of registrations in the vestibular (Figure 9).

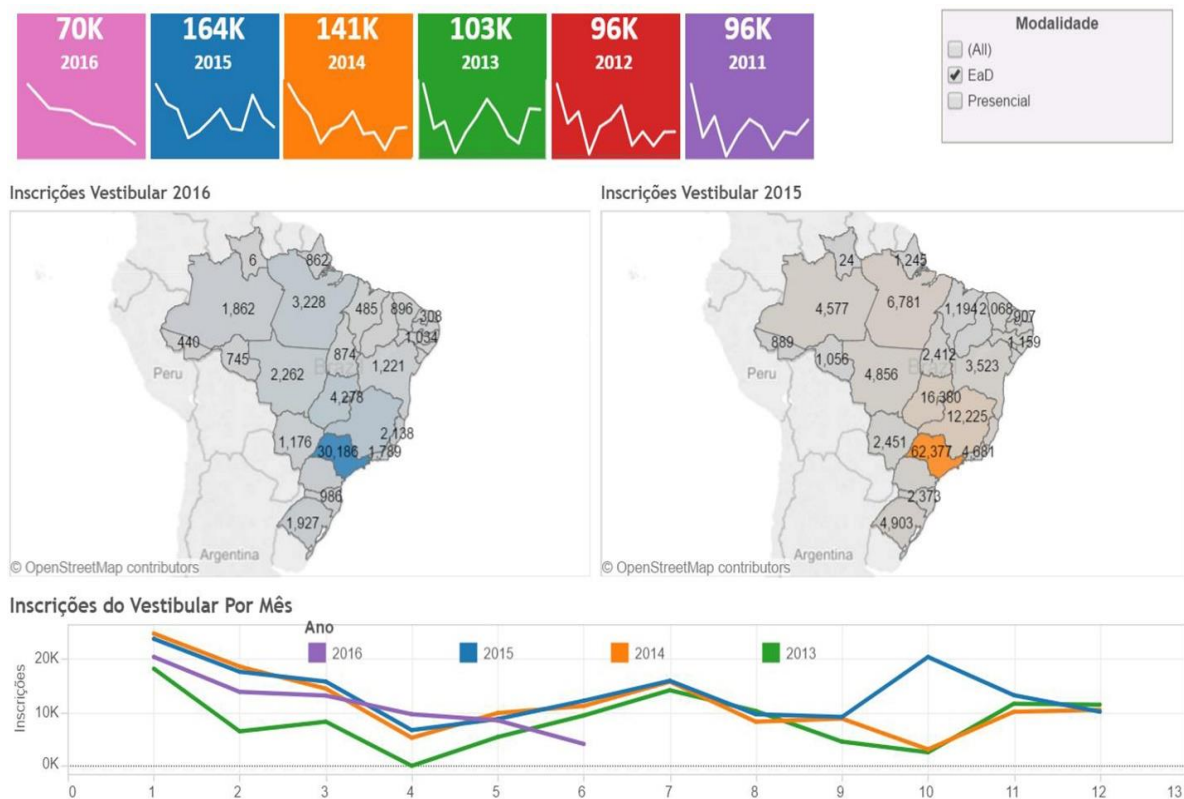


Figure 9 - Comparison of the number of registrations by year, month and location

The tool also allows the client to interact with the data, so that the user himself can select data for comparisons; for figure 10 a simple media calculation was performed and it is also possible to visualize the types of filters released.

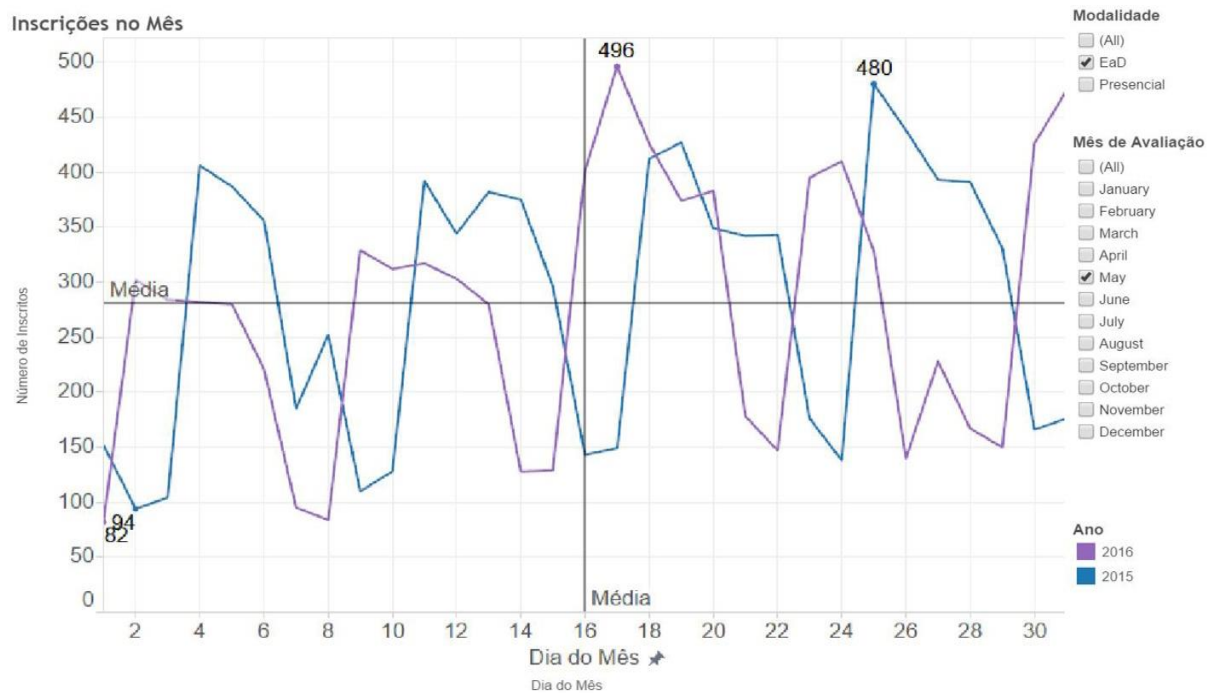


Figure 10 - Monthly comparisons among registered.

It is possible the interaction with the tool, where several filters are unrestricted so that it is possible to perform information crossings, for example: comparisons among years, courses and centers and therefore verify which courses have the best performance in the entrance exam and on which learning centers; what is the candidate's age profile by course, since it is possible to visualize the registered candidates per course (Figure 11 and figure 12).

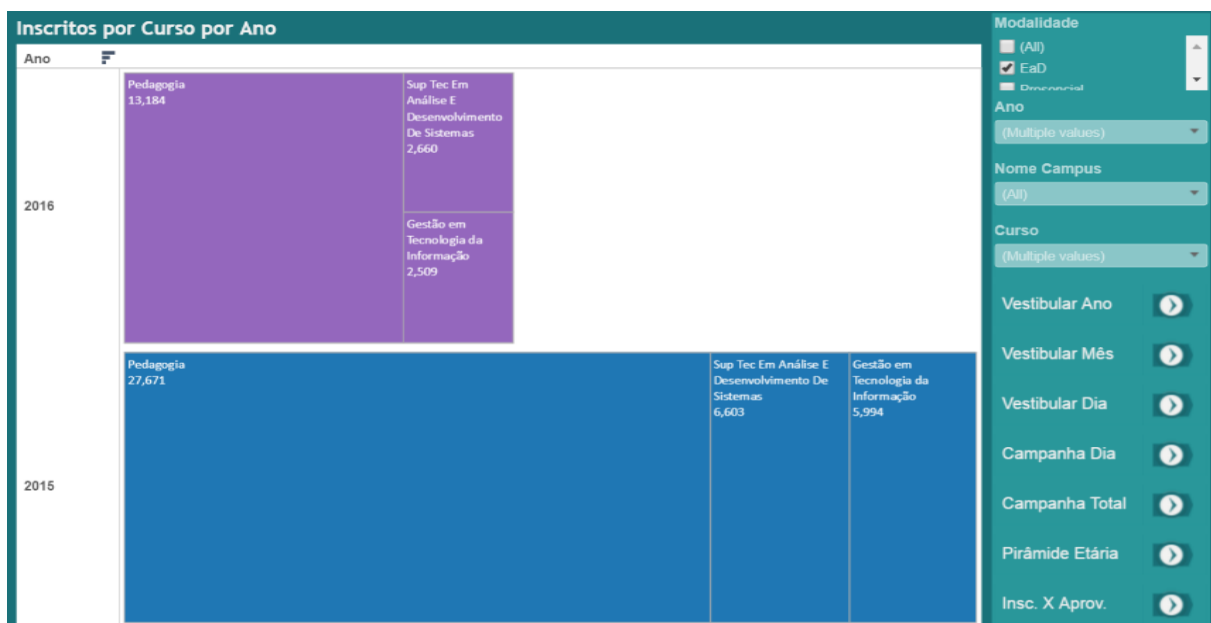


Figure 7 - Performance comparisons among courses

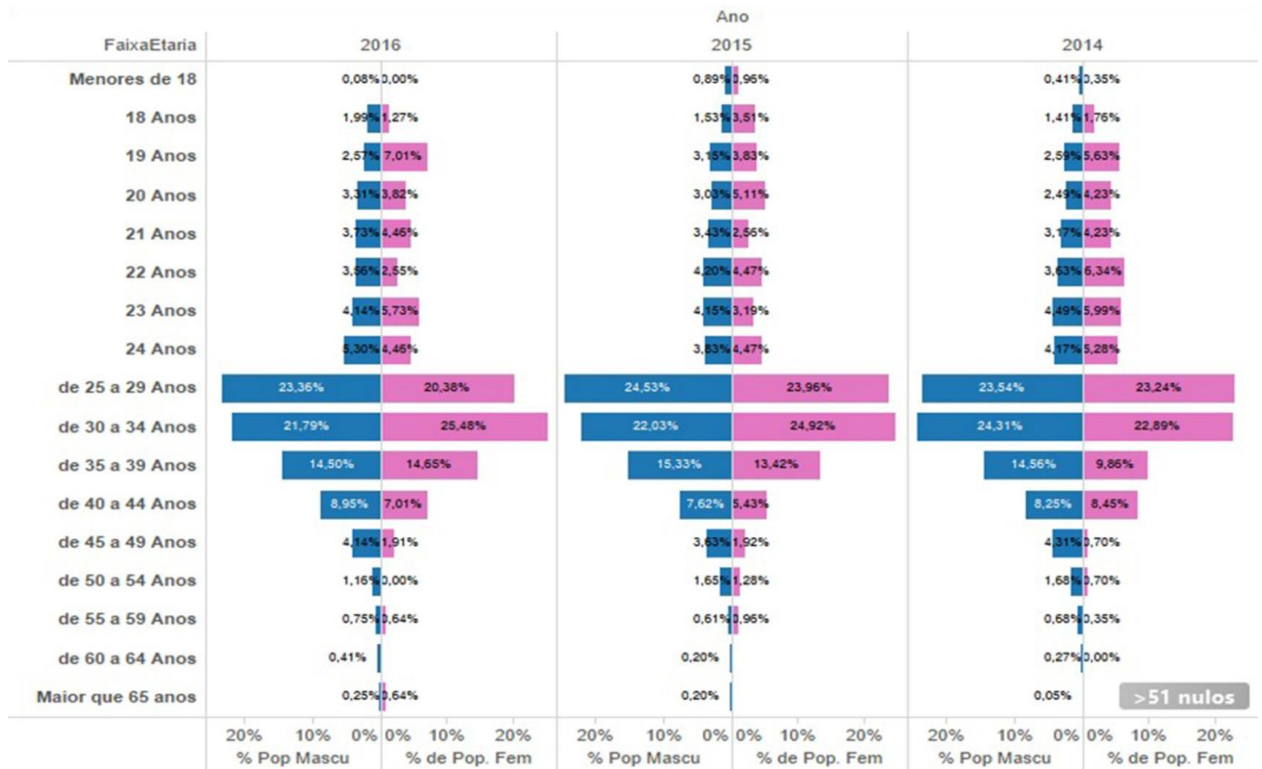


Figure 12 - Performance comparisons among age profile

In this study, problems were encountered; some requested information was not in the database on the correct manner, but they were as a description field. It was necessary that the Manager responsible for this system performed the necessary changes in the system so that the data visualization could be presented correctly.

For the study called learning centers, the result of the analysis was presented to the Managers of the learning centers restrictively, where each Manager view only the information of his/her center. For administrative and pedagogical Directors, the access is not restricted, and then it is possible the data visualization of all learning centers. In this study problems were encountered in the consolidation of the data, the HE has learning centers that are allies, and some are its own learning centers; these differences were not entered in the database, which made it difficult at the beginning of this study, to address this situation the responsible Manager made the appropriate parameterization on the system.

Hence for the study of the learning centers the objectives were: reflection, monitoring and analysis, the data visualization technique was used and the information was presented. The visualization of the locations where HE has learning centers and their distribution by types, resulted in a visualization easily identified where there is a higher concentration of learning centers; it is possible the interaction and consultation by state or just searching for a

specific learning center (figure 13), therefore, going into detail on a chosen learning center and visualize the number of students by course and distribution by gender (figure 14).



Figure 13 Geographical distributions of the centers in the HE

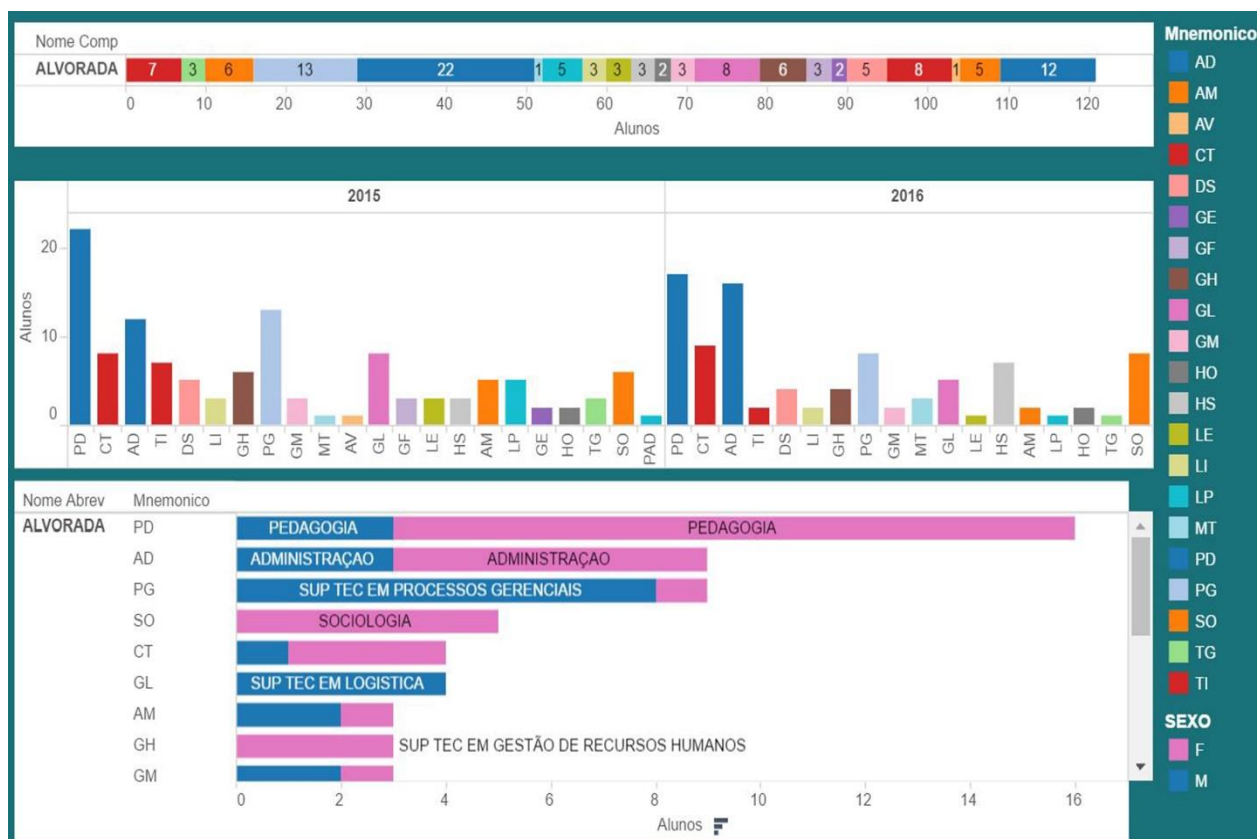


Figure 8- Quantity of students by course and gender of a center

5. Conclusion

Neste trabalho, foi possível, utilizando o modelo de referência, realizar a implantação do Learning Analytics em IES utilizando ferramentas de integração, de banco de dados e de análise de dados. A implantação de uma ferramenta automatizada para alcançarem a informação para as tomadas de decisão, foi realizada com sucesso. Os participantes da equipe desempenharam papéis fundamentais para que os dados extraídos se transformassem em informações relevantes.

In this work, it was possible; using the reference model, perform the implementation of Learning Analytics in the HE using integration, database and data analysis tools. The implementation of an automated tool to achieve information for decision-making was successfully performed. The team members played key roles so that the extracted data were transformed into relevant information.

The results presented show the tool's effectiveness for the HE and brought benefits such as: agility in decision-making, information accessible by online means, consolidation of data in a single tool, profile knowledge of its population. The treatment on data quality turned out to be a valuable byproduct in this deployment.

The contribution of this work is on the demonstration of implementation of LA and its results; it is worth noting that the selected tools showed excellent performance and ease of use, and that the deployment method can serve as a complement to the use of other tools. Finally, as an academic contribution, it is expected that this work arouse interest from others academics for research in the implementation of LA.

Suggestions for future work: expand of data source and use of others data mining techniques.

Referências

Campbell, J.P., DeBlois, P.B. & Oblinger, D.G. (2007). Academic Analytics: A New Tool for a New Era. *EDUCAUSE Review*, 42(4), 40-42.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318-331.

Fox, J., & Andersen, R. (2005). Using the R statistical computing environment to teach social statistics courses. Department of Sociology, McMaster University.

Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M.,... & Arnab, S. (2014). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*.

Johnson, L., R. Smith, H. Willis, A. Levine, Haywood, K. (2011). Learning Analytics. The 2011 Horizon Report. Austin, Texas: The New Media Consortium.

Lias, T. E., & Elias, T. (2011). Learning Analytics: The Definitions, the Processes, and the Potential.

Moraes, E.M., Silva, M.T., Souza, M.C. (2016). Models to Implement Learning Analytics: A Literature Review. Pomsmeeting, 065-0931.

Siemens, G. (2013). Learning analytics the emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.

Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE review*, 46(5), 30.

Shum, S. B. (2012). Policy Brief - Learning Analytics. Moscow, Russian Federation: UNESCO.

Van Barneveld, A., Arnold, K. & Campbell, J. (2012). Analytics in Higher Education: Establishing a Common Language. *EDUCAUSE Learning Initiative (ELI)*, 1, 1-11.

6 CONSIDERAÇÕES FINAIS

Muitas tecnologias estão surgindo na educação e na modalidade a distância o tema se torna mais complexo, pois tem como característica a utilização de variados recursos das TICs para os processos de gestão acadêmica, administrativa e o desenvolvimento das atividades para o processo de ensino e aprendizagem.

O desafio para as IES está em organizar para a análise de forma agrupada as informações e assim conseguir encontrar as que serão adequadas e permitirão um maior apoio no processo de tomada de decisão.

6.1 Resultados

Nesta dissertação, foi possível verificar que a capacidade de ler, analisar e consolidar os dados para a extração do conhecimento das múltiplas fontes de dados, auxilia na gestão da educação e pode apoiar a IES no gerenciamento de várias atividades que fazem parte do ciclo de vida do aluno, como na captação, ingresso e realização do curso. Desta forma, o LA foi implantado e aplicável na IES, por meio do modelo de referência, atingindo-se o objetivo geral proposto.

Um dos objetivos específicos desta dissertação foi alcançado com o mapeamento sistemático da literatura, para buscar modelos de implementação do LA e assim auxiliar na escolha de possíveis direções. Poucos modelos foram encontrados e artigos não foram localizados artigos que indicassem o sucesso da implantação desses modelos. O modelo de Chatti et al. (2012), por sua fácil adoção e entendimento, foi adotado para implantação na IES e elaboração desse estudo.

O modelo ajudou com sucesso o entendimento do LA e também norteou a implantação na IES. Na etapa seguinte outro objetivo específico foi alcançado: o modelo foi aplicado e foram escolhidos os sistemas de gestão acadêmica, o sistema de vestibular e o ambiente virtual de aprendizagem da IES. Por meio das ferramentas *Talend*, *Vertica* e *Tableau* e da utilização de técnicas de visualização foi possível a extração do conhecimento. As escolhas das ferramentas trouxeram benefícios, como informações acessíveis de forma *on-line*, consolidação dos dados de diferentes sistemas em uma única visualização e rapidez no acesso ao grande volume de dados.

Na etapa seguinte, outros dois objetivos específicos foram alcançados por meio da identificação das necessidades de recursos humanos e escolha da equipe, que com seus

conhecimentos técnicos e em conjunto com a pesquisadora, realizaram a implantação de uma aplicação analítica. Havia a necessidade de as pessoas envolvidas conhecerem profundamente os sistemas que foram utilizados na pesquisa, o processo da IES cada qual em sua área.

A identificação das dificuldades na implantação do LA tornou-se um subproduto neste estudo, pois foi preciso qualificar os dados e, em alguns casos, foi necessário a inserção de parâmetros para uma melhor compreensão.

Ainda nesta etapa foi desenhado o fluxo dos dados para o repositório de dados integrado, que facilitou a compreensão e periodicidade com que estes são atualizados para a ferramenta analítica, por meio de uma técnica de mineração de dados, a árvore de decisão para o LA.

Por fim, percebe-se que a tecnologia de LA pode ser utilizada em todos os níveis da IES e tornar-se uma ferramenta importante para a gestão da educação. Para isso, a IES necessitará de uma equipe dedicada e multidisciplinar e a confiabilidade das informações dependerá do entendimento correto da solicitação dos clientes pela equipe e que os dados fornecidos estejam corretos.

6.2 Contribuições

Este trabalho demonstra os resultados da implantação de um modelo de referência de LA encontrado na literatura que contribui no auxílio à gestão da educação. A principal contribuição é demonstrar a possível melhora da gestão educacional quando se implanta o LA. As ferramentas escolhidas mostraram desempenho e facilidade de uso e espera-se que o método de implantação possa servir de complemento para a utilização de outras ferramentas. Por fim, como contribuição acadêmica, espera-se que esse trabalho desperte interesse de outros acadêmicos pela pesquisa na implantação do LA.

6.3 Limitações e recomendações para trabalhos futuros

Quanto às limitações da pesquisa destaca-se a não abordagem ampla sobre ética neste trabalho. Segundo Slade e Prinsloo (2013), o uso dos dados enfrenta alguns desafios éticos como: consentimento informado, privacidade, classificação e gestão dos dados. Claro que o estudo apresentado é de uso restrito, no qual cada cliente tem acesso às informações pertinentes a sua área de atuação na IES, mas o assunto não é abordado amplamente nessa dissertação.

Outro ponto é que, como requereria um tempo considerável para que todos os sistemas pudessem ser analisados, este estudo ficou restrito a três importantes sistemas da IES. Não foi possível, ainda, recolher os resultados da implantação da árvore de decisão, pois só seriam aferidos no próximo semestre, ocasião em que será possível fazer ajustes nas variáveis.

Para trabalhos futuros sugere-se o mapeamento de outros modelos para a implantação do LA e dimensões a serem avaliadas, o que permitiria comparar a eficácia entre os modelos. Para a continuidade deste estudo, pode-se considerar a utilização de mais sistemas da própria IES e ampliar a área de atuação do LA. Além disso, cabe buscar outras técnicas de mineração de dados e ampliar a utilização em outras modalidades de ensino.

REFERÊNCIAS

BICHSEL, Jacqueline. Analytics in higher education: Benefits, barriers, progress, and recommendations. Louisville, Co: **EDUCAUSE Center for Applied Research**, 2012.

CAMPBELL, John P.; DEBLOIS, Peter B.; OBLINGER, Diana G. Academic analytics: A new tool for a new era. **EDUCAUSE review**, v. 42, n. 4, p. 40, 2007.

CASTELS, Manuel; EM REDE, A. Sociedade. A era da informação: economia, sociedade e cultura. **São Paulo: Paz e Terra**, v. 3, 1999.

CHATTI, Mohamed Amine et al. A reference model for learning analytics. **International Journal of Technology Enhanced Learning**, v. 4, n. 5-6, p. 318-331, 2012.

FERREIRA, Sergio André; ANDRADE, António - **Analytics na educação: Da gestão personalizada da atividade de ensino a definição de políticas institucionais**. In I Seminário Internacional “Educação, Territórios e Desenvolvimento Humano”, Porto, Portugal, 23-24 Julho 2015.

FREITAS, Sara et al. Foundations of dynamic learning analytics: Using university student data to increase retention. **British Journal of Educational Technology**, v. 46, n. 6, p. 1175-1188, 2015.

GRAJEK, Susan; ROTMAN, David. Top-Ten IT Issues, 2014: Be the Change You See. **Educause Review**, v. 49, n. 2, p. 10, 2014.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Inep libera **informações sobre ensino superior**. Brasília: Inep, 2014. Disponível em: Acesso em: 20 mai. 2016.

MARTINS, Antonio Roberto; MELLO, Carlos Henrique Pereira; TURRIONI, João Batista. **Guia para elaboração de monografia e TCC em Engenharia de Produção**. São Paulo: Editora Atlas, 2013.

MENDONÇA, José Ricardo Costa et al. Competências Eletrônicas de Professores para Educação à distância no Ensino Superior no Brasil: discussão e proposição de modelo de análise. **Fórum da Gestão do Ensino Superior nos países e regiões de língua portuguesa, II**, 2012.

MOORE, Michel G. et al. **Educação a distância: uma visão integrada**. São Paulo: Cengage Learning, 2011.

PAPAMITSIOU, Zacharoula K.; ECONOMIDES, Anastasios A. Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. **Educational Technology & Society**, v. 17, n. 4, p. 49-64, 2014.

SHARKEY, Mike. Academic analytics landscape at the University of Phoenix. In: **Proceedings of the 1st International Conference on Learning Analytics and Knowledge**. ACM, p. 122-126, 2011.

SHUM, S. B. Policiy brief-Learning Analytics. **Moscow, Russian Federation: UNESCO**, 2012.

SIEMENS, George; LONG, Phil. Penetrating the Fog: Analytics in Learning and Education. **EDUCAUSE review**, v. 46, n. 5, p. 30, 2011.

SLADE, Sharon; PRINSLOO, Paul. Learning analytics: ethical issues and dilemmas. **American Behavioral Scientist**, 57(10) pp. 1509–1528, 2013.

VAN BARNEVELD, Angela; ARNOLD, Kimberly E.; CAMPBELL, John P. Analytics in higher education: Establishing a common language. **EDUCAUSE learning initiative**, v. 1, p. 1-11, 2012.

THIOLLENT, Michel. **Metodologia da Pesquisa-ação**. 14º edição. Editora São Paulo: Cortez; 2005.

YARMOHAMMADIAN, Mohammad Hossein et al. Evaluation of distance education programs based on the NADE-TDEC 2009-2010. **Procedia-Social and Behavioral Sciences**, v. 28, p. 117-119, 2011.

ANEXO A: IDENTIFICANDO O DESEMPENHO DO ALUNO DA EAD: RELAÇÃO ENTRE AS APROVAÇÕES E AS INTERAÇÕES NO AVA

Artigo apresentado e publicado nos anais do XXXV Encontro Nacional de Engenharia de Produção: Perspectivas Globais para a Engenharia de Produção, realizado em Fortaleza CE, em 14/ 10/ 2015.

XXXV ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO
Perspectivas Globais para a Engenharia de Produção
Fortaleza, CE, Brasil, 13 a 16 de outubro de 2015.

IDENTIFICANDO O DESEMPENHO DO ALUNO DA EAD: RELAÇÃO ENTRE AS APROVAÇÕES E AS INTERAÇÕES NO AVA

Elisangela Monaco de Moraes (UNIP)
epmonaco@gmail.com
Marcia Terra da Silva (UNIP)
mtdsilva@usp.br
Eudes Azevedo Batista (UNIP)
eudesazevedo@uol.com.br



Este artigo apresenta um estudo sobre o desempenho dos alunos em um curso oferecido na modalidade a distância. Nesse sentido o objetivo deste trabalho é verificar se a relação entre o desempenho dos alunos nas realizações das atividades propostas no ambiente virtual de aprendizagem (AVA) está relacionada com as aprovações e reprovações, bem como se tem relação com a continuação dos alunos no curso de um semestre para o outro. Para isto foram coletados dados obtidos dos sistemas de informação utilizados na EAD, especificamente do sistema de gestão acadêmica e do AVA e a partir da análise feita utilizar alguns indicadores para mapear a permanência do aluno no curso. Foram utilizados os indicadores de desempenho (número de aprovações no semestre, número de atividades realizadas no ambiente virtual de aprendizagem e as notas obtidas nas avaliações realizadas no ambiente virtual de aprendizagem) e as suas relações analisando os dados gerados pelas ações dos alunos. Os dados indicaram que os alunos com alto índice de aprovação, considerados como sucesso, têm participação efetiva no AVA uma vez que a porcentagem de entrega das atividades disponibilizadas no AVA é muito alta e as notas destas atividades acima da média e consequentemente são os alunos que continuam de um semestre para o outro. Diante destes resultados, pode-se perceber que, os dados obtidos pelos logs do AVA são significantes indicadores de sucesso em um curso uma vez que foram relacionados com os altos índices de aprovação nas disciplinas e a continuação dos alunos no curso.

Palavras-chave: alunos persistentes, educação a distância, mineração de dados educacionais



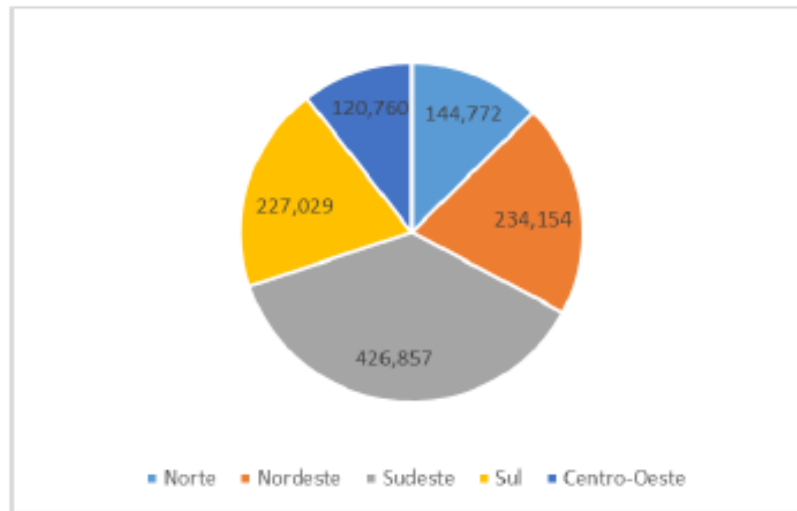
1. Introdução

A mudança de uma sociedade industrial para uma sociedade orientada às informações Castells (2000) criou novos estilos de vida, de consumo e de aprendizado. Nesta nova sociedade, a Educação a distância está situada como uma modalidade em expansão que atende demandas crescentes.

A educação a distância (EAD) é uma modalidade em expansão com a chegada da internet e com o desenvolvimento das tecnologias da informação e comunicação (TICs) e é vista como uma modalidade com potencial para viabilizar a demanda por ensino superior do Brasil, pois a EAD democratiza o ingresso à educação superior para a parcela da população que não tinha oferta de cursos de graduação presencial em suas localidades. A expansão desta modalidade fica clara nos censos de educação superior realizado pelo INEP e pela ABED.

De acordo com o resumo técnico do censo da educação superior de 2013 publicado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep/MEC, 2013), que apresenta informações sobre a educação superior no Brasil, houve um crescimento no número de matrículas em cursos de EAD de 3,6% no período de 2012 a 2013 e já contam a participação de mais de 15% na matrícula de graduação. Em 2013, 1.153.572 milhão de alunos estavam matriculados em cursos de graduação a distância divididos conforme figura 1 segundo a Unidade da Federação e a maior frequência dos discentes está em IES da rede privada.

Figura 1- Número de matrículas segundo a Unidade da Federação



Dados apresentados pela Associação Brasileira de Educação a Distância (ABED), relatório estatístico anual de atividades de EAD (2013) nas comunidades acadêmicas, profissionais e corporativas do Brasil, em 2013 foram realizadas 692.279 matrículas em cursos autorizados/credenciados totalmente a distância, com a oferta de 1772 cursos sendo a maior parte delas (38%) em licenciatura, seguindo-lhe os cursos tecnológicos (27%) vale ressaltar que entre os cursos os de maior sucesso identificados como grupos isolados são os que ensinam as engenharias (Civil, Elétrica, de Produção e outras), uma área relativamente nova na modalidade a distância. O perfil etário e ocupacional reforça o diferencial do aluno EAD que na maioria é do sexo feminino (57,5%), entre 31 e 40 anos que estuda e trabalha.

O crescimento do número de alunos dificulta o acompanhamento de sua aprendizagem no ambiente virtual, exigindo um grande esforço do professor visto que o número de informações é bem grande. No entanto, este acompanhamento pode auxiliar o aluno a obter êxito visando a sua permanência no curso. Gerenciar estes dados produzidos pelos alunos de forma a torná-los em informação qualificada pode ajudar a identificar característica de comportamento colaborando para a identificação prematura de alunos em risco permitindo orientá-los a como obter melhores resultados.

Em qualquer curso a meta primordial é que o aluno obtenha sucesso e que consiga se formar. Nesse sentido segundo Mill e Carmo (2012) a evasão/ permanência do aluno e os sistemas informáticos para a gestão pedagógica da EAD são alguns dos principais desafios dos gestores de sistemas da EAD.



Neste contexto alguns trabalhos Er (2012); Morris *et al.* (2005); Crampton *et al.* (2012) analisam as características e as variáveis, conforme tabela1, de desempenho do aluno que podem se relacionar com os motivos de sucesso ou fracasso nos cursos da Educação a Distância.

Tabela 1 - Características e variáveis estudadas

Trabalho	Características e variáveis
Er	Assiduidade de cada semana no AVA Notas nos exames parciais Notas das tarefas realizadas no AVA Notas dos exames finais
Morris <i>et al.</i>	Frequência (acesso ao conteúdo, número de interações no fórum de discussão) Duração (tempo gasto na visualização de conteúdo, tempo gasto com as interações no fórum de discussão)
Crampton <i>et al.</i>	Notas dos alunos Logs de acesso ao site Logs do fórum de aprendizagem

Nesse sentido o objetivo deste trabalho é verificar se a relação entre o desempenho dos alunos nas realizações das atividades propostas no ambiente virtual de aprendizagem (AVA) está relacionada com as aprovações e reprovações, bem como se tem relação com a continuação dos alunos no curso de um semestre para o outro.

Para isto serão coletadas as informações geradas pelo sistema de gestão acadêmica e pelo AVA e serão verificadas se a continuação dos alunos de um semestre para o outro, tem relação com o número de aprovações nas disciplinas e que estas aprovações têm como fator importante a participação mais efetiva do aluno nas atividades propostas no AVA.

2. Fundamentação teórica

A bibliografia aponta para dois tipos de métodos principais utilizados em pesquisas que buscam identificar características de desempenho dos alunos na Educação a Distância. Trabalhos com instrumentos de coleta de dados como questionários, formulários. Entrevistas e banco de dados e que depois são analisados de um lado por métodos estatísticos Morris *et al.* (2005); Souza *et al.* (2012); Chaves & Andreoli (2013) e por outro lado a mineração de dados



Hamalainen *et al* (2004); Romero e Ventura (2013) outra importante vertente para examinar o desempenho do estudante.

Dentre as pesquisas envolvendo o estudo de fatores que são determinantes para acompanhar o desempenho do aluno na EAD pode-se destacar o trabalho de Gottardo *et al* (2014) que usou com referência a “Teoria de Interação em Educação a Distância” e a partir das interações relevantes, escolheram-se três dimensões para representar os estudantes em um AVA: perfil de uso do AVA, interação estudante-estudante e interação bidirecional estudante-professor.

No estudo realizado por Er (2012) os seguintes atributos foram avaliados: assiduidade de cada semana no ambiente virtual de aprendizagem, notas nos exames parciais, notas nas tarefas e notas no exame final. Os alunos então foram classificados em dois grupos de acordo com o seguinte critério nota ≥ 60 é sinal de sucesso e ≤ 60 é sinal de fracasso.

Alguns estudos investigaram as variáveis demográficas como Byrne e Flood (2008) afirmam que não há mais espaço na educação para sustentação da tese de que há diferenças significativas decorrentes do gênero, que possam vir a se manifestar sobre o desempenho acadêmico do aluno já Levy (2007) descreve que estas variáveis influenciam apenas indiretamente na situação.

3. Procedimentos metodológicos

Este estudo utilizou dados obtidos dos sistemas de informação utilizados na EAD, especificamente do sistema de gestão acadêmica e do AVA. Foram realizadas análises estatísticas para explorar a relação entre o número de aprovações nas disciplinas com a participação dos alunos e realização das atividades no AVA e com sua persistência no curso.

Segundo Byrne e Flood (2008) dados demográficos ajudam a conhecer o perfil socioeconômico do aluno não seu desempenho, neste estudo serão analisados os seguintes dados para composição do perfil demográfico: região, ano de conclusão do ensino médio, idade e sexo.

Neste trabalho optou-se a partir da análise feita das variáveis encontradas na literatura utilizar alguns indicadores para mapear a permanência do aluno no curso. Primeiramente será verificado o total de disciplinas aprovadas no semestre 1 e serão divididos em três grupos: alta aprovação, aprovação mediana e baixa aprovação conforme mostra a figura 2. Serão verificadas as porcentagens de matrículas efetuadas para o semestre 2 de cada um dos grupos. De acordo com o número de aprovações neste semestre 2 os alunos serão divididos



novamente em grupos: alta aprovação, aprovação mediana e baixa aprovação. Serão verificadas as porcentagens de matrículas por grupo para o semestre 3.

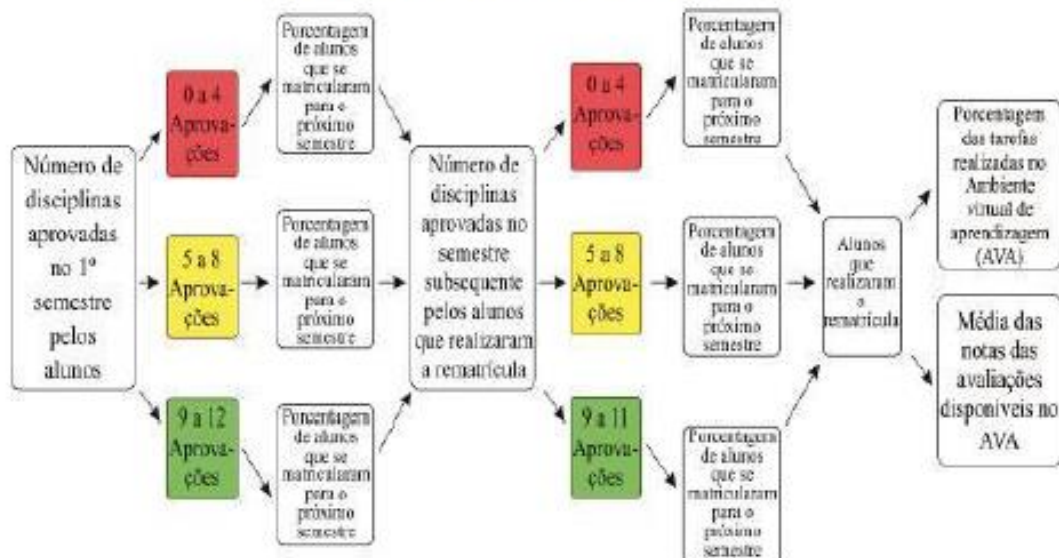
Para estes alunos que estão no semestre 3 do curso serão avaliadas as demais variáveis adotadas neste estudo que são a porcentagem de tarefas realizadas no AVA e as médias de notas das avaliações realizadas no AVA.

A seleção dos dados armazenados é constituída pelos mesmos atributos e alguns atributos irrelevantes foram eliminados. A partir da seleção feita, optou-se por utilizar os seguintes atributos:

- Total de disciplinas aprovadas por semestre: verifica-se a quantidade total de disciplinas na grade curricular que há no semestre e a partir desta informação busca-se o número de aprovações e o número de dispensas nas disciplinas e com estes dados sabe-se o número total de aprovações/ dispensas no semestre.
- Porcentagem das tarefas realizadas no AVA: verifica-se a quantidade máxima de atividades disponíveis no semestre e a quantidade total realizada pelos alunos e calcula-se a porcentagem de atividades realizadas do máximo possível.
- Média das notas das avaliações disponíveis no AVA: calcula-se a nota média das atividades que foram configuradas para gerarem nota ao aluno. Como o conjunto das atividades por disciplina pode ter nota máxima diferenciada (entre 0 e 10), aqui se verifica o quanto ele conseguiu obter do máximo disponível. Logo, é feita uma média dessas notas considerando também os respectivos pesos.

Considerando estes atributos os alunos foram divididos em três grupos de acordo com o número de aprovações no semestre 1, sendo que os números máximos são de 12 aprovações, da seguinte forma: baixa aprovação são os alunos que têm de 0 a 4 aprovações, aprovação mediana são alunos que têm de 5 a 8 aprovações e alta aprovação que são os alunos que têm de 9 a 12 aprovações. Foi verificado a porcentagem, por grupo, do número de alunos que realizaram a rematrícula para o semestre 2. Estes alunos foram novamente divididos da mesma forma em três grupos, sendo que o número máximo de aprovações para o semestre 2 é de 11 disciplinas. Novamente foi verificada, por grupo, a porcentagem de alunos que realizaram a matrícula para o semestre 3. Para este grupo de alunos foram verificadas a porcentagem de realizações das atividades no AVA e a média das notas das avaliações disponíveis no AVA.

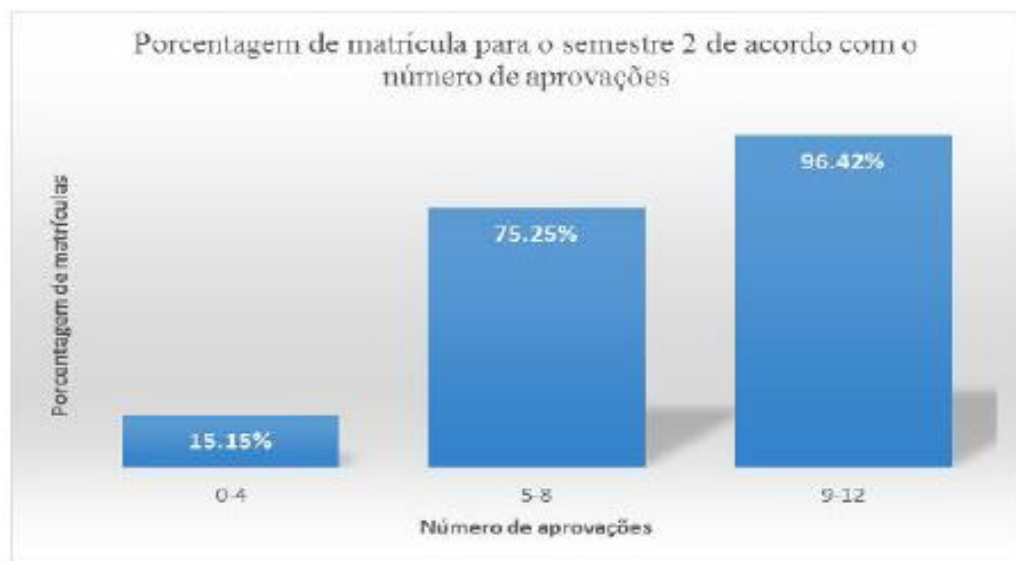
Figura 2 - Árvore de decisão para chegar ao aluno de sucesso



4. Resultados

Os resultados demonstram que os alunos que têm muitas disciplinas reprovadas, ou seja, que têm baixo número de aprovações, de 0 a 4 no semestre 1, são os que menos realizaram rematricula para o semestre 2 com apenas 15,15 % dos alunos conforme demonstra a figura 3, foram utilizados para este resultado a variável nota final do aluno nas disciplinas.

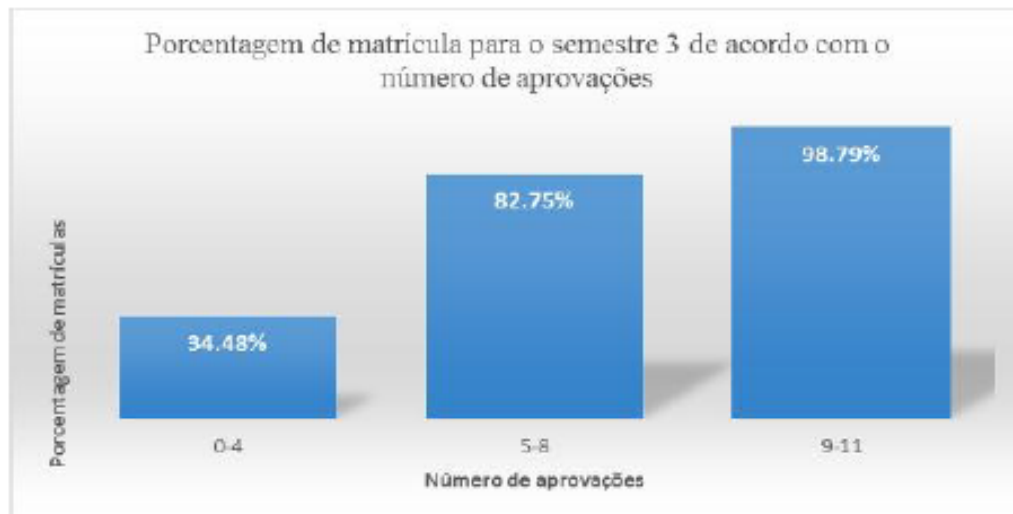
Figura 3 - Porcentagem de rematricula de acordo com o número de aprovações





Dos alunos, de todos os grupos, que realizaram a rematrícula para o semestre 2 foi realizado novamente a mesma divisão que é de acordo com o número de aprovações, utilizando a variável nota do aluno. Percebe-se que o mesmo acontece, ou seja, quem tem baixo índice de aprovação são os que menos realizaram a rematrícula sendo que dos alunos que tiveram de 0 a 4 aprovações apenas 34.48% foram para o semestre 3 conforme mostra a figura 4.

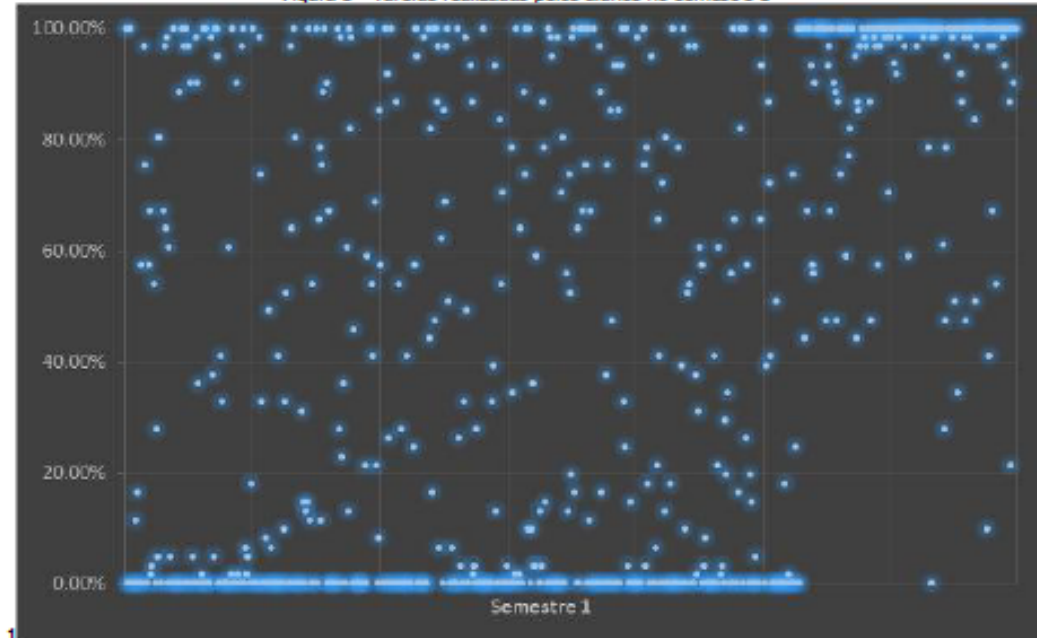
Figura 4 - Porcentagem de rematrícula de acordo com o número de aprovações



Os alunos no semestre 1 realizaram em média 45,49% das atividades disponibilizadas no ambiente virtual de aprendizagem conforme mostra a figura 5 e a média de notas obtidas nas atividades que valem nota é de 3,57, foram utilizadas as variáveis quantidade de atividades realizadas pelos alunos e as notas obtidas nas avaliações disponíveis no AVA.



Figura 5 - Tarefas realizadas pelos alunos no semestre 1



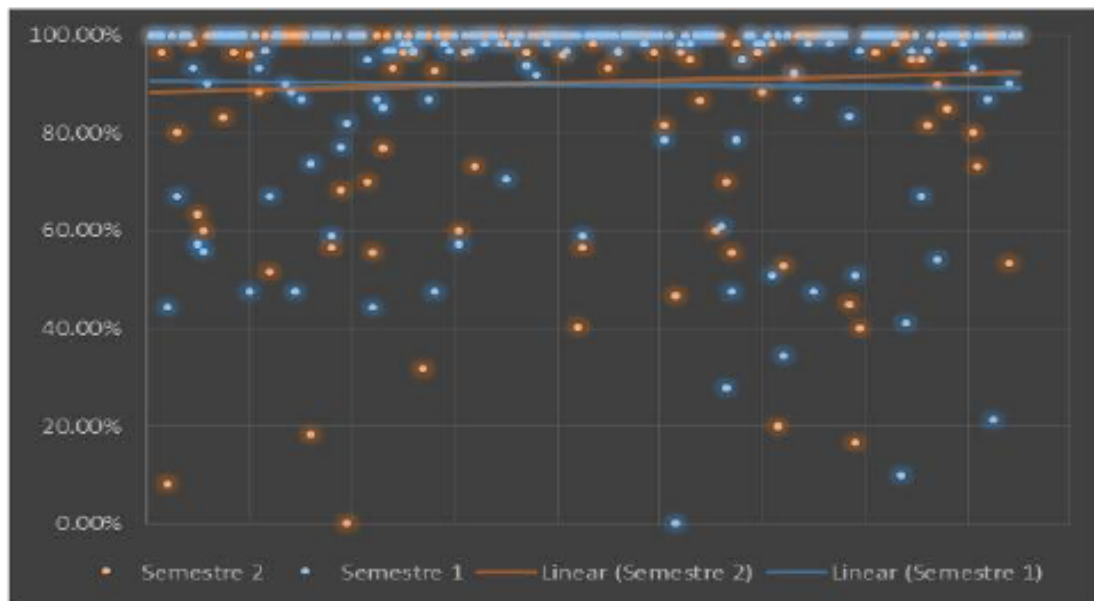
1

Fica notório que os alunos que obtiveram o sucesso até esta fase do curso têm altas porcentagens de realizações nas atividades disponíveis no AVA como pode-se observar na figura 6, em média os alunos realizaram 89.87% das atividades disponíveis no AVA no semestre 1 e 90.38% no semestre 2, para este resultado a variável quantidade de atividades realizadas pelos alunos foi utilizada.

Para os alunos com baixo índice de aprovação nas disciplinas no semestre 1 foram realizadas em média 31.23% das atividades disponíveis no AVA.

Da mesma forma se repete em relação as notas obtidas pelos alunos nas avaliações disponíveis no AVA, onde as médias de notas no semestre 1 para os alunos com alto índice de aprovação é de 7.23 e no semestre 2 é 7.27, a variável notas obtidas nas avaliações disponíveis no AVA foi utilizada para este resultado.

Figura 6 - Porcentagem das tarefas realizadas no AVA pelos alunos com alto índice de aprovação



Os dados demográficos demonstram que os alunos em média têm 32 anos, terminaram o ensino médio a 12 anos, a maioria do sexo masculino e se localiza na região sudeste. O mesmo perfil se repete para os alunos com índice alto de aprovação que portanto são os considerados alunos de sucesso.

5. Discussões e considerações finais

O principal objetivo deste estudo, era medir três indicadores de desempenho e as suas relações (número de aprovações no semestre, número de atividades realizadas no ambiente virtual de aprendizagem e as notas obtidas nas avaliações realizadas no ambiente virtual de aprendizagem) através do mapeamento destas variáveis, foram extraídos e analisados os dados gerados pelas ações dos alunos, estes dados indicam que os alunos com alto índice de aprovação, considerados como sucesso, têm participação efetiva no AVA uma vez que a porcentagem de entrega das atividades disponibilizadas no AVA é muito alta e as notas destas atividades acima da média e consequentemente são os alunos que continuam de um semestre para o outro.



Diante dos resultados apresentados, pode-se perceber que, os dados obtidos pelos logs do AVA são significantes indicadores de sucesso em um curso uma vez que foram relacionados com os altos índices de aprovação nas disciplinas e a continuação dos alunos no curso.

Nesse sentido após as análises dos indicadores Hamalainen *et al* (2004) sugere um modelo para a predição de alunos em risco e Macfadyen e Dawson (2010) o desenvolvimento de relatórios para alertar professores sobre os alunos em risco.

Como este estudo utilizou técnicas de estatísticas como ponto de partida, apenas coletando os dados e fazendo as análises dos logs e suas correlações, nos próximos trabalhos sugere-se a utilização de mineração de dados para descobrir padrões de comportamento analisando a combinação de vários outros aspectos. Adicionalmente é necessário ampliar a mais de um curso e inclusive a mais de uma Instituição de Ensino Superior.

REFERÊNCIAS

- BYRNE, Marann; FLOOD, Barbara. **Examining the relationships among background variables and academic performance of first year accounting students at an Irish University.** *Journal of Accounting Education*, v. 26, n. 4, p. 202-212, 2008.
- CASTELLS, Manuel; GERHARDT, Klauss Brandini. **A sociedade em rede.** São Paulo: Paz e Terra, 2000.
- CENSO EaD.br: **relatório analítico da aprendizagem a distância no Brasil 2013** = Censo EaD.br: analytic report of distance learning in Brazil/[traduzido por Maria Thereza Moss de Abreu]. – Curitiba: Ibpex, 2014.
- CRAMPTON, Andrea; RAGUSA, Angela T.; CAVANAGH, Heather. **Cross-discipline investigation of the relationship between academic performance and online resource access by distance education students.** *Research in Learning Technology*, v. 20, 2012.



DE PAULA, Eduardo; CHAVES, Silva; ANDREOLI, Célia Regina. **Qual o impacto do conhecimento de informática no desempenho acadêmico dos alunos de EaD?**. Revista de Administração IMED, v. 3, n. 2, 2013.

ER, Erkan. **Identifying At-Risk Students Using Machine Learning Techniques: A Case Study with IS 100**. International Journal of Machine Learning and Computing, v. 2, n. 4, 2012

GOTTARDO, Ernani; KAESTNER, Celso Antônio Alves; NORONHA, Robinson Vida. **Estimativa de Desempenho Acadêmico de Estudantes: Análise da Aplicação de Técnicas de Mineração de Dados em Cursos a Distância**. Revista Brasileira de Informática na Educação, v. 22, n. 01, p. 45, 2014.

HAMALAINEN, W., SUHONEN, J., SUTINEN, E. and TOIVONEN, H.: **Data mining in personalizing distance education courses**. World Conference on Open Learning and Distance Education, Hong Kong (pp. 1-11). (2004).

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Inep libera informações sobre ensino superior**. Brasília: Inep, 2013. Disponível em: . Acesso em: 05 maio. 2015.

LEVY, Yair. **Comparing dropouts and persistence in e-learning courses**. Computers & education, v. 48, n. 2, p. 185-204, 2007.

MACFADYEN, Leah P.; DAWSON, Shane. **Mining LMS data to develop an “early warning system” for educators: A proof of concept**. Computers & Education, v. 54, n. 2, p. 588-599, 2010.

MILL, Daniel; CARMO, Hermano. **Análise das dificuldades de educadores e gestores da educação a distância virtual no Brasil e em Portugal**. SIED: EnPED-Simpósio Internacional de Educação a Distância e Encontro de Pesquisadores em Educação a Distância 2012, 2012.

MORRIS, Libby V.; FINNEGAN, Catherine; WU, Sz-Shyan. **Tracking student behavior, persistence, and achievement in online courses**. The Internet and Higher Education, v. 8, n. 3, p. 221-231, 2005.

ROMERO, Cristobal; VENTURA, Sebastian. **Data mining in education**. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 3, n. 1, p. 12-27, 2013.

SOUZA, Thiago IA et al. **Avaliando o Desempenho Discente em um AVA: Um Estudo de Caso Utilizando Estatística Multivariada**. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2013.