



Programa Interdisciplinar de Pós-Graduação em

# Computação Aplicada

Mestrado Acadêmico

Wagner Luiz Cambruzzi

GVWISE: Uma aplicação de *learning analytics* para a  
redução da evasão na educação a distância

São Leopoldo, 2014

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS  
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA  
NÍVEL MESTRADO

WAGNER LUIZ CAMBRUZZI

GVWISE: UMA APLICAÇÃO DE *LEARNING ANALYTICS* PARA A REDUÇÃO DA  
EVASÃO NA EDUCAÇÃO A DISTÂNCIA

SÃO LEOPOLDO  
2014



Wagner Luiz Cambruzzi

GVWISE: UMA APLICAÇÃO DE *LEARNING ANALYTICS* PARA A REDUÇÃO DA  
EVASÃO NA EDUCAÇÃO A DISTÂNCIA

Dissertação apresentada como requisito parcial  
para a obtenção do título de Mestre pelo  
Programa de Pós-Graduação em Computação  
Aplicada da Universidade do Vale do Rio dos  
Sinos — UNISINOS

Orientador:  
Prof. Dr. Sandro José Rigo

Co-orientador:  
Prof. Dr. Jorge Luis Victória Barbosa

São Leopoldo  
2014



Wagner Luiz Cambruzzi

GVWISE: UMA APLICAÇÃO DE LEARNING ANALYTICS PARA A REDUÇÃO DA  
EVASÃO NA EDUCAÇÃO A DISTÂNCIA)

Dissertação apresentada à Universidade do Vale do  
Rio dos Sinos – Unisinos, como requisito parcial para  
obtenção do título de Mestre em Computação  
Aplicada.

Aprovado em quinze de abril de 2014

BANCA EXAMINADORA

---

Prof. Dr. Sandro José Rigo – UNISINOS  
Orientador - Presidente

---

Prof. Dr. Jorge Luis Victória Barbosa – UNISINOS  
Co-orientador

---

Prof. Dr. Cristiano André da Costa – UNISINOS  
Membro da Banca

---

Prof. Dr. Eliseo Berni Reategui – UFRGS  
Membro da Banca



Para Valentina, Sandra, Jean, Gema e Vilson.

*If I have seen farther than others,  
it is because I stood on the shoulders of giants.*  
— SIR ISAAC NEWTON





## **AGRADECIMENTOS**

Agradeço ao meu orientador Prof. Dr. Sandro José Rigo e ao meu co-orientador Prof. Dr. Jorge Luis Victória Barbosa, pela liberdade e confiança, além da indiscutível amizade.

Agradeço ao Gilmar Luiz Piaia, diretor da GVDASA Sistemas pela confiança e amizade.

Agradeço a todos os amigos do projeto GVwise pela enorme convivência e aprendizado.

Agradeço meus pais Gema e Vilson pelo apoio incondicional ao longo de toda minha vida.

Agradeço ao meu irmão Jean pela parceria.

Agradeço minha esposa Sandra e minha filha Valentina, pelo amor, cumplicidade e por estarem ao meu lado sempre.

Muito obrigado.



## RESUMO

Aplicações que fazem uso de tecnologias como Mineração de Dados Educacionais (MDE) e *Learning Analytics* (LA) vêm sendo adotadas na mitigação da evasão escolar, disponibilizando informações sobre os alunos que são utilizadas em intervenções pedagógicas. Os trabalhos estudados sobre a implementação destas aplicações priorizam a descrição das técnicas empregadas e existem poucas avaliações da sua utilização em larga escala, além da falta de detalhamento sobre as causas da evasão. Este trabalho apresenta um estudo de fatores envolvidos no fenômeno de evasão escolar e descreve a utilização de um sistema para MDE e LA durante 18 meses em cursos de graduação na modalidade de Educação a Distância. É ampliada a análise dos fatores tradicionalmente monitorados e utilizados nos sistemas de MDA e LA, com a inclusão de elementos associados ao papel exercido pelos docentes e pelo conjunto de aspectos metodológicos de cada instituição. O sistema possui como diferencial a flexibilidade na integração e utilização dos dados gerados no processo de mediação digital, o que permite que necessidades de diferentes ferramentas de apoio sejam disponibilizadas. Resultados positivos destacados são a identificação de perfis de alunos evasores e a realização de intervenções pedagógicas, com redução das médias da evasão.

**Palavras-chave:** Educação - Processamento de Dados. Mineração de dados (Computação). Ensino à distância.



## **ABSTRACT**

Educational Data mining (EDM) and Learning Analytics (LA) applications have been adopted in mitigation of dropout, providing information about students who are employed in pedagogical interventions. The most papers about the implementation of these systems describe the techniques employed, there are few evaluations of their large-scale use, apart from the lack of detail about the causes of dropout. This work presents a study of factors involved in dropout and describes the use of a system for EDM and LA during 18 months for undergraduate courses in distance education. The analysis of the factors traditionally monitored and used in EDM and LA systems is extended, with the inclusion of elements associated with the role exercised by the teachers and by institutional methodological aspects. The system has flexibility in integration and use of data generated in the process of digital mediation, which allows different support tools to be available. Some results are the identification of evaders students profiles and the realization of pedagogical actions with reducing evasion.

**Keywords:** Learning Analytics. Educational Data Mining.



## LISTA DE FIGURAS

Figura 1:	Processo de descoberta de conhecimento. Adaptado de Fayyad . . . . .	28
Figura 2:	Dimensões críticas da <i>Learning Analytics</i> segundo Greller e Drachsler (2012)	34
Figura 3:	Modelo de Referência de <i>Learning Analytics</i> segundo Chatti et al (2012) . .	35
Figura 4:	Processo de <i>Learning Analytics</i> . . . . .	39
Figura 5:	Interações originando múltiplas trilhas . . . . .	40
Figura 6:	Múltiplas trilhas e contexto sistêmico . . . . .	41
Figura 7:	Exemplo de contexto sistêmico do primeiro semestre de 2009. . . . .	41
Figura 8:	Exemplo de contexto da disciplina Estrutura de Dados. . . . .	42
Figura 9:	Diagrama conceitual de Múltiplas Trilhas . . . . .	43
Figura 10:	Modelo Conceitual da Arquitetura . . . . .	44
Figura 11:	Componentes da camada Objetivos . . . . .	44
Figura 12:	Componentes da camada Dados Educacionais . . . . .	46
Figura 13:	Componentes do Cliente Coletor de Dados . . . . .	46
Figura 14:	Servidor de Dados . . . . .	47
Figura 15:	Componentes da camada Técnicas de LA . . . . .	47
Figura 16:	Interface de Lista de Turmas . . . . .	50
Figura 17:	Interface com Lista de Estudantes . . . . .	50
Figura 18:	Interface de Gráficos . . . . .	51
Figura 19:	Interface de Avaliações . . . . .	52
Figura 20:	Interface de Histórico de Ações Realizadas . . . . .	53
Figura 21:	Gráfico de Predição de Risco de Evasão . . . . .	53
Figura 22:	Interface de Histórico de Ações Realizadas . . . . .	54
Figura 23:	Fluxo de importação de dados . . . . .	55
Figura 24:	Visualização de trilhas para acompanhamento pedagógico . . . . .	57
Figura 25:	Componentes do estudo de caso (Ensino Médio) . . . . .	58
Figura 26:	Visualização de interações entre alunos em fórum . . . . .	59
Figura 27:	Alunos por disciplina das classes Não Evadidos e Evadidos . . . . .	61
Figura 28:	Acerto de predição da disciplina 1 . . . . .	62
Figura 29:	Acerto de predição da disciplina 2 . . . . .	63
Figura 30:	Acerto de predição da disciplina 3 . . . . .	63
Figura 31:	Acerto de predição da disciplina 4 . . . . .	64





## LISTA DE TABELAS

Tabela 1:	Comparativo entre os trabalhos relacionados . . . . .	37
Tabela 2:	Comparativo de taxas de evasão e aprovação antes e depois da utilização do modelo proposto . . . . .	65
Tabela 3:	Resultado do Teste-T para Evasão . . . . .	66
Tabela 4:	Resultado do Teste-T para Aprovação . . . . .	66



## LISTA DE SIGLAS

DT	Distância Transacional
EaD	Educação a Distância
EDM	<i>Educational Data Mining</i>
LA	<i>Learning Analytics</i>
MD	Mineração de Dados
MDE	Mineração de Dados Educacionais
RNA	Redes Neurais Artificiais
TEL	<i>Technology Enhances Learning</i>



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>21</b>
1.1	Questão de Pesquisa	23
1.2	Objetivos	23
1.3	Método de Trabalho	23
1.4	Organização do Texto	24
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>25</b>
2.1	Educação a Distância	25
2.2	Learning Analytics	27
2.3	Mineração de Dados	28
2.3.1	Mineração de Dados Educacionais	29
2.4	Trilhas	31
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>33</b>
3.1	Modelos de referência de Learning Analytics	33
3.1.1	Trabalho de GRELLER; DRACHSLER (2012)	33
3.1.2	Trabalho de CHATTI et al. (2012)	35
3.2	Organização de Dados em Trilhas	36
3.2.1	Comparação entre os trabalhos de Organização de Dados em Trilhas	37
3.3	Considerações sobre os trabalhos relacionados	37
<b>4</b>	<b>MODELO PROPOSTO</b>	<b>39</b>
4.1	Elementos Básicos do Modelo de <i>Learning Analytics</i>	39
4.2	Elementos Básicos do Modelo de Tratamento de Múltiplas Trilhas	40
4.3	Arquitetura	43
4.3.1	Camada de Objetivos	44
4.3.2	Camada de Dados Educacionais	45
4.3.3	Camada de Técnicas de LA	47
4.3.4	Camada de Interessados	48
<b>5</b>	<b>IMPLEMENTAÇÃO</b>	<b>49</b>
5.1	Camada de Objetivos	49
5.1.1	Monitoramento e Análises	49
5.1.2	Predição e Intervenção	53
5.2	Camada de Dados Educacionais	53
5.2.1	Cliente Coletor de Dados	54
5.2.2	Servidor de Dados	55
<b>6</b>	<b>AVALIAÇÃO</b>	<b>57</b>
6.1	Visualização Gráfica das Múltiplas Trilhas	57
6.2	Visualização da Interação	59
6.3	Predição da Evasão	60
6.3.1	Disciplina 1	61
6.3.2	Disciplina 2	62
6.3.3	Disciplina 3	62
6.3.4	Disciplina 4	64
6.3.5	Considerações	64

<b>6.4 Reversão de Evasão</b> . . . . .	<b>65</b>
<b>7 CONCLUSÃO</b> . . . . .	<b>69</b>
<b>7.1 Contribuição</b> . . . . .	<b>69</b>
<b>7.2 Trabalhos Futuros</b> . . . . .	<b>70</b>
<b>REFERÊNCIAS</b> . . . . .	<b>71</b>

## 1 INTRODUÇÃO

A Educação a Distância (EaD) é uma modalidade de ensino onde professores e alunos podem estar afastados geograficamente e temporalmente. Essa modalidade tem crescido muito no Brasil (CENSOEAD.BR, 2012) e juntamente com a ampla difusão do uso de sistemas informatizados nos ambientes acadêmicos, colabora no crescimento do volume de dados gerados e armazenados em bases digitais (KOEDINGER et al., 2008). Dados de acesso nos Ambientes Virtuais de Aprendizagem (AVA), histórico escolar, aproveitamento acadêmico, participação dos estudantes em atividades complementares, interação em redes sociais, bem como diversos outros conjuntos de dados disponíveis são potencialmente ricos e podem responder a muitas questões e auxiliar os educadores a entender como os estudantes aprendem (MOSTOW; BECK, 2006). Deve ser destacado também que além do volume considerável de dados gerados nos processos de ensino-aprendizagem com base em mediação digital, existe também um outro aspecto a considerar, que é a sua diversidade. Esta pode ser observada em formatos diferenciados e em origem diferenciadas, a partir de diferentes sistemas ou dispositivos utilizados.

A análise desses dados pode fornecer informações sobre os estudantes que estão em risco de evasão ou que precisam de apoio para obterem sucesso (SIEMENS; LONG, 2011). Esta análise também pode fornecer aos gestores e educadores o conhecimento para melhorarem o ensino, a aprendizagem, a eficiência organizacional e a tomada de decisões. Essas possibilidades de análise têm promovido um crescente interesse na área de Mineração de Dados Educacionais (MDE), do inglês *Educational Data Mining* (EDM) (ROMERO; VENTURA, 2013, 2010; KAMPFF, 2009), e mais recentemente na área de *Learning Analytics and Knowledge* (LAK) também referenciado como *Learning Analytics* (LA) (SIEMENS; BAKER, 2012; CHATTI et al., 2012; GRELLER; DRACHSLER, 2012).

Nos trabalhos observados nestas duas áreas podem ser encontradas iniciativas voltadas para geração de alertas (KAMPFF, 2009), apoio à sistemas de recomendação (DURAND; LAPLANTE; KOP, 2011; TOSCHER; JAHNER, 2010) ou a captura de perfis (LI; MATSUDA, 2011). Ou ainda podem ser identificados trabalhos como o de KOTSIANTIS (2012), que se propõe a prever a nota do estudante em uma disciplina ministrada à distância.

Com base em iniciativas existentes (KAMPFF, 2009; BRA et al., 2013) observa-se um cenário de interesse para a aplicação de recursos diversos de MDE e LA para o apoio em alguns contextos educacionais. Dentro desta perspectiva, pode ser relacionado o fenômeno da evasão, que vem sendo apontado em diversos estudos como um fenômeno de importância e que gera consequências desfavoráveis, tanto na modalidade de ensino de EaD como em outras (LEVY, 2007; ADACHI, 2009).

Com base neste contexto geral, evidencia-se a necessidade de construção de uma arquitetura de software que flexibilize o registro e a utilização dos dados históricos educacionais dos estudantes tendo em vista a aplicação de técnicas de EDM e LA focada no tratamento do fenômeno da evasão. A investigação aqui proposta visa identificar as condições em que a utilização de



recursos de LA pode ser empregada como forma de apoiar processos de mitigação de evasão na EaD, tendo em conta que esta modalidade de ensino utiliza de forma ampla a mediação digital, gerando dados sobre a utilização de recursos pelos alunos durante o acesso aos materiais e durante a sua interação.

Além do aumento no volume de dados disponíveis sobre as ações realizadas e os resultados obtidos pelos alunos, a crescente utilização de dispositivos de computação móvel em ambientes educacionais permite que informações complementares, como a localização física, sejam usadas no registro de contextos. Contexto é qualquer informação que possa ser utilizada para caracterizar a situação de entidades (pessoa, lugar ou objeto) que sejam consideradas relevantes para interação entre um usuário e uma aplicação (DEY, 2001). Desta forma, todos os dados registrados poderiam ser considerados contextos. O registro histórico dos contextos e da atuação da entidade normalmente recebe a denominação de Trilha (DRIVER; CLARKE, 2008; LEVENE; PETERSON, 2002). Com o objetivo de possibilitar a utilização dos dados de trilhas em ambiente educacionais é proposto também neste trabalho um modelo para gerenciamento de múltiplas trilhas denominado MultiTrail. Com base na proposta de gerenciamento e manipulação de trilhas desenvolvida no modelo UbiTrail (SILVA et al., 2009), o modelo proposto atua no sentido de gerenciar os dados gerados pelas entidades que estão sendo acompanhadas e na disponibilização e aplicação para sistemas de apoio à Educação. Na versão inicial do modelo (CAMBRUZZI; RIGO; BARBOSA, 2012) foi implementado um protótipo onde múltiplas aplicações podem compartilhar dados com finalidades específicas e dois estudos de caso foram realizados. O primeiro é contextualizado no nível de Ensino Médio, voltado para acompanhamento individualizado de atividades e visualização de dados. O segundo deu-se no contexto do Ensino Superior, voltado para a análise e visualização de interações e relacionamento de dados. Em (CAMBRUZZI et al., 2012) o modelo foi ampliado para permitir a realização do terceiro estudo de caso em que as trilhas foram utilizadas no registro de informações de um grupo de estudantes da modalidade EAD no Ensino Superior com o objetivo de prever os alunos com tendência à evasão ainda durante o andamento da disciplina.

A presente proposta está sendo desenvolvida em um contexto de um projeto de pesquisa entre a empresa GVDASA Informática LTDA e o Programa Interdisciplinar de Pós-graduação em Computação Aplicada (PIPICA) e com recursos do CNPQ no programa RHAE (Edital MCT/CNPq n.75/2010 RHAE Pesquisador na Empresa, período 2012 - 2014). O objetivo principal da GVDASA é desenvolver um produto que utilize descoberta de conhecimento em bases de dados educacionais para a redução de evasão. Além disso existe um convênio firmado com a Unisinos para interação com o setor pedagógico responsável pelos cursos em EaD, o que proporciona a execução de estudos de caso nesse contexto.

## 1.1 Questão de Pesquisa

A questão de pesquisa que deverá ser respondida por este trabalho é a seguinte: *"Como deve ser um modelo computacional para apoio a EAD que considere aspectos de Learning Analytics, tais como visualização e mineração de dados educacionais que permita aos educadores atuarem de forma pró-ativa na redução da evasão?"*

## 1.2 Objetivos

Considerando o contexto e a motivação apresentados, este trabalho tem como objetivo principal desenvolver um modelo computacional que considere aspectos de LA com o objetivo de reduzir a evasão na EAD .

Como objetivos específicos podem ser relacionados:

- Desenvolver o modelo respeitando algum modelo de referência para aplicações de LA.
- Avaliar a utilização de técnicas de MDE para predição de alunos com risco de evasão durante o andamento de uma disciplina.
- Desenvolver uma arquitetura capaz de suportar o conceito de múltiplas trilhas aplicada a educação.
- Avaliar o uso potencial da organização de dados educacionais em múltiplas trilhas.

## 1.3 Método de Trabalho

Para atingir os objetivos elencados e responder a questão de pesquisa foi adotada a seguinte metodologia:

- Revisão bibliográfica das áreas.
- Estudo de aplicações relacionadas.
- Estudo e entrevistas com profissionais da educação para avaliar as necessidades de atuação para prevenção e reversão de evasão.
- Estudo das técnicas de visualização e mineração de dados.
- Definição da arquitetura.
- Implementação.
- Delimitação de estudos de caso.
- Realização dos estudos de caso.
- Avaliação.

Com esse estudo pretende-se demonstrar o quanto a arquitetura proposta favorece a flexibilidade na coleta, tratamento de dados e aplicações técnicas de LA para a redução da evasão na EAD.

## 1.4 Organização do Texto

Este trabalho apresenta no capítulo 2 os conceitos educação a distância, *Learning Analytics*, mineração de dados e trilhas. Trabalhos relacionados com o tema escolhido estão descritos no capítulo 3. No capítulo 4 é detalhado o modelo proposto com a explicação do seu funcionamento. A implementação do modelo está descrita no capítulo 5. A avaliação, capítulo 6 apresenta os experimentos realizados e resultados obtidos. Por fim o capítulo 7 aborda as considerações finais, contribuições e trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta considerações acerca de áreas que fundamentam o trabalho proposto. Seus aspectos são apresentados em linhas gerais, sem a intenção de esgotamento dos assuntos em profundidade. Entretanto são destacados os elementos que proporcionam a base para a elaboração desta proposta.

### 2.1 Educação a Distância

A Educação a Distância (EaD) é uma modalidade de ensino onde professores e alunos podem estar afastados geograficamente e temporalmente. A mediação é realizada com recursos didáticos disponibilizados em diversos meios de comunicação. Segundo (KEEGAN, 1996) o que diferencia a EaD de um estudo autodidata é a possibilidade de comunicação de via dupla, onde alunos e professores podem interagir e dialogar por meios eletrônicos e em encontros ocasionais.

Essa modalidade tem crescido muito no Brasil e somente no ano de 2012 houve um aumento de 52,5% das matrículas na modalidade EaD, somando 5.772.466 de matrículas (CEN-SOEAD.BR, 2012).

Conforme citado por (KAMPFF, 2009) o êxito dos cursos na EAD depende de sistemas eficazes de acompanhamento e avaliação, pois os professores poderão identificar dificuldades específicas dos alunos e oferecer o suporte adequado. Segundo (MOORE; KEARSLEY, 2007) o monitoramento eficaz exige uma rede de indicadores que disponibilizem os dados sobre o desempenho do aluno e do professor e que isso precisa ser feito frequente e rotineiramente.

Os altos índices de evasão apresentados na EAD são uma preocupação presente tanto no Brasil (CENSOEAD.BR, 2010), como em outros países (SCOTT, 2011; LEVY, 2007). No estudo apresentado por Carr (CARR, 2000) o índice de evasão girava em torno de 50% nos Estados Unidos. Em 2004 no estudo apresentado por (MAIA; MEIRELLES; PELA, 2004), onde foram analisadas 22 instituições brasileiras e 51 mil alunos, os índices de evasão ficavam em torno de 30%. O trabalho de (LONGO, 2009) apresenta índices internacionais com valores acima de 65% de evasão.

A EAD possui diversos modelos de implementação, sendo que os primeiros a serem conhecidos estão ligados com o uso de material impresso entregue através do correio, ou então os recursos de comunicação por rádio e, depois, por TV. Com a melhoria dos mecanismos de comunicação passam a ser exploradas também as possibilidades de cursos nesta modalidade com apoio de redes de transmissão de vídeo e áudio através de satélites de comunicação e ainda o suporte proporcionado pela Internet, tanto para transmissão de materiais, em especial os materiais em vídeo, como também para a implementação de ferramentas de interação (MOORE, 1993). Os diversos sistemas que integram as possibilidades de interação através da Internet, voltados para o uso em EAD, são também conhecidos como Ambientes Virtuais de Aprendizagem. Estes

ambientes permitem a realização de atividades de organização do ensino, com a integração de dados acadêmicos tais como a composição de alunos e professores em uma turma. Além disso, são oportunizados nestes ambientes a organização de materiais instrucionais e ferramentas diversas de interação, tais como ferramentas de interação síncrona, como o chat, e as ferramentas de interação assíncronas, como o fórum de discussão. Estes ambientes são responsáveis pela geração de uma grande quantidade de dados, tais como os dados de acesso ao ambiente e os dados de resultados de atividades, além dos dados de interação do tipo textual. Esta grande quantidade de dados pode servir como apoio em diversos tipos de análise quanto ao andamento e ao processo de ensino-aprendizagem.

O termo evasão escolar permite diversas interpretações e é utilizado em diferentes contextos com significados ligeiramente distintos. Também são diversificadas as iniciativas para a definição e aplicação de índices que permitem realizar comparações e avaliações a seu respeito. Em alguns casos considera-se como evasão a desistência do curso pelo estudante, independentemente da quantidade de participações efetuadas (FAVERO; FRANCO, 2006); em outras situações diferencia-se evasão de acordo com períodos médios para conclusão de curso e períodos anuais (SILVA FILHO et al., 2007); bem como identificam-se como evasão as situações de desistência definitiva após determinado contato com o curso.

A utilização de índices para medição de evasão (ADACHI, 2009), pode variar entre a medição de desistência em uma determinada disciplina de um curso, ou a desistência de um curso de graduação. Também são considerados em alguns casos situações de alunos que desistem de um curso, mas ingressam em outro curso dentro de uma mesma instituição, o que se diferencia de situações em que o aluno se desliga completamente de uma determinada instituição. São elencadas, portanto, a evasão no âmbito da disciplina, do curso, da instituição e do próprio sistema de Ensino.

A busca de referenciais teóricos que expliquem as questões relacionadas a evasão pode ser identificada em trabalhos diversos (TINTO, 1975; ANDRIOLA; ANDRIOLA; MOURA, 2006) nos quais é destacada a natureza complexa de relacionamentos envolvidos, tais como características pessoais, expectativas e eventos motivacionais. Estes modelos envolvem a condição social do aluno, atributos como gênero, idade, habilidades pessoais, experiências escolares anteriores, juntamente com suas expectativas de desenvolvimento pessoal e de carreira, associadas com a motivação para o desempenho acadêmico e o seu reconhecimento. Estas características são consideradas dentro de um contínuo temporal, de modo que a importância e influência de cada uma delas muda de acordo com o tempo no ambiente universitário.

Alguns destes estudos relacionam também aspectos mais específicos, ligados às características pessoais e institucionais, destacando-se fatores que se farão presentes ao longo de todo o período do curso, tais como questões ligadas à escolha do curso, divulgação de características da metodologia de ensino, adequação de atividades, atendimento e materiais instrucionais, bem como a atuação docente e o apoio institucional no atendimento à diversidade dos alunos e suas necessidades.

Como objetivo associado aos estudos desenvolvidos, nos mais diversos âmbitos, percebe-se a iniciativa das instituições de ensino, em geral, na promoção de ações para identificar variáveis associadas com o comportamento de evasão. Estas informações são posteriormente utilizadas em ações preventivas, de modo a minimizar o seu efeito. Em alguns casos esta identificação pode ser feita com informações de pouca variação e compostas pelo histórico social, motivacional e educacional do aluno. Já em outras situações as informações são bastante dinâmicas, como observa-se em relação à competências docentes ou interação e colaboração discente ao longo de um semestre letivo.

Os recursos de mineração de dados possibilitam, desde que sejam utilizados de modo adequado, o tratamento de ambas as situações acima, para geração de modelos que permitem ações de diagnóstico precoce e encaminhamento de ações preventivas, complementando desta forma o trabalho realizado em estudos periódicos de evasão escolar.

## 2.2 Learning Analytics

Mesmo possuindo uma grande quantidade de dados sobre os estudantes, as instituições de ensino superior tem sido tradicionalmente ineficientes no uso desses dados, muitas vezes realizando análises com atrasos substanciais, retardando ações e perdendo oportunidades de intervenções (SIEMENS; LONG, 2011). Isso se deve ao fato que o crescimento contínuo na quantidade de dados cria um ambiente em que novas abordagens se tornam necessárias para entender os padrões de valor que existem dentro dos dados.

Este contexto torna-se mais evidente ao ser constatado que a utilização de recursos de mediação digital e a melhoria dos sistemas de suporte, tais como os sistemas acadêmicos, repositórios digitais e ambientes virtuais de aprendizagem, é um fato que participa do cotidiano dos cursos de graduação em nível universitário, além de ser observado em menor escala em outros níveis de ensino. Soma-se a este fato a larga e crescente adoção dos dispositivos móveis (*tablets* e *smartphones*) na educação com a disponibilização de conteúdos de forma mais ampla e aberta, além da incorporação gradual de tecnologias de visualização, interação e imersão, tais como a realidade aumentada, ambientes de imersão e ambientes de interação multimodal, integrando gestos e voz ao processo de mediação digital. Este contexto define implicações positivas para a educação em geral, sendo enfaticamente destacado em diversos estudos na área, como o Horizon Report (New Media Consortium) (JOHNSON et al., 2013) e o relatório do Office of Educational Technology (US Department of Education), (BIENKOWSKI; FENG; MEANS, 2012). Entretanto este contexto traz à tona o problema de manuseio adequado e produtivo destes dados, disponíveis em grande quantidade e diversos formatos.

Esse contexto tem promovido um crescente interesse na análise automática dos dados que recentemente foi referenciada como *Learning Analytics* (LA) (CHATTI et al., 2012; GRELLER; DRACHSLER, 2012). Vale ressaltar que *Learning Analytics* não é uma nova área de pesquisa, mas ela pode ser considerada uma síntese de técnicas existentes em diversas áreas de

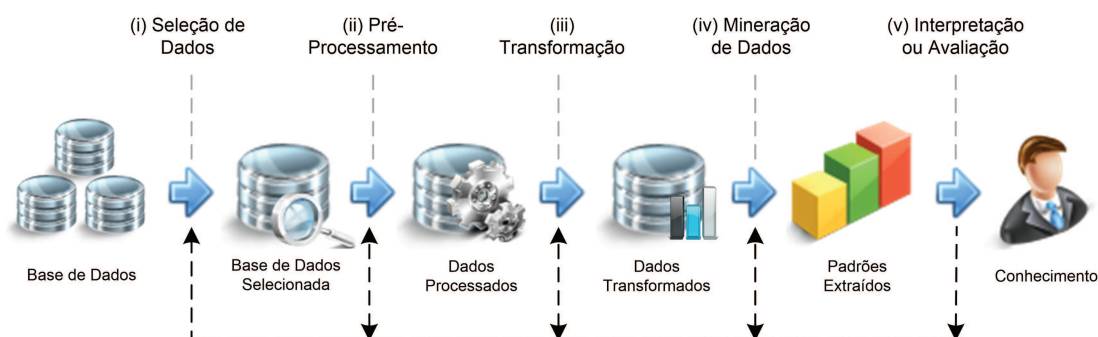
pesquisa convergentes com o uso da tecnologia para melhoria da aprendizagem, ou, em inglês *Technology Enhances Learning* (TEL). A relação entre LA e as área de pesquisa correlatas foram referenciadas no trabalho de CHATTI et al. (2012) e incluem *Learning Analytics, action research*, mineração de dados educacionais (EDM) do inglês *Educational Data Mining*, sistemas de recomendação e aprendizagem personalizada e adaptativa. Desta forma, as iniciativas de LA pode envolver combinações de recursos disponibilizados a partir de diversas outras áreas, tais como o Aprendizado de Máquina, a Inteligência Artificial, o resgate de informações, recursos de Estatística ou de Visualização de Dados, entre outros.

Em Chatti et al. (2012) os autores sugerem um modelo de referência para LA baseado em quatro dimensões com o objetivo de identificar tudo que é necessário para construir uma análise em LA. No trabalho de Greller e Drachsler (2012) são exploradas as dimensões chaves do LA, as questões que devem ser consideradas para minimizar perigos em potencial para uma exploração benéfica de dados educacionais. Eles propõem um framework que serve como guia para criação de serviços de LA de apoio à pratica educacional e de orientação ao aluno, na garantia de qualidade do aprendizado, desenvolvimento do currículo e melhoria da eficiência e eficácia do professor.

## 2.3 Mineração de Dados

A mineração de dados (MD), também conhecida como descoberta de conhecimento de bases de dados, é uma técnica computacional que vem sendo usada para identificar padrões novos e relevantes a partir de bases de dados. Para FAYYAD et al. (1996) trata-se de um processo não trivial, interativo e iterativo, composto por várias etapas, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados. Pode ser dividido em cinco etapas conforme Figura 1.

**Figura 1:** Processo de descoberta de conhecimento. Adaptado de Fayyad



Fonte: Adaptador de (FAYYAD et al., 1996)

A etapa (i), de Seleção dos dados, tem como objetivo a seleção de uma base de dados observando sua natureza. A etapa (ii), de Pré-Processamento, é considerada a etapa mais custosa do processo e tem como objetivo a limpeza dos dados e a eliminação de informações desneces-



sárias para o conhecimento que se deseja extrair. A etapa (iii), de Transformação, apresenta como objetivo a formatação dos dados para atender aos requisitos da técnica se será aplicada na próxima etapa. Na etapa (iv), de Mineração de Dados, são aplicadas as técnicas de mineração voltadas ao aprendizado de máquina. Por fim na etapa (v), de Interpretação ou Avaliação, os resultados do processo são avaliadas através de métricas calculadas. Ao final de cada etapa, os resultados devem ser avaliados e caso necessário, deve-se realizar alterações e iniciar um novo ciclo do processo a partir do ponto onde foram realizadas melhorias.

Destaca-se que este grande volume de dados educacionais tem estimulado o interesse na sua utilização, junto com técnicas de MD, na busca de respostas para perguntas específicas da educação, relacionadas com processos de aprendizagem, desenvolvimento de materiais instrucionais, acompanhamento e previsões, entre outros (BAKER; ISOTANI; CARVALHO, 2011; MANHÃES et al., 2011), a partir da obtenção de informações e padrões de comportamento importantes para apoiar determinadas práticas pedagógicas.

### 2.3.1 Mineração de Dados Educacionais

A Mineração de Dados Educacionais trata da aplicação de técnicas da Mineração de Dados junto aos novos conjuntos de dados obtidos nos diversos contextos educacionais. A natureza destes dados é mais diversa do que a observada nos dados tradicionalmente utilizados, demandando adaptações e novas técnicas. Ao mesmo tempo, esta diversidade nos dados representa um potencial de implementação de recursos fundamentais para auxílio na melhoria da Educação, (ROMERO; VENTURA, 2010). Alguns exemplos de aplicação destas técnicas são a geração de alertas (KAMPFF, 2009), o apoio à sistemas de recomendação (DURAND; LAPLANTE; KOP, 2011; TOSCHER; JAHRER, 2010) ou a captura de perfis (LI; MATSUDA, 2011). KOTSIANTIS (2012) se propõe a prever a nota do estudante em uma disciplina ministrada à distância, utilizando para isso seis técnicas de regressão existentes e comparando seus resultados. Já o trabalho de OLADOKUN; ADEBANJO; CHARLES-OWABA (2008) apresenta o problema da seleção de alunos para a universidade da Nigéria. Segundo os autores, uma abordagem para prever o desempenho de um estudante pode se dar a partir de dados histórico e suas performances associadas, sendo que com esse objetivo o estudo de caso apresenta a aplicação de redes neurais artificiais para prever o desempenho do estudante antes de admiti-lo.

Esta é uma área de pesquisa em expansão, tendo como principais enfoques os trabalhos relacionados com predição, agrupamento, mineração de relações, descoberta com modelos e tratamento de dados para apoio à decisão. Em todos estes casos, em maior ou menor grau, podem ser vislumbradas aplicações ligadas às questões discutidas no âmbito dos modelos teóricos sobre evasão escolar e ligadas aos processos de aquisições de modelos gerais e de antecipação de diagnósticos. Questões ligadas ao histórico educacional e aspectos sociais, que se apresentam com características estáticas e constituem dados históricos podem ser tratadas para a obtenção de informações que atendem às iniciativas institucionais de prevenção com ações amplas e edu-



cativas. Já dados originados na interação e desenvolvimento das atividades semestrais dentro de uma disciplina podem ser utilizados para a geração de diagnósticos e indicação de ações mais imediatas, dentro de um escopo menor de tempo, mas igualmente importantes, tratando de situações de possível evasão imediata.

O estudo do problema de evasão escolar possibilita identificar a sua relação com uma demanda importante na sociedade, onde tanto universidades públicas como universidades particulares apresentam índices considerados altos. Os seus efeitos estão relacionados com questões financeiras e com a diminuição do número de alunos formados no ensino superior, portanto com impactos para toda a cadeia produtiva do país. As situações associadas com origens e motivadores para ações de evasão escolar são bastante diversas e envolvem fatores externos às instituições, fatores ligados à aspectos pessoais dos alunos e fatores dependentes de políticas e metodologias adotadas institucionalmente. Alguns destes fatores podem ser identificados com apoio de técnicas de Mineração de Dados Educacionais, de modo a garantir facilidades para ações de prevenção e de atendimento.

Entretanto, identifica-se a necessidade de uma abordagem que integre de forma mais ampla o estudo prévio dos fatores a serem monitorados por técnicas de Mineração de Dados Educacionais, de modo a compor cenários coerentes com os conhecimentos acerca do processo de evasão escolar. A escolha dos atributos a serem utilizados para processos de mineração pode ser tanto mais efetiva quanto maior for o envolvimento de equipe multidisciplinar, que garanta a melhor utilização do conhecimento acerca deste fenômeno. Embora a mineração de dados tenha o potencial de identificar padrões relevantes, é muito difícil o processo de seleção das regras mais interessantes para o usuário que não é especialista em Educação e conhecedor profundo dos aspectos motivadores de evasão (ROMERO; VENTURA, 2010).

Quando os resultados são adequadamente gerados, entra em questão a utilização destas informações, que idealmente devem estar associadas com ações afirmativas institucionais. Um dos exemplos é o uso destas informações para a geração de alertas a serem utilizados pelos professores. Conforme (KAMPPFF, 2009) a integração das áreas de Mineração de Dados Educacionais e sistemas de alertas permite que o professor deixe de ser reativo, respondendo ao aluno apenas quando solicitado, e passe a ser proativo, ampliando as possibilidades de aprendizagem do aluno.

Aspectos ainda sendo investigados, como a integração e relacionamento de dados, podem ser tratados com recursos semânticos integrados às técnicas de mineração de dados (LIU, 2010). Por sua vez, os conjuntos textuais disponíveis permitem a identificação de estados emocionais dos estudantes e podem ser tratados com abordagens linguísticas de apoio à mineração (MALIN et al., 2009), ou então, dados textuais em redes sociais também podem ser utilizados a partir de técnicas de mineração (MACHADO et al., 2010).

## 2.4 Trilhas

Além do aumento no volume de dados disponíveis, constata-se um aumento e uma diversidade nos dispositivos que originam os dados. Isso torna-se evidente com a crescente utilização de dispositivos de computação móvel, que em boa parte propiciam também a inclusão de recursos capazes de gerar informações complementares tais como a informação de localização. Estes dispositivos permitem que informações diferenciadas sejam usadas para geração de contextos. Segundo DEY (2001), contexto é qualquer informação que possa ser utilizada para caracterizar a situação de entidades (pessoa, lugar ou objeto) que sejam consideradas relevantes para interação entre um usuário e uma aplicação.

O registro histórico dos contextos visitados e da atuação da entidade normalmente recebe a denominação de Trilha (DRIVER; CLARKE, 2008; LEVENE; PETERSON, 2002). Ou seja, o conceito de trilhas é descrito a partir destas referências como uma abordagem que visa proporcionar melhores condições de tratamento de dados que possuem uma certa diversidade e que podem ser relacionados com dados gerados por outros dispositivos ou sistemas e que podem também estar caracterizados pela geração em diferentes períodos de tempo. Além desta visão, observa-se em outros trabalhos (SCHOONENBOOM et al., 2007; TATTERSALL et al., 2006) o conceito de trilhas com o mesmo aspecto geral, mas associado de forma mais efetiva com os aspectos de aprendizagem de alunos, tendo em vista o contexto diversificado quanto aos materiais, formatos e disponibilidade destes. A identificação de trilhas pode ser então associada com o acompanhamento de ações diversas realizadas por um estudante, ou com um conjunto específico de materiais e recursos que foi utilizado ou acessado, ou ainda, por um conjunto de assuntos mais gerais que foi identificado a partir de ações e materiais.

Com o referido crescimento e transformação do conjunto de recursos e com as opções encontradas nos ambientes de aprendizagem, o tratamento de trilhas pode ser avaliado como uma opção para a identificação também de aspectos mais gerais, tais como os aspectos cognitivos e pedagógicos associados com o comportamento descrito nos registros processados.

Devido às possibilidades geradas pela utilização destas informações, diversos trabalhos estão sendo pesquisados a partir de um objetivo geral ligado à aquisição e utilização dos dados referentes às trilhas de entidades. Porém, estes trabalhos apresentam focos distintos, tais como: o auxílio à navegação e acesso a conteúdos (GAMS; REICH, 2002; WANG et al., 2008), o registro de experiências de aprendizagem (LEVENE; PETERSON, 2002; SILVA et al., 2009), a anotação de localizações e atividades (SMITH, 2008; DRIVER; CLARKE, 2008). Entretanto vários destes dados armazenados em trilhas distintas podem ser úteis de forma integrada para atender as demandas de análises com finalidades distintas. Por exemplo, na área educacional existem diversas demandas de análise de dados gerados pelos alunos, tais como a relação entre frequência e resultados de avaliação. Portanto seria vantajoso para a área educacional tratar esses dados em um contexto unificado.



### 3 TRABALHOS RELACIONADOS

Neste capítulo são descritos dois trabalhos que propõem modelos de referência para LA e trabalhos de representação de dados educacionais em trilhas. O objetivo foi identificar nos modelos de referência de LA alguns aspectos gerais que representam necessidades a serem supridas. Já os trabalhos sobre trilhas são analisados com o objetivo de justificar esta abordagem para a representação dos dados educacionais dos estudantes. Essas duas áreas foram utilizadas para a definição dos aspectos principais da arquitetura aqui proposta.

#### 3.1 Modelos de referência de Learning Analytics

Os trabalhos de GRELLER; DRACHSLER (2012) e CHATTI et al. (2012) apresentam modelos de referência para LA. Como esta área é recente considera-se que a análise dos dois trabalhos elencados reflete de forma consistente as iniciativas na área. A seguir são tecidos comentários descrevendo e detalhando estes modelos estudados.

##### 3.1.1 Trabalho de GRELLER; DRACHSLER (2012)

No trabalho de GRELLER; DRACHSLER (2012) os autores exploram as dimensões chave da LA e alguns perigos potenciais para exploração benéfica dos dados na educação. Apontam que o crescimento das pesquisas na área de LA é justificado principalmente devido a aplicações de LA conseguirem abranger todo o universo de informações e apresentarem um resultado real e ininterrupto sobre o comportamento dos estudantes.

Alertam também que, apesar do grande entusiasmo, as questões de privacidade e ética dever ser consideradas. Ainda sobre essa questão, sugerem que precisa acontecer um trabalho substancial nas universidades e poder público na definição de uma legislação em um novo guia de ética para reduzir os riscos envolvidos na manipulação desses dados e para proteger o sujeito dos dados, normalmente os estudantes. Relatam também as novas competências que são necessárias aos educadores para interpretar e realizar ações sobre as reflexões que as aplicações de LA permitem.

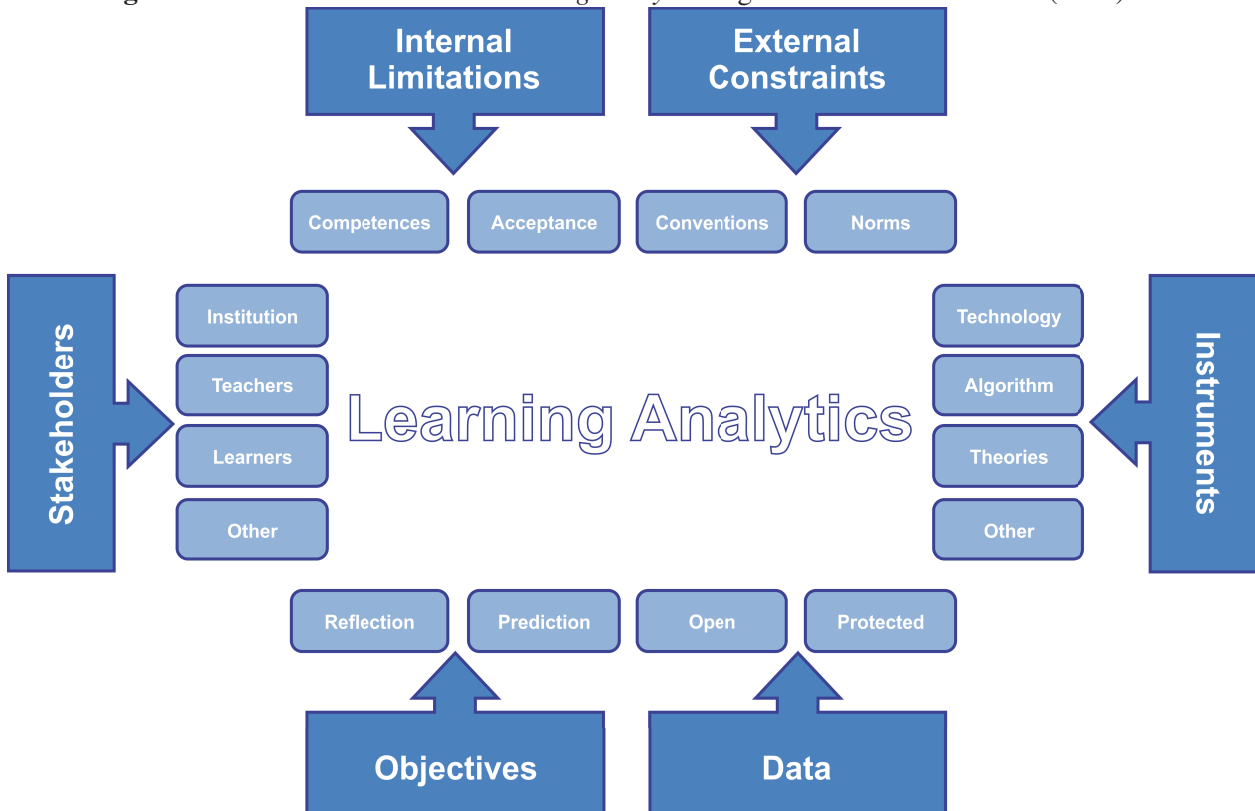
Apresentam como perigo potencial o fato que a LA pode não ser usada somente com o objetivo de fornecer informações aos estudantes sobre suas atuais necessidades e ajudar as instituições a terem uma experiência educacional altamente personalizada, mas sim que isso possa servir como forma de manipulação dos estudantes, trabalhadores, cidadãos pelos governos ou instituições.

No trabalho é proposto um *framework* para LA, apresentado na Figura 2, onde são consideradas seis dimensões críticas: Interessados (do inglês *Stakeholders*), Objetivos (do inglês *Objective*), Dados (do inglês *Data*, Instrumentos (ou do inglês *Instruments*), Limitações Externas (do inglês *External limitations*) e Limitações Internas (do inglês *Internal limitation*). Estas

dimensões são descritas e comentadas a seguir.

- Interessados: Inclui os sujeitos dos dados e os clientes dos dados, que podem ser os estudantes e educadores respectivamente.
- Objetivos: O objetivo pode ser reflexão ou predição. A reflexão ocorre quando os dados são utilizados para entender comportamentos e reestruturar a estrutura do curso para novos estudantes. A predição ocorre quando sobre o passado é criado um modelo preditivo sobre os estudantes atuais.
- Dados: Essa dimensão se refere a origem dos dados, que pode ser o ambiente de aprendizado e outras fontes internas e externas da instituição.
- Instrumentos: Podem se referir às tecnologias que podem ser empregadas nas aplicações de LA, como MDE, aprendizado de máquina, técnicas clássicas de estatística, análise das interações sociais e outras. Além disso, também podem se referir a teorias pedagógicas.
- Limitações Externas: Envolve as questões de privacidade e ética.
- Limitações Internas: Refere-se às competências requeridas para a interpretação dos dados e realização de ações fundamentadas nas análises obtidas.

**Figura 2:** Dimensões críticas da *Learning Analytics* segundo Greller e Drachsler (2012)



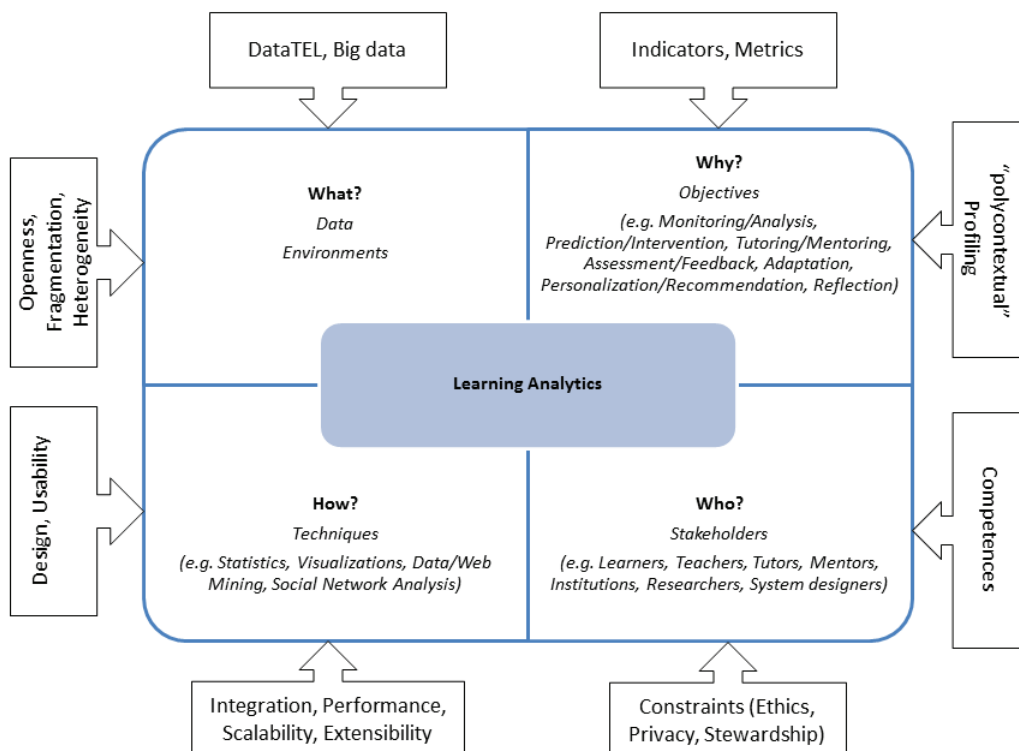
O objetivo do *framework* é demonstrar as conexões inerentes entre as seis diferentes dimensões e o impacto dos processos analíticos no usuário final e nos dados. Qualquer mudança em uma das dimensões pode alterar o resultado final e portanto a consideração a respeito destas seis dimensões pode levar a uma ótima exploração da LA.

### 3.1.2 Trabalho de CHATTI et al. (2012)

No trabalho de CHATTI et al. (2012) os autores exploram as conexões existentes entre LA e as áreas relacionadas e descrevem um modelo de referência de LA, representado na Figura 3, que é baseado em quatro dimensões: "O que?", "Quem?", "Porque?" e "Como?".

- "O que?" (do inglês *What?*) - responsável pela definição do tipo de dados que deve ser coletado, gerenciado e utilizado nas análises.
- "Quem?" (do inglês *Who?*) - responsável pela definição dos interessados nas análises.
- "Porque?" (do inglês *Why?*) - responsável pela definição dos objetivos das análises.
- "Como?" (do inglês *How?*) - responsável pela definição das técnicas que são utilizadas para realizar as análises dos dados coletados.

**Figura 3:** Modelo de Referência de *Learning Analytics* segundo Chatti et al (2012)



Os autores também apresentaram uma revisão bibliográfica das publicações recentes na área de LA e mapearam os trabalhos nessas quatro dimensões. Nessa revisão identificaram que: (1) a maioria dos dados utilizados em ambientes de LA vem de ambientes de aprendizagem baseados na web; (2) que a maioria dos ambientes de LA são orientadas para tutores inteligentes ou pesquisa; (3) que o objetivo mais comum é adaptação e monitorament; (4) a técnica mais frequentemente utilizada é classificação e predição. Os autores concluem dizendo que acreditam que no futuro este padrão vai mudar e que o seu foco tende a ser mais aberto.

### **3.2 Organização de Dados em Trilhas**

Essa seção apresenta trabalhos relacionados com foco na representação e utilização de trilhas para aplicações diversas.

O projeto TrailTRECer (GAMS; REICH, 2002) é voltado ao auxílio na navegação e na busca de conteúdo, possibilitando que um usuário acompanhe a trilha de navegação de outro usuário, desta forma facilitando a busca de informações. O conceito de trilha adotado pelo projeto define esta como um conjunto de documentos digitais que um usuário acessou ao executar uma determinada atividade.

No trabalho desenvolvido por (LEVENE; PETERSON, 2002), denominado projeto Trail Records, são utilizados dispositivos móveis para registro de experiências de aprendizagem usando diferentes mídias, tais como textos, vídeos e áudio, durante a visita em diferentes localizações. O conceito de trilha adotado pelo projeto define esta como a sequência dos registros vinculando conteúdo com localização, sendo utilizada para recompor a experiência do aprendiz no trajeto visitado.

Em Wang et al (WANG et al., 2008) o principal objetivo do trabalho é a disponibilização de conteúdo multimídia. O projeto considera uma trilha composta por um conjunto de registros que indicam os lugares onde cada dispositivo realizou acesso. Já nos trabalhos de Smith (SMITH, 2008; SMITH et al., 2008) observa-se que é proposta a anotação das localizações percorridas por usuários de dispositivos móveis, compondo uma trilha de locais visitados. Além disso, esse modelo prevê que sejam registradas todas as informações possíveis relacionadas com cada local. O objetivo é o registro das atividades de um usuário, permitindo consultas no futuro.

O projeto Hermes (LEVENE; PETERSON, 2002) é voltado para o gerenciamento de atividades e considera que a trilha é uma coleção de atividades programadas de forma contextual. Suas funcionalidades são direcionadas para a geração automática da ordem de execução das atividades levando em conta informações de contexto.

No projeto Startrack, descrito em (ANANTHANARAYANAN et al., 2009), é proposto um conjunto de operações para manipulação de trilhas. Seu conceito de trilha considera que esta representa uma rota contínua baseada em eventos discretos. Desta forma, cada trilha é uma coleção de lugares visitados ordenada no tempo. As trilhas são armazenadas em um servidor e o usuário proprietário da trilha pode compartilhar suas trilhas com outros usuários. Cada

aplicação pode definir quais informações vai guardar em cada trilha.

Silva (SILVA et al., 2009) propôs um modelo denominado UbiTrail, que possui como principal objetivo gerenciar informações sobre a trilha de uma entidade. O conceito de trilha adotado pelo projeto define que trilha é uma representação digital do histórico dos contextos visitados por uma entidade. Não apenas a localização é considerada, mas também o contexto, indicando o que ela está fazendo e o ambiente a seu redor. Todas as informações relacionadas com o deslocamento de uma entidade são organizadas em uma trilha única e contínua.

### 3.2.1 Comparação entre os trabalhos de Organização de Dados em Trilhas

Os trabalhos apresentados estão relacionados na Tabela 1. Embora os projetos estudados possuam suporte a trilhas, as características tratadas são diferentes, de acordo com o foco específico de cada um deles. Todos os trabalhos registram suas informações em uma única trilha. Os critérios de comparação analisados são os seguintes:

- Contexto: indica o registro de informações do contexto;
- Histórico: este quesito indica se o modelo avaliado possui armazenamento de informações históricas relacionadas;
- Níveis de Dados: indica a disponibilidade de suporte para o tratamento de diferentes níveis de detalhamento para os dados dos contextos;
- Múltiplas Trilhas: avalia se diferentes tipos de trilha podem ser armazenados e tratados.

**Tabela 1:** Comparativo entre os trabalhos relacionados

Projetos	Contexto	Histórico	Níveis de Dados	Múltiplas Trilhas
TrailTRECer	Sim	Sim	Não	Não
Trail Records	Sim	Sim	Não	Não
Nidaros	Sim	Sim	Não	Não
Life Annotation	Sim	Sim	Não	Não
Hermes	Sim	Não	Não	Não
Startrack	Não	Sim	Não	Não
UbiTrail	Sim	Sim	Não	Não

Fonte: Elaborado pelo autor.

Conforme evidenciado pela análise realizada, os conceitos de múltiplas trilhas e o conceito de níveis de granularidade de dados não são explorados nestes trabalhos.

### 3.3 Considerações sobre os trabalhos relacionados

Os modelos de referência estudados descrevem um cenário onde diversos elementos e dimensões estão envolvidos. Os resultados do processo de LA é dependente da quantidade de



elementos considerados, sendo que os dados que descrevem cada um destes elementos são atualmente originados em diversos sistemas de informação diferentes. Existe uma relação importante entre a capacidade dos sistemas de LA e a sua possibilidade de obtenção e tratamento de dados. Por sua vez, a abordagem de tratamento de trilhas e a abordagem sugerida e adotada neste trabalho, de múltiplas trilhas, colabora exatamente no sentido de proporcionar maior flexibilidade nas operações de coleta e tratamento destes diversos dados.

Considerando-se ainda que as fontes de dados podem apresentar comportamentos dinâmicos, como no caso de surgimento de novos recursos em áreas de interação e de sistemas de informação, então deve ser ampliada a consideração da importância de uma escolha de uma infraestrutura flexível para acomodar estes dados, o que é possível a partir da abordagem de múltiplas trilhas.

## 4 MODELO PROPOSTO

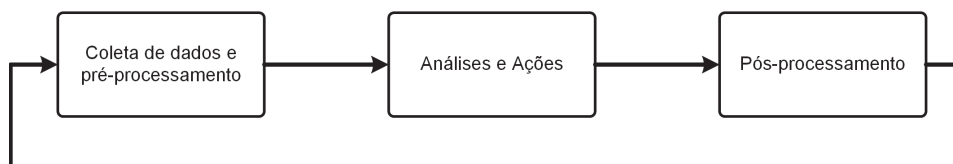
Neste capítulo é apresentado o modelo proposto, nessa descrição foi utilizada a anotação TAM/FMC (SAP AG, 2007). Inicialmente destacam-se alguns detalhes para os elementos básicos do modelo de LA e de gerenciamento de múltiplas trilhas. Em seguida é detalhada a arquitetura do modelo detalhando os componentes cada camada.

### 4.1 Elementos Básicos do Modelo de *Learning Analytics*

O processo de LA representado na Figura 4 foi definido por CHATTI et al. (2012) e está dividido em três grandes etapas:

- Coleta de dados e pré-processamento: Responsável pela coleta de dados nos diversos sistemas e ambientes educacionais e pela transformação dos dados em um formato que possa ser usado de entrada para as técnicas específicos de LA.
- Análises e Ações: Com base nos dados obtidos na etapa anterior, deve ser escolhida uma técnica de LA que permita descobrir padrões escondidos nos dados com o objetivo de proporcionar uma experiencia de aprendizado mais efetiva. Além das análises e visualizações das informações, essa etapa pode incluir ações de monitoramento, predição, intervenção, adaptação, personalização, recomendação e outras.
- Pós-Processamento: Essa etapa é responsável pela melhoria contínua no processo de LA, pois envolve a inclusão e exclusão dados e atributos, redefinição de indicadores e identificação de novos métodos de análises.

**Figura 4:** Processo de *Learning Analytics*



Fonte: Adaptado de CHATTI et al. (2012)

O modelo de LA proposto tem o objetivo de atender a etapa de coleta de dados e pré-processamento e a etapa de análises e ações. A etapa de pós-processamento não será tratada neste trabalho, ficando para uma extensão futura a sua automatização. Com a utilização desse modelo, essa etapa deve ser realizada de forma manual, onde devem ser analisados os resultados e realizadas as alterações necessárias para refletir as melhorias. Por fim, o modelo de gerenciamento de múltiplas trilhas está inserido na etapa de coleta de dados e pré-processamento.

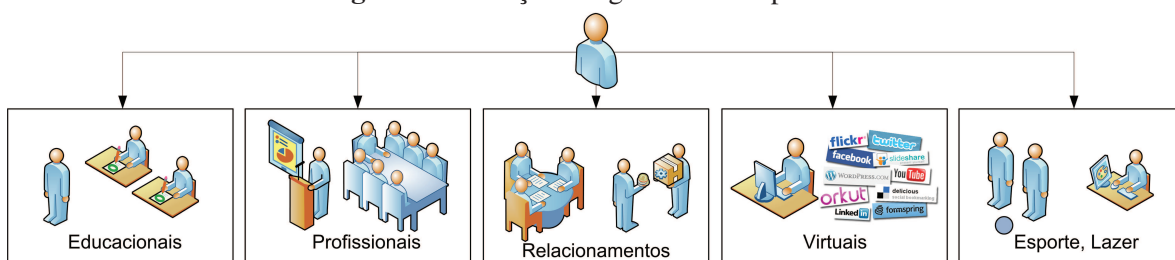
## 4.2 Elementos Básicos do Modelo de Tratamento de Múltiplas Trilhas

O termo entidade deve ser entendido como a representação de um objeto computacional ou de um usuário. Com isso, as trilhas registram as diferentes informações geradas por esta entidade. No modelo proposto cada entidade não é representada em uma única trilha, tal como observa-se em trabalhos relacionados (SILVA et al., 2009). Ao invés disso, cada entidade é representada em múltiplas trilhas, de acordo com as características das informações, de modo que possam ser agrupadas ou diferenciadas, para os casos em que ocorrem diferentes necessidades de análise e uso.

Sendo assim, este modelo proporciona duas vantagens que estão fortemente relacionadas com esta abordagem. A primeira é a flexibilidade para a aquisição dos dados, tendo em vista que diferentes fontes e diferentes aplicações podem ser integradas ao modelo como origem de dados de trilhas específicas. A segunda é a flexibilidade para utilização dos dados, já que aplicações com necessidades distintas podem ter apenas as trilhas de interesse consultadas. Um exemplo desta situação pode ser melhor entendido ao ser considerado o conjunto de registros de diversos aspectos da vida de uma pessoa, que poderiam envolver o seu histórico educacional, atuações profissionais, o registro das atividades realizadas em ambientes virtuais de aprendizagem, ou ainda, o registro de suas ações em comunidades virtuais de relacionamento.

Na Figura 5 é destacada a representação de uma entidade interagindo em diversas destas situações, sendo que cada uma delas pode estar representada em uma trilha, na qual serão armazenadas adequadamente as informações pertinentes. Essa abordagem permite flexibilidade na utilização destes diversos aspectos da vida desta pessoa, possibilitando que sejam levados em conta para representar o contexto de suas interações e atuações.

**Figura 5:** Interações originando múltiplas trilhas



Fonte: Elaborado pelo autor.

Para realizar o mapeamento e registro dos contextos que farão parte de cada trilha, são utilizados componentes de software específicos, sendo cada um deles responsável por registrar um conjunto de informações em uma trilha ou em partes dela. No exemplo da Figura 5, diferentes componentes de software registram o contexto da forma mais detalhada e adequada possível para cada situação.

Em geral, as aplicações que utilizam trilhas fazem uso de seus dados em consultas com base em uma dimensão temporal, em um espaço de tempo. No modelo proposto, é denominado

**Figura 6:** Múltiplas trilhas e contexto sistêmico

Fonte: Elaborado pelo autor.

contexto sistêmico a composição resultante do agrupamento de contextos de diversas trilhas (aquelas de interesse para a aplicação específica) em uma dimensão temporal. O esquema representado na Figura 6 ilustra o conceito de contexto sistêmico, onde pode-se visualizar diversas trilhas representadas de forma independente e relacionadas pela dimensão temporal. Na figura também estão exemplificadas trilhas para diversos aspectos relacionados com um estudante. Os exemplos destacam a trilha 'Educação formal' composta por registros de cursos de formação obrigatória e cursos de ensino superior ou pós-graduação. Em sequência é destacada a trilha 'Educação complementar', composta por cursos de extensão realizados. Por fim, a trilha 'Profissional' foi utilizada para destacar as diferentes atividades profissionais realizadas.

A flexibilidade de consulta aos dados mapeados nas diferentes trilhas é também estendida aos aspectos relacionados com a coleta de dados, que pode ser feita com base em diferentes

**Figura 7:** Exemplo de contexto sistêmico do primeiro semestre de 2009.

Contexto Sistêmico			
01/01/2009 e 30/06/2009			
Ensino Superior	Local: Universidade X Curso: Ciência da Computação	<disciplina> Estrutura de Dados Aprovado (8,0)	<disciplina> Sistemas Operacionais Aprovado (8,80)
Inglês	Local: Escola de Idiomas Nível: Avançado Carga Semanal: 4 horas Professor: Fulado de Tal	Conteúdos	Colegas
Programador II	Local: Empresa X Carga Semanal: 44 horas Salário: R\$ 2.000,00 Projetos: Projeto 1, Projeto 2, ... Atividades: Atividade 1 e Atividade 2, ...		Colegas de Trabalho

Fonte: Elaborado pelo autor.

dispositivos. Novas fontes e novos dispositivos podem ser incluídos e relacionados ao conjunto de trilhas existentes sem a necessidade de reorganização das aplicações que estão em uso.

**Figura 8:** Exemplo de contexto da disciplina Estrutura de Dados.

Contexto			
Modalidade EAD de 16/02/2009 a 26/06/2009			
Estrutura de Dados	Conteúdo	Histórico de Interação	Professor
	Trabalhos e Avaliações	Aspectos financeiros	Colegas
	Bibliografia retirada da biblioteca	Bibliografia Indicada	...

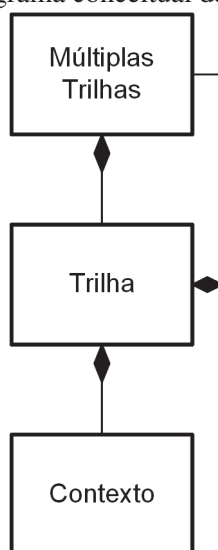
Fonte: Elaborado pelo autor.

Em particular, o detalhamento de cada trilha também é tratado neste modelo, tendo em vista que, muitas vezes, as aplicações interessadas em dados de uma mesma trilha podem apresentar necessidades diversas relacionadas com a granularidade dos dados consultados. Isso pode ser necessário em casos de aplicações que precisem verificar em detalhes a situação descrita em um contexto. A Figura 7 exibe um exemplo onde a trilha ‘Ensino Superior’ possui um contexto para a disciplina ‘Estrutura de Dados’, onde é exibida a nota final. Caso uma aplicação necessite, é possível o acesso a todos os detalhes dessa disciplina, tal como exibido na Figura 8, onde podem ser verificados os diversos elementos possíveis de recuperação e armazenamento em relação à esta disciplina.

No modelo conceitual de múltiplas trilhas, exibido na Figura 9, é possível verificar que as múltiplas trilhas de uma entidade são compostas de trilhas, e por sua vez, as trilhas são compostas de contextos. Uma trilha pode conter outro conjunto de múltiplas trilhas, garantindo assim que os dados possam ser organizados em diversos níveis.

Cada registro de contexto é composto pela seguinte estrutura:

- Identificação do Contexto: Código único identificador do contexto;
- Entidade: Entidade que o contexto pertence;
- Trilha: Trilha que o contexto pertence;
- Período: Período que o contexto é ocorreu;
- Descrição: Descrição do contexto;

**Figura 9:** Diagrama conceitual de Múltiplas Trilhas

Fonte: Elaborado pelo autor.

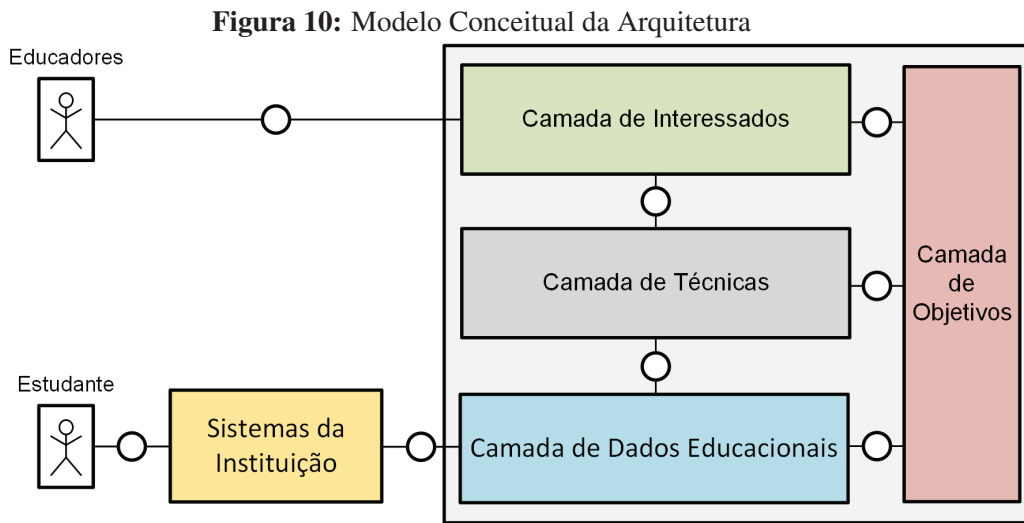
- Dados do Contexto: Dados que detalham o contexto;
- Informações do Registro: Definição do responsável pelo registro, data e hora do registro, origem do dado.

Deste modo, ilustra-se a flexibilidade para a montagem de visões diferenciadas do conjunto de dados armazenados de acordo com as necessidades de cada aplicação. Neste sentido, as Figuras 6, 7 e 8 apresentam a composição de quadros de informações que são obtidas em diversas fontes e com uso de diferentes componentes de software. Portanto, todas estas informações são pertinentes e contribuem para a composição de um contexto mais detalhado e completo para uma determinada entidade. Sendo assim, a flexibilidade de composição e coleta de dados é uma vantagem para a implementação das aplicações que irão utilizar estes dados. Essa flexibilidade está associada a mecanismos que facilitem a integração e interoperabilidade de cada trilha.

### 4.3 Arquitetura

A arquitetura do modelo contempla as quatro dimensões do Modelo de Referência de LA (Figura 3) proposto por CHATTI et al. (2012). Quanto ao *framework* para LA (Figura 2) proposto por (GRELLER; DRACHSLER, 2012), este contempla quatro das seis dimensões, que são: Interessados, Objetivos, Dados e Instrumentos. As dimensões Limitações Internas e Limitações Externas não foram tratadas na arquitetura por se tratarem de políticas, diretrizes e questões ligadas à competência da equipe envolvida.

Na Figura 10 é possível observar o modelo conceitual da arquitetura proposta. A Camada de Objetivos implementa dimensão 'Porque?'. A Camada de Dados Educacionais implementa a dimensão 'O que?'. A Camada de Técnicas implementa a dimensão 'Como?'. E por fim, a



Fonte: Elaborado pelo Autor.

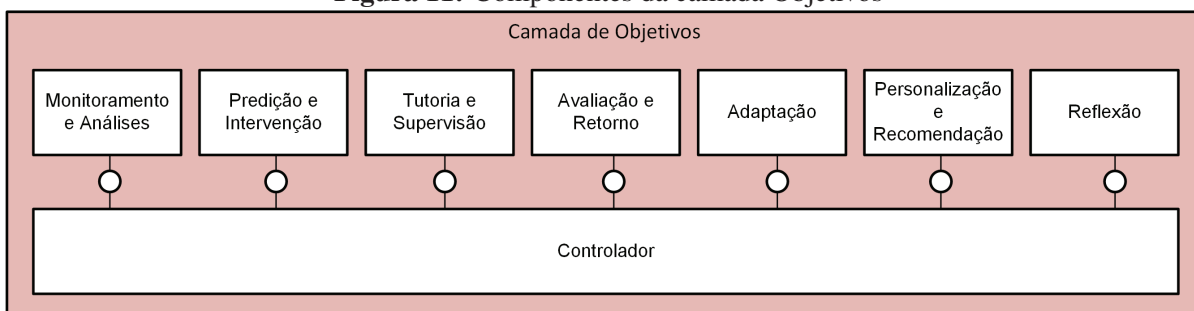
Camada de Interessados implementa a dimensão 'Quem?'.

#### 4.3.1 Camada de Objetivos

A Camada de Objetivos, detalhada na Figura 11, é responsável pelas aplicações que atendem aos objetivos da aplicação de LA. É composta pelos seguintes componentes: Controlador, Monitoramento e Análises, Predição e Intervenção, Tutoria e Supervisão, Avaliação e Retorno, Adaptação, Personalização e Recomendação, Reflexão. Os componentes Controlador, Monitoramento e Análises, Predição e Intervenção foram projetados e desenvolvidos neste trabalho, os demais foram inserido para expansão do modelo em trabalhos futuros.

- **Controlador:** é o ponto de entrada para os serviços oferecidos pela camada de objetivos. A sua operação inicia a partir de uma solicitação de serviço, que será delegada para o componente mais apropriado para seu atendimento.
- **Monitoramento e Análises:** esse componente implementa as funcionalidades que pos-

**Figura 11: Componentes da camada Objetivos**



Fonte: Elaborado pelo Autor.

suem o objetivo de mapear as atividades que os estudantes estão realizando e como eles estão utilizando o ambiente. Ou seja, permite visualizar seu comprometimento. O objetivo é auxiliar os educadores a identificarem padrões, melhorarem o ambiente e realizarem ações sobre os estudantes.

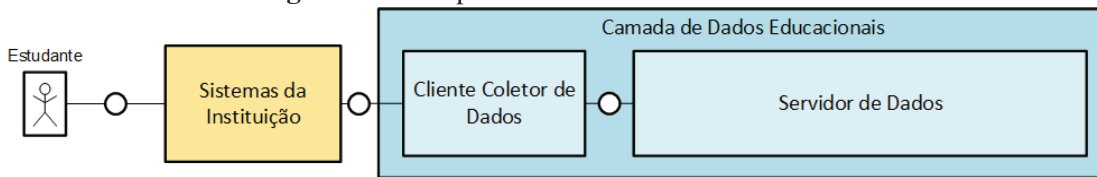
- **Predição e Intervenção:** esse componente é responsável pelas funcionalidades de predição do resultado futuro, baseado nas atividades correntes e no comprometimento dos estudantes. O objetivo é permitir uma abordagem pró-ativa voltada para os estudantes que necessitam de orientação.
- **Tutoria e Supervisão:** esse componente é responsável pelas funcionalidades que permitem auxiliar os estudantes em seu aprendizado através de ferramentas que permitam ao tutor acompanhar o progresso do aluno e a fornecer orientações sobre seu desempenho.
- **Avaliação e Retorno:** esse componente é responsável pelas funcionalidades que apoiam a avaliação ou auto-avaliação com o objetivo de melhorar a eficiência e eficácia do processo de aprendizagem.
- **Adaptação:** esse componente é responsável pelas funcionalidades de organização de forma adaptativa dos recursos de aprendizagem de acordo com as necessidades de cada estudante.
- **Personalização e Recomendação:** esse componente é responsável pelas funcionalidades de personalizar como são apresentados os recursos e objetivos conforme o perfil de cada estudante e suas preferências.
- **Reflexão:** esse componente é responsável pelas funcionalidades que permitem, tanto aos educadores, quanto aos estudantes, realizarem uma reflexão através de análise de dados no contexto da turma, disciplina, curso ou mesmo da instituição.

#### 4.3.2 Camada de Dados Educacionais

A Camada de Dados Educacionais, detalhada na Figura 12, é responsável pela extração e gestão de dados do modelo. Ela é baseada na arquitetura cliente/servidor. O funcionamento ocorre a partir da interação dos estudantes com a instituição. Essa interação pode ser direta, quando o estudante utiliza um sistema da instituição como o AVA, ou indireta, quando o estudante utiliza algum serviço da instituição e esse serviço registra informações através de um sistema, por exemplo o ERP. Sempre que ocorre uma interação são registrados dados nos bancos de dados próprios dos sistemas e esses dados são a base para o funcionamento do sistema de LA.

Os dados podem ser organizados diretamente no banco de dados ou utilizando múltiplas tabelas.



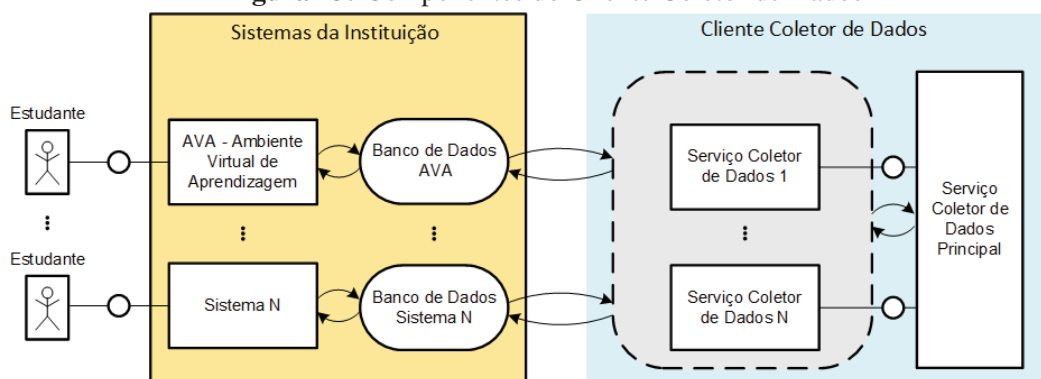
**Figura 12:** Componentes da camada Dados Educacionais

Fonte: Elaborado pelo Autor.

Ela é composta pelos seguintes componentes: Cliente Coletor de Dados e Servidor de Dados.

#### 4.3.2.1 Cliente Coletor de Dados

O cliente coletor de dados, detalhado na Figura 13, é responsável pela extração de dados dos bancos de dados dos sistemas da instituição. Ele possui dois tipos de componentes, o serviço coletor de dados principal e um serviço de dados para cada fonte de dados. Cada serviço coletor de dados pode optar por registrar na estrutura de múltiplas trilhas ou não. Caso optem por utilizar o modelo de múltiplas trilhas, devem enviar, juntamente com os dados obtidos que desejam armazenar, um conjunto de metadados necessário para registro na trilha.

**Figura 13:** Componentes do Cliente Coletor de Dados

Fonte: Elaborado pelo Autor.

O serviço coletor de dados principal gerencia a periodicidade de execução de todos os serviços coletores de dados instanciados no cliente. Para cada sistema deve haver pelo menos um serviço coletor de dados, que periodicamente efetua a coleta, limpeza e carga dos dados para o serviço gerenciador de dados que está no servidor de dados.

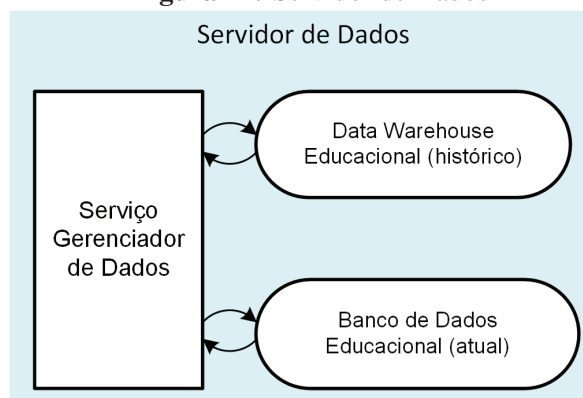
#### 4.3.2.2 Servidor de Dados

O servidor de dados, detalhado na Figura 14, é responsável pelo gerenciamento dos dados da solução. Ele possui três componentes: Serviço gerenciador de dados, data warehouse

Educacional e banco de dados educacional.

O serviço gerenciador de dados é responsável por atender as solicitações dos clientes coletores de dados e armazenar os dados em seus respectivos lugares. Quando os dados são históricos eles são armazenados em um data warehouse educacional e quando são dados atuais eles são armazenados no banco de dados educacional. O armazenamento dos dados pode ser efetuado em forma de múltipla trilhas, quando solicitado pelo cliente coletor.

**Figura 14:** Servidor de Dados

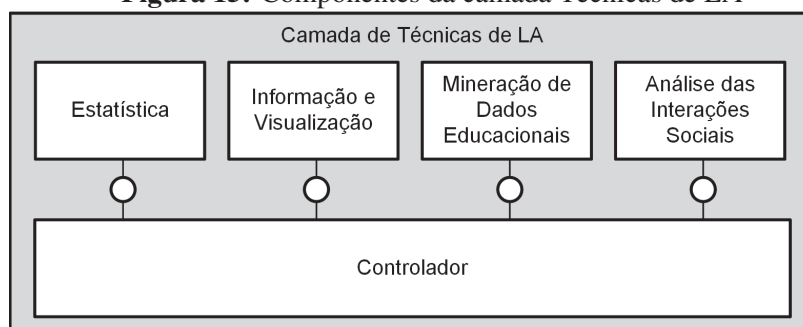


Fonte: Elaborado pelo Autor.

#### 4.3.3 Camada de Técnicas de LA

A Camada de Técnicas de LA, detalhada na Figura 15, é responsável pela implementação das técnicas que são utilizadas nas camadas de Objetivos e Interessados. É composta pelos seguintes componentes: Controlador, Estatística, Informação e Visualização, Mineração de Dados Educacionais e Análise das Interações Sociais.

**Figura 15:** Componentes da camada Técnicas de LA



Fonte: Elaborado pelo Autor.

A seguir são destacados alguns detalhes de seus componentes:

- **Controlador:** é o ponto de entrada para os serviços oferecidos pela camada de técnicas. O funcionamento dele inicia a partir de uma solicitação de serviço, que é delegada para o componente mais apropriado.

- **Estatística:** esse componente é responsável por implementar as técnicas de análise estatística sobre os dados educacionais e gerar indicadores para cada estudante, tais como o Tempo total on-line, o número total de visitas, o número de páginas visitadas, o número de postagens em fóruns, o número de postagens lidas em fóruns, percentual de material lido e outros possíveis.
- **Informação e Visualização:** esse componente é responsável por permitir que as informações estatísticas e outros dados sejam exibidos de maneira gráfica para facilitar a análise e interpretação. Podem ser utilizados gráficos, mapas, representações em 3D, grafos para que a informação seja apresentada de forma clara e objetiva.
- **Mineração de Dados Educacionais:** esse componente é responsável por implementar as técnicas de mineração de dados educacionais. As duas principais etapas são: Descoberta de padrões e classificação. No serviço de descoberta de padrão são implementadas as técnicas de árvore de decisão e redes neurais artificiais, ambas com aprendizagem supervisionada. Em ambas as técnicas os dados históricos que estão no Data Warehouse Educacional são utilizados para a etapa de treinamento. Ao final do processo de descoberta de padrões obtém-se um modelo preditivo. Já no serviço de classificação, é dado um conjunto de dados e um modelo preditivo e o serviço é responsável por classificar os dados de entrada conforme o modelo. O resultado gerado é a lista de dados da entrada classificada.
- **Análise das Interações Sociais:** esse componente é responsável por implementar as técnicas de análise das interações sociais que permitem aos educadores visualizar e analisar as redes de interação dos estudantes. A partir de um conjunto de dados sobre a interação dos estudantes, em um fórum por exemplo, esse componente desenha um grafo orientado indicando onde há interação, e a quantidade de interação entre todos os envolvidos.

#### 4.3.4 Camada de Interessados

A Camada de Interessados é responsável pela implementação das funcionalidade de configuração e envio de alertas automático conforme interesse da instituição e dos usuários. Por exemplo, os alunos tem interesse em saber como está seu desempenho em relação a turma e a turmas passadas. Os professores e tutores podem estar interessados em saber como podem melhorar seu processo de ensino, quais são os materias que estão sendo acessados, como estão sendo acessados e se estão gerando dúvidas. As instituições podem estar interessadas em saber quais as ações de reversão estão dando resultado e como podem melhorar seus índices de retenção. Devido ao foco desse trabalho e da abrangência deste tema, seu detalhamento se dará em uma extensão futura.

## 5 IMPLEMENTAÇÃO

Neste capítulo são descritos os aspectos relacionados a implementação para avaliação do modelo. A implementação foi realizada com o envolvimento de equipe de pedagogos responsáveis pela EAD na UNISINOS e os requisitos foram definidos de forma conjunta. Este processo ocorreu durante o período de um ano, envolvendo a análise de necessidades da equipe de professores e coordenadores e também os resultados obtidos com os dados disponibilizados.

### 5.1 Camada de Objetivos

Para esta dissertação foram implementados os seguintes componentes da camada de objetivos: Monitoramento e Análises e Predição e Intervenção. Porém, a arquitetura da aplicação comporta a agregação dos demais componentes pois foi desenvolvida uma aplicação em Microsoft .NET utilizando o framework MVC para implementação da camada de Objetivos.

#### 5.1.1 Monitoramento e Análises

A camada de objetivos foi implementada na forma de um sistema web e o funcionamento ocorre a partir do login do usuário na aplicação. Ao identificar-se no sistema, o usuário é direcionando para a lista de turmas (exibida na Figura 16) com as quais ele possui vínculo ou autorização de acesso. Essa lista apresenta as seguintes informações:

- Turma: nome que identifica a turma.
- Estudantes: A quantidade de estudantes, a quantidade de estudantes em risco de evasão e o percentual de estudantes em risco. Esse risco de evasão é configurado e nesse exemplo é considerado em risco todo estudante com predição de evasão maior ou igual a 60 %.
- Última Predição: Data da última predição de evasão realizada.
- Ação: Ícone que representa se a turma possui ações e se existe alguma ação pendente de alguma tarefa.
- Disciplina: nome da disciplina.
- Equipe: nome do tutor presencial, nome do tutor EAD e nome do professor.
- Data de Início: Data de início da turma.

Ao escolher uma turma o usuário é direcionado para a lista de estudantes (Figura 17) onde são listados os estudantes da turma escolhida com as seguintes informações:

- Ação: Ícone que representa se o estudante possui alguma ação e se existe ação pendente de intervenção do usuário logado.

**Figura 16:** Interface de Lista de Turmas

#	Turma	Estudantes			Última Predição	Ação	Disciplina	Equipe			Data de Início
		Total	Em Risco	%				Tutor EAD	Tutor Presencial	Professor	
1	Banco de Dados I	44	16	36 %	03/02/14		Saúde Coletiva: estratégias assistenciais	Tutor EAD 2	Tutor Presencial 2	Professor 2	30/12/13
2	Lógica	44	16	36 %	03/02/14		Lógica	Tutor EAD 1	Tutor Presencial 1	Professor 1	30/12/13

Fonte: Elaborado pelo Autor.

- Painel: Ícone para acesso ao painel do estudante.
- Estudante: Nome do estudante.
- Acessos: Apresenta o módulo ou semana em que o estudante efetuou o primeiro acesso. Essa informação serve para identificar quando o estudante começou a participar da comunidade da disciplina. Na coluna último acesso, é exibida a informação sobre a quanto tempo o estudante não acessa mais. Essa informação permite que o tutor verifique como está o acesso dos estudantes da turma e faça alguma intervenção nos estudantes que estão a muito tempo sem acessar.
- Nota GA: Nota parcial.
- Risco de Evasão: Percentual atual do risco de evasão calculado pelo modelo preditivo. A coluna diferença apresenta a variação, para que o usuário identifique se o risco está

**Figura 17:** Interface com Lista de Estudantes

#	<input type="checkbox"/>	Ação	Painel	Estudante	Acessos		Nota GA	Risco de Evasão			Última Ação
					1º Acesso	Último Acesso		Atual	Diferença	Histórico	
1	<input type="checkbox"/>			César Dutra	Módulo 1	8 dias	0,00	94 %	5,50 % ▲		4 dias
2	<input type="checkbox"/>			Giovani Santos	Módulo 2	11 dias	6,67	88 %	1,67 % ▲		4 dias
3	<input type="checkbox"/>			Guilherme Zilli	Nunca	Nunca	0,00				4 dias
4	<input type="checkbox"/>			Gustavo Bonisoni	Nunca	Nunca	0,00				4 dias
5	<input type="checkbox"/>			Issis Sibirino	Módulo 2	13 dias	47,00	22 %	13,00 % ▼		6 dias
6	<input type="checkbox"/>			Liege Fontoura	Módulo 1	9 dias	30,67	78 %	14,50 % ▲		4 dias
7	<input type="checkbox"/>			Lucas Belchior	Módulo 1	14 dias	37,33	85 %	12,75 % ▲		6 dias
8	<input type="checkbox"/>			Milene Falk	Nunca	Nunca	0,00				4 dias
9	<input type="checkbox"/>			Paulo Jardim	Módulo 1	39 dias	0,00	88 %	7,33 % ▲		4 dias
10	<input type="checkbox"/>			Rossana Pereira	Módulo 1	46 dias	0,00	94 %	1,50 % ▲		4 dias
11	<input type="checkbox"/>			Tiago Malta	Módulo 1	17 dias	38,00	54 %	6,25 % ▲		6 dias
12	<input type="checkbox"/>			Diego Abreu	Módulo 2	7 dias	58,67	36 %	12,17 % ▼		5 dias
13	<input type="checkbox"/>			Fernando Almeida	Módulo 2	7 dias	67,33	34 %	11,58 % ▼		6 dias
14	<input type="checkbox"/>			Marluce Canal	Módulo 2	10 dias	33,67	15 %	10,33 % ▼		6 dias

Fonte: Elaborado pelo Autor.

aumentando ou diminuindo. Por fim, na coluna histórico é exibido um gráfico com o risco de evasão predito. Nesse gráfico uma linha base é exibida indicando o corte de alto risco de evasão.

- Última Ação: tempo decorrido desde a última ação efetuada com esse estudante.

**Figura 18:** Interface de Gráficos



Fonte: Elaborado pelo Autor.

Quando o usuário deseja uma análise mais detalhada de um estudante, ele acessa o painel do estudante. Esse painel possui abas com informações organizadas em grupos. Conforme os direitos de acesso do usuário, as abas ficam visíveis ou não.

Na aba de gráficos (Figura 18), o usuário tem acesso aos gráficos configurados pela instituição. Quando o gráfico tem dados de referência, estes são plotados em cinza e representam a média dos estudantes que foram aprovados nas edições anteriores. A quantidade de gráficos disponíveis também está disponível conforme o perfil do usuário. Nesse exemplo, são mostrados quatro gráficos do perfil de acesso do estudante na comunidade:

- **Uso por turno:** Indica a cada módulo ou semana em que turno o estudante acessou o AVA.
- **Tempo de uso:** Permite ao usuário analisar o tempo que o estudante está dedicando aos estudos no AVA. A linha azul representa o tempo do estudante e a linha cinza representa a referência média da turma.



- Quantidade de acessos: Esse gráfico mostra a cada módulo a quantidade de acessos que o estudante efetuou e a referência média da turma.
- Quantidade de dias de acesso: Apresenta a quantidade de dias da semana em que o estudante está fazendo atividades na comunidade e a referência. Essa informação permite ao usuário identificar se o estudante dedica um ou vários dias por semana nas atividades da comunidade.

Na Figura 19 é apresentada a interface onde são exibidas as avaliações com as notas do estudante. Através dessa interface o usuário pode identificar atividades nas quais o estudantes está tendo melhor ou pior desempenho e se o mesmo está entregando as atividades.

**Figura 19:** Interface de Avaliações

The screenshot shows the 'Painel do Estudante' interface. At the top left, there is a navigation arrow and the title 'Painel do Estudante'. Below this, a student profile card for 'Ingrid Paz' is displayed, including fields for E-mail, RA, Telefone, Cidade, UF, and Polo. To the right of the profile, there are buttons for 'Turma', 'Banco de Dados I', and 'Lógica'. Below the profile, there is a '+ Adicionar ação' button. At the bottom, there are tabs for 'Gráficos', 'Avaliações', and 'Ações'. The 'Avaliações' tab is active, showing a table with columns for 'Módulo', 'Tarefa', 'Nota Final', and 'Nota Máxima'.

Módulo	Tarefa	Nota Final	Nota Máxima
0	Questões sobre fórmulas equivalentes	10,00	10,00
0	Questionário sobre Tabela-verdade	10,00	10,00
0	Tarefa - Questões sobre regras dos Quantificadores	7,00	20,00
0	Tarefa sobre Sentenças Abertas	2,00	10,00
0	Argumentos - Prova por Tabela-Verdade	10,00	10,00
0	Formalização de Sentenças	9,00	10,00
0	Tarefa sobre Dedução Natural	16,00	20,00
0	Tarefa sobre Enunciados Categóricos	7,00	10,00

Fonte: Elaborado pelo Autor.

Por fim na Figura 20 é apresentada a tela de históricos de ações realizadas para este estudante. Para cada ação são apresentadas as seguintes informações:

- Data: data que ação foi realizada.
- Indício: Motivo ou indício que levou o usuário a efetuar a ação.
- Ação: Ação em si, indica o que foi feito.
- Meio: Meio em que a ação foi executada. Se envolveu contato com o aluno, é como o contato foi feito. Se foi algum encaminhamento interno na instituição, permite identificar como foi feito para possível rastreamento.
- Descrição: Descrição da ação pelo usuário.
- Data de Registro: data que a ação foi registrada no sistema.

**Figura 20:** Interface de Histórico de Ações Realizadas

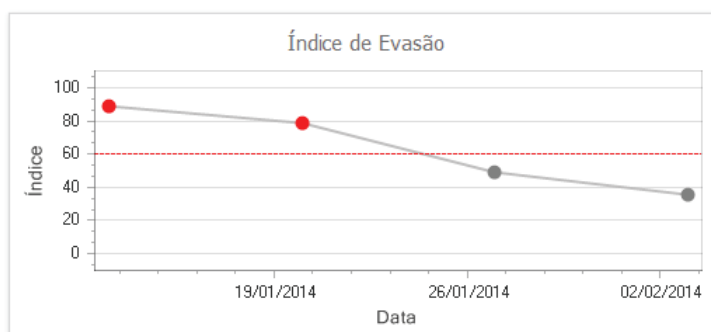
The screenshot shows a web interface for a student named Giovanni Santos. The profile includes contact information and a list of actions. The 'Ações' tab is active, displaying a table of actions with columns for Date, Start, Action, and Medium. Below the table, there is a detailed view of an email action from 25/02/2014, including the email content and a 'Continuar' button.

Data	Início	Ação	Meio
25/02/2014 17:20:37	Não acesso à comunidade	Envio de e-mail para o estudante	Envio de e-mail para o estudante
<p>Olá Giovanni Santos:</p> <p>lembro que as atividades de aula do seu curso EAD iniciaram nesta última segunda-feira. É importante que você realize seu primeiro acesso à comunidade e inicie o acompanhamento das atividades propostas.</p> <p>Obrigado</p> <p>Continuar</p> <p>criado por GVwise em 25/02/2014 17:21:06</p>			
27/02/2014 15:16:16	Baixo desempenho	Envio de e-mail para o estudante	Envio de e-mail para o estudante

Fonte: Elaborado pelo Autor.

### 5.1.2 Predição e Intervenção

A predição é executada através de um serviço que classifica cada estudante da turma conforme o modelo especificado para a turma. Essa predição gera um índice de risco de evasão que é exibido em forma de gráfico (Figura 21). O gráfico de predição mostra a data em que a predição foi feita e o índice predito.

**Figura 21:** Gráfico de Predição de Risco de Evasão

Fonte: Elaborado pelo Autor.

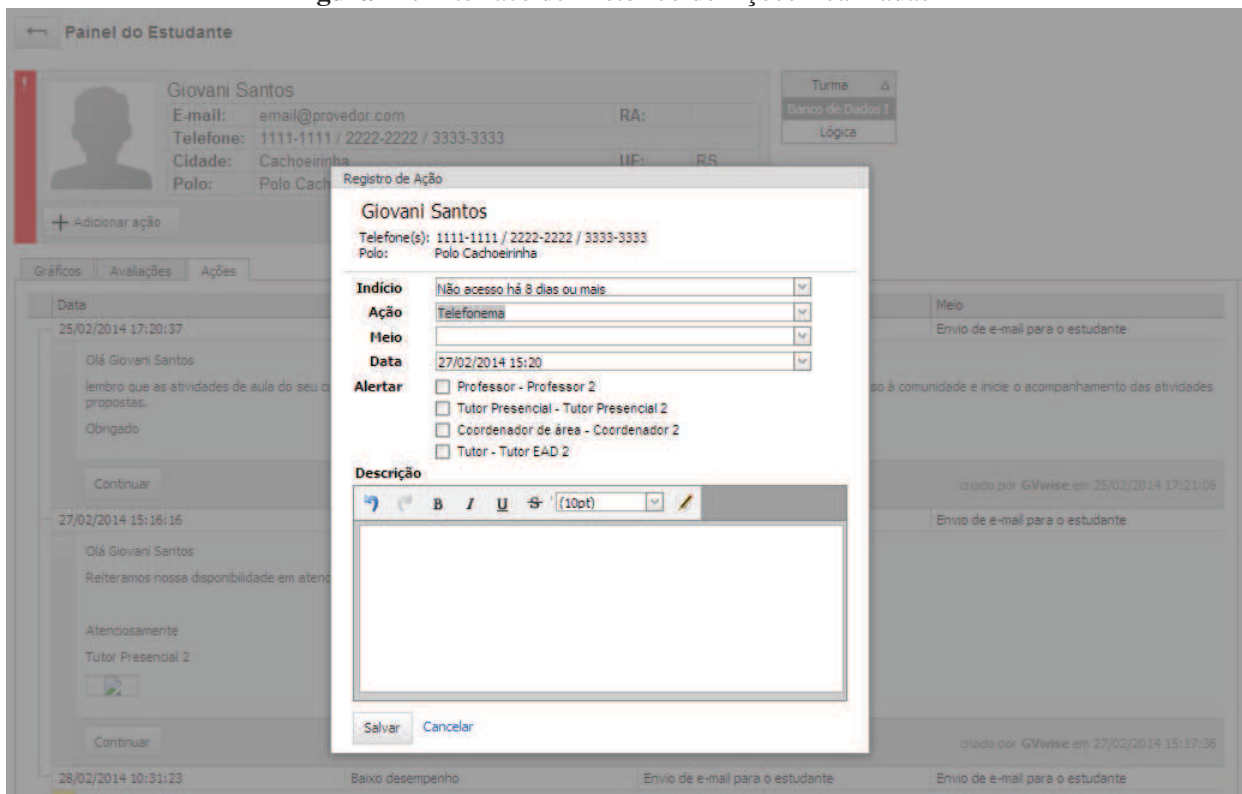
Além disso, é possível registrar novas ações associadas a um estudante. A tela de registro de ações (Figura 22) solicita que o usuário preencha algumas informações para guardar no histórico de ações.

## 5.2 Camada de Dados Educacionais

A integração de dados foi implementada na plataforma Microsoft Integration Services 2012 (SSIS) e como banco de dados foi utilizado o Microsoft Sql Server 2012 (MSSQL).



**Figura 22:** Interface de Histórico de Ações Realizadas



Fonte: Elaborado pelo Autor.

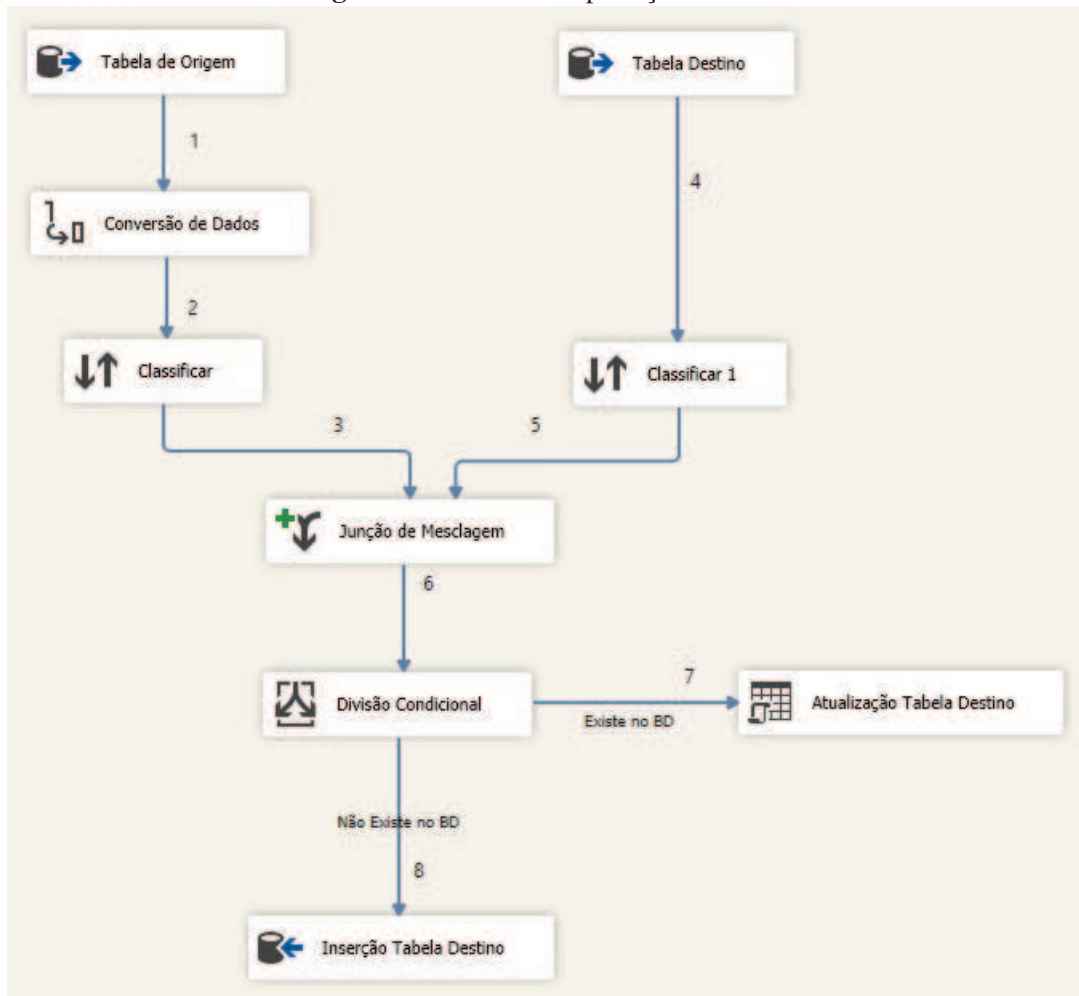
### 5.2.1 Cliente Coletor de Dados

Cada cliente coletor é um projeto do SSIS que faz a leitura, conversão e sincronização dos dados. Os dados registrados em múltiplas trilhas possuem coletores de dados específicos desenvolvidos para essa finalidade. O fluxo de sincronização dos clientes coletores implementados no SSIS, representado na Figura 23, contém oito passos:

- Passo 1 - Os dados são lidos da tabela de origem.
- Passo 2 - Cada atributo é convertido para o formato compatível com a tabela de destino para poder ser comparado.
- Passo 3 - Os dados já convertidos são classificados para otimizar o processo de junção e mesclagem.
- Passo 4 - Os dados são lidos da tabela de destino.
- Passo 5 - Os dados são classificados para otimizar o processo de junção e mesclagem.
- Passo 6 - Na junção de mesclagem é possível separar os dados que já existem no destino, dos que não existe.
- Passo 7 - Os dados que já existiam são atualizados na tabela de destino.

- Passo 8 - Os dados que não existiam são inseridos na tabela de destino.

**Figura 23:** Fluxo de importação de dados



Fonte: Elaborado pelo Autor.

### 5.2.2 Servidor de Dados

O servidor de dados foi implementado em MSSQL. A integração de dados ocorre a partir de uma tarefa agendada no MSSQL que periodicamente executa os projetos de SSIS implantados nela. Ou seja, periodicamente os clientes coletores são chamados para uma atualização de dados. A cada execução os dados do banco de dados educacional e o data warehouse educacional são atualizados.



## 6 AVALIAÇÃO

Para validação do modelo proposto foram realizados quatro estudos de caso, sendo que todos fazem o registro das informações em formas de trilhas e possuem aplicações distintas com o objetivo de validar a flexibilidade do modelo.

### 6.1 Visualização Gráfica das Múltiplas Trilhas

O primeiro estudo de caso realizado teve como objetivo avaliar a possibilidade de registro dos dados em múltiplas trilhas. Assim, foram coletados dados de estudantes do ensino médio, os quais tiveram duas aplicações: (1) apoio na orientação pedagógica, onde os funcionários tinham acesso a uma interface Web (ilustrada na Figura 24) para acompanhamento da situação dos alunos dentro da perspectiva de contextos sistêmicos; (2) mineração de dados e análise estatística para avaliação de possíveis correlações entre diferentes aspectos do contexto de um aluno, tais como, por exemplo, a idade e aprovação escolar, com objetivo de indicar possíveis melhorias com a identificação antecipada destas situações.

O estudo de caso utilizou o conceito de múltiplas trilhas para organizar informações históricas em três trilhas: “Histórico Escolar”, “Atividades Complementares” e “Ocorrências Pedagógicas”. As trilhas de 5.921 alunos foram registradas. O modelo ilustrado na Figura 25 representa o resumo dos componentes utilizados, onde dados históricos de três contextos foram coletados e organizados em trilhas. Cada componente coletor foi responsabilizado por coletar

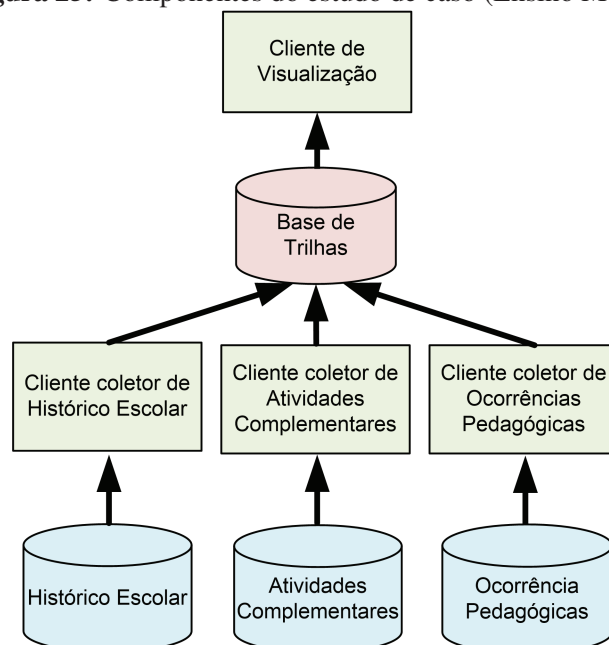
**Figura 24:** Visualização de trilhas para acompanhamento pedagógico

10		11		12		13	
4ª Série		5ª Série		6ª Série		7ª Série	
2004	2005	2006	2007	2004	2005	2006	2007
9,1 - Ciências	8,5 - Ciências	7,4 - Arte	7,9 - Arte	9,1 - Ciências	8,5 - Ciências	7,4 - Arte	7,9 - Arte
9,5 - Educação Artística	8,5 - Educação Artística	7,3 - Ciências Naturais	7,8 - Ciências Naturais	9,5 - Educação Artística	8,5 - Educação Artística	7,3 - Ciências Naturais	7,8 - Ciências Naturais
8,9 - Educação Física	8,8 - Educação Física	9,8 - Educação Física	9,9 - Educação Física	8,9 - Educação Física	8,8 - Educação Física	9,8 - Educação Física	9,9 - Educação Física
9,8 - Ensino Religioso	8,4 - Ensino Religioso	8,8 - Ensino Religioso	9,8 - Ensino Religioso	9,8 - Ensino Religioso	8,4 - Ensino Religioso	8,8 - Ensino Religioso	9,8 - Ensino Religioso
9,3 - Estudos Sociais	7,9 - Geografia	7,1 - Geografia	9,0 - Geografia	9,3 - Estudos Sociais	7,9 - Geografia	7,1 - Geografia	9,0 - Geografia
9,6 - Língua Estrangeira	8,3 - História	7,8 - História	8,8 - História	9,6 - Língua Estrangeira	8,3 - História	7,8 - História	8,8 - História
9,0 - Língua Portuguesa	8,8 - Língua Inglesa	8,3 - Língua Estrangeira - Inglês	8,0 - Língua Estrangeira - Inglês	9,0 - Língua Portuguesa	8,8 - Língua Inglesa	8,3 - Língua Estrangeira - Inglês	8,0 - Língua Estrangeira - Inglês
9,5 - Matemática	8,0 - Língua Portuguesa	7,5 - Língua Portuguesa	7,6 - Língua Portuguesa	9,5 - Matemática	8,0 - Língua Portuguesa	7,5 - Língua Portuguesa	7,6 - Língua Portuguesa
	9,5 - Matemática	8,1 - Matemática	9,4 - Matemática		9,5 - Matemática	8,1 - Matemática	9,4 - Matemática
2004	2005	2006	2007	2004	2005	2006	2007
Futebol Salão	Futebol Salão	Futebol Salão	Robótica	Futebol Salão	Futebol Salão	Futebol Salão	Robótica
		Robótica					

Fonte: Elaborado pelo autor.

e registrar o histórico de um domínio específico e os componentes coletores armazenaram as informações obtidas na camada servidor da arquitetura.

**Figura 25:** Componentes do estudo de caso (Ensino Médio)



Fonte: Elaborado pelo autor.

O estudo de caso inicial está relacionada com as atividades de aconselhamento pedagógico e com o apoio aos professores no relacionamento com os alunos. No dia a dia de uma instituição de ensino é necessária a intervenção frequente de professores e coordenadores no acompanhamento e aconselhamento dos alunos. Muitas vezes a falta de informações e, em especial, do relacionamento de informações diversas, pode levar a uma intervenção de menor qualidade da que seria possível caso todo contexto sistêmico do aluno estivesse disponível para consulta. Para atender a esta demanda, o histórico registrado nas trilhas foi exibido em uma linha de tempo, permitindo a análise do desempenho do aluno em determinado contexto, de acordo com a necessidade de cada situação.

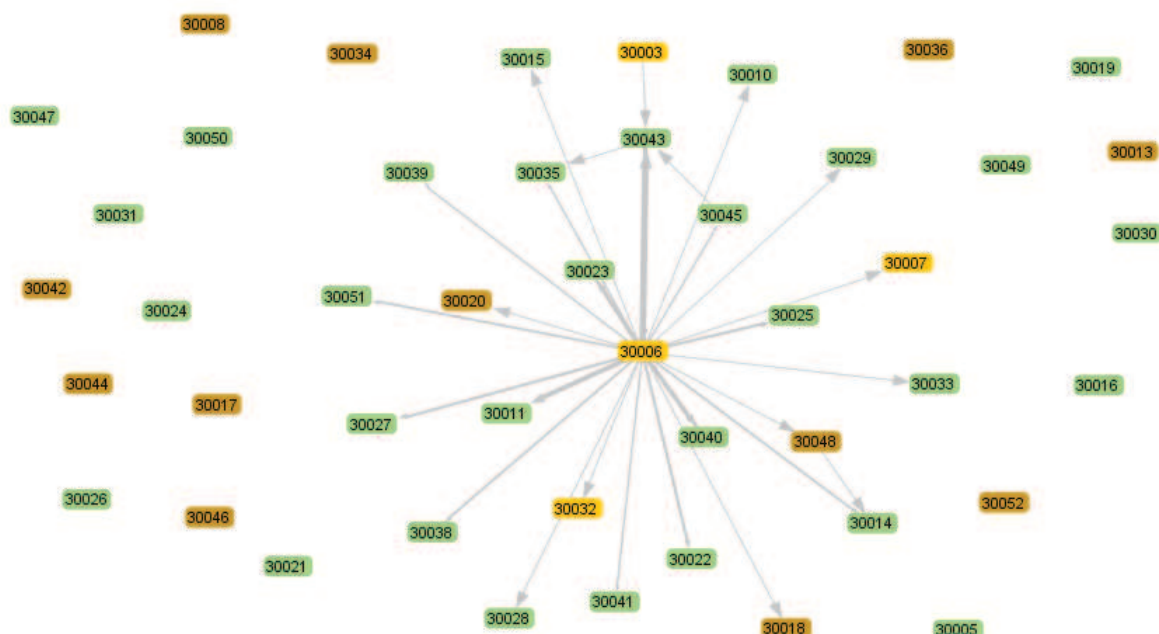
A Figura 24 apresenta trechos parciais da situação histórica de um aluno, onde é possível verificar a idade em que o mesmo cursou cada ano do ensino fundamental e médio, o desempenho do mesmo em cada disciplina, atividades complementares que foram realizadas e ocorrências pedagógicas. Através desta interface Web é possível a análise de fatores importantes, tais como a influência de atividades complementares no desempenho escolar e no comportamento do aluno. A avaliação empírica e preliminar desta aplicação deu-se através de entrevistas com profissionais da educação, que indicaram a sua expectativa de possibilidades de apoio desta ferramenta para atividades de aconselhamento pedagógico e acompanhamento.

## 6.2 Visualização da Interação

O segundo estudo de caso realizado teve por objetivo utilizar dados obtidos no Ensino Superior, com a seguinte aplicação: Analisar dados de uma turma de alunos na modalidade EAD para a visualização gráfica da interação relacionada com situações de evasão ou baixo desempenho. O estudo de caso aqui descrito consiste do acompanhamento de dados gerados por 377 alunos de graduação durante uma ocorrência da disciplina, sendo que foram acompanhados e tratados os aspectos de interação e de realização de tarefas de avaliação, bem como troca de mensagens e realização de intervenções em um ambiente virtual de aprendizagem.

Em um dos aspectos evidenciados como importantes para o acompanhamento de possíveis situações demandando atenção, está a interação entre os alunos nas diversas ferramentas de ambientes virtuais de aprendizagem. Para este acompanhamento, foi desenvolvido um componente de visualização que disponibiliza, em forma de um grafo, aspectos relacionados com a interação e desempenho, a partir do qual é possível analisar a relação entre o desempenho dos estudantes e a participação nos fóruns da disciplina. No exemplo da figura 26, o resultado obtido com a análise parcial de uma turma de alunos representa a interação destes alunos de uma disciplina de Matemática para Administração, em um curso de graduação de Administração, realizado na modalidade de Educação a Distância. Os nodos representados em amarelo são os tutores da disciplina, os nodos em verde são os alunos aprovados e os nodos em marrom são os alunos reprovados. As arestas do grafo representam a interação entre os diversos participantes na ferramenta fórum de discussão.

**Figura 26:** Visualização de interações entre alunos em fórum



Através de ferramentas de visualização como este grafo é possível prover o suporte para análises relacionando a interação dos alunos no ambiente virtual de aprendizagem e o seu resultado final nas disciplinas. Nos experimentos realizados, em diversos contextos resta evidenciado que a participação ativa nos fóruns desta disciplina está relacionada positivamente com o desempenho, pois alguns dados indicam que enquanto 15% dos alunos que interagiram com maior intensidade reprovaram, o índice de reprovação aumentou para 45% entre os alunos que não interagiram significativamente.

Deste modo, destaca-se a identificação de aspectos positivos tais como o apoio para atividades de acompanhamento pedagógico também no caso de Ensino Superior, com a descrição de uma ferramenta para visualização de possíveis relações entre atividades e desempenho.

### 6.3 Predição da Evasão

O terceiro estudo de caso teve por objetivo a validação do modelo proposto de gerenciamento de múltiplas trilhas com o objetivo de identificar os alunos com potencial risco de evasão escolar em cursos EAD com a utilização de técnicas de MD).

Para este estudo de caso, foi realizada a coleta de dados de dezenove turmas de quatro disciplinas da Universidade do Vale do Rio dos Sinos (UNISINOS). Se tratando de disciplinas de EAD, os dados coletados são formados por registros de interação com a ferramenta Moodle e dados do ERP. Na coleta realizada foram obtidas dados de 603 alunos de dezenove turmas em que as aulas semestrais ocorreram entre o período de 2011 à 2012. A partir disso, os dados foram modeladas de forma a considerar as interações semanais com as diferentes ferramentas utilizadas para o ensino.

Com a coleta e modelagem das informações foi possível então a aplicação de técnicas de MD de forma que, ao final de cada semana foi gerado um relatório indicando quais alunos foram classificados com risco de evasão. Ou seja, na modelagem utilizada, não há necessidade do término do semestre para a predição dos alunos com potencial de evasão, o que possibilita a intervenção de tutores e professores no decorrer do desenvolvimento da disciplina, aumentando as possibilidades de reversão destes possíveis quadros.

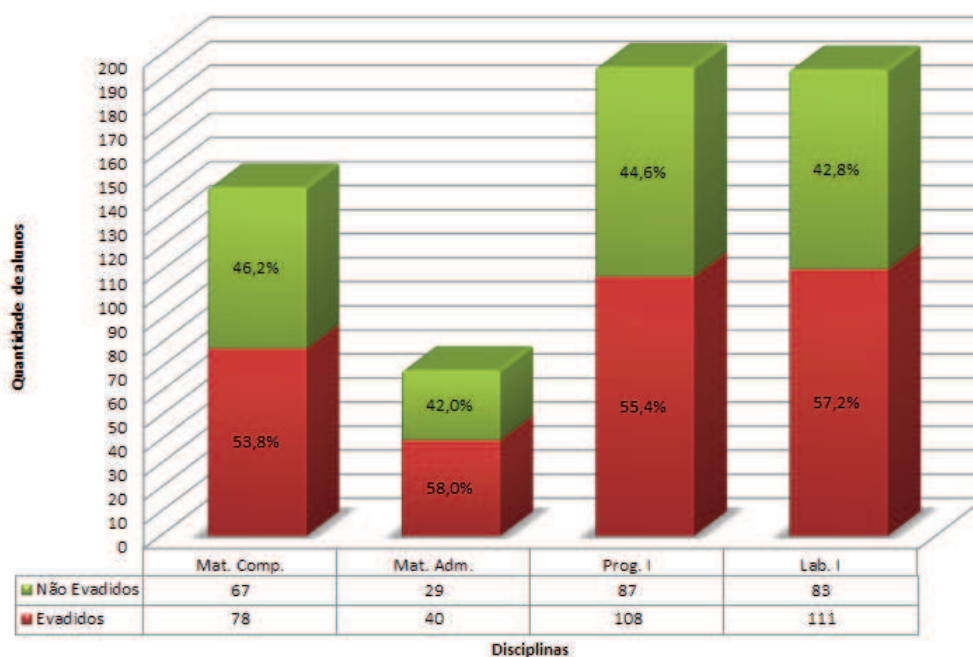
Para a aplicação de técnicas de MD a técnica de Redes Neurais Artificiais (RNAs) foi escolhida por ser clássica na área de mineração e apresentar um desempenho satisfatório em outros problemas de mineração de dados (BISHOP, 2007; HAYKIN, 2001). Por ser adotada uma técnica de aprendizado de máquina supervisionado (ALPAYDIN, 2010), em que dados históricos são utilizados para o treinamento da técnica, os dados dos alunos pertencentes às turmas que ocorreram anteriormente ao período em que o estudo de caso foi realizado foram divididos em duas partes, a primeira para o treinamento das RNAs e o restante utilizado para a avaliação do modelo de mineração obtido.

A Figura 27 apresenta o gráfico das quantidades de alunos e o percentual referente a evasão e não evasão observados na conclusão das disciplinas acompanhadas. O acompanhamento iniciou



a partir das primeiras semanas de aula, quando os dados de acesso dos alunos foram avaliados com auxílios de técnicas de aprendizado de máquina. Com isso, o objetivo foi identificar dois grupos de alunos, alunos com chance de evasão da disciplina e alunos que permaneceriam até o final do período lecionado. Desta forma, semanalmente relatórios foram enviados aos coordenadores responsáveis contendo uma indicação de evasão para cada aluno de cada uma das disciplinas.

**Figura 27:** Alunos por disciplina das classes Não Evadidos e Evadidos



Fonte: Elaborado pelo autor.

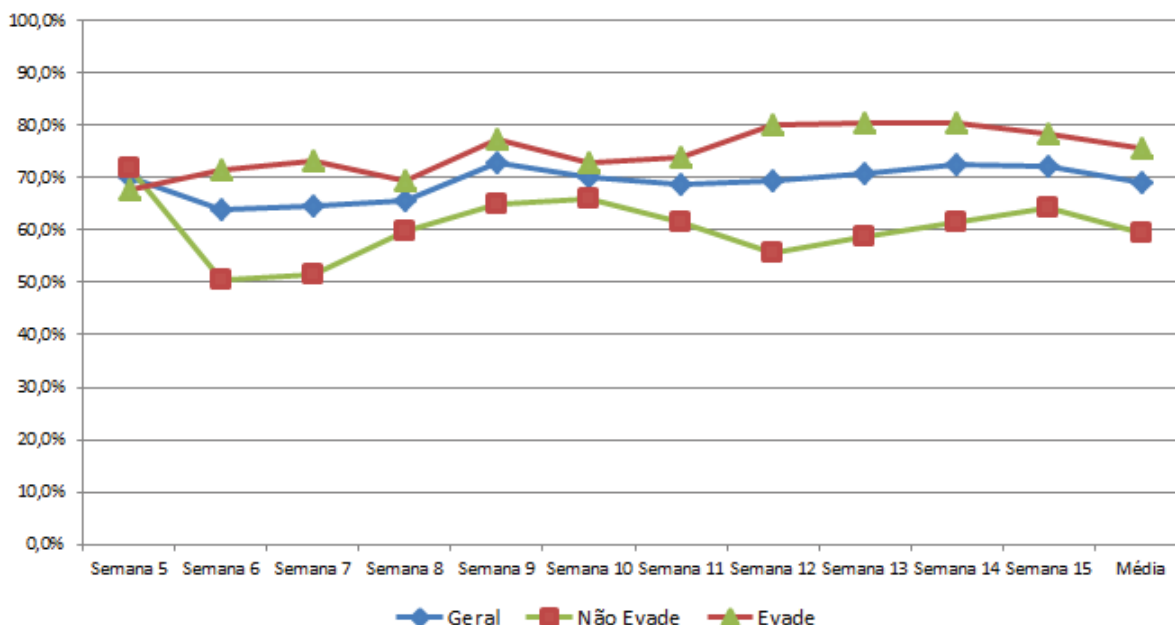
Ao final do estudo de caso, os indicadores de evasão foram avaliados semanalmente considerando a taxa de acerto tanto de cada um dos grupos de alunos, como a média de acerto obtida em cada uma das disciplinas, já que cada uma delas apresenta características particulares de estrutura do curso, formatos de materiais e avaliação.

### 6.3.1 Disciplina 1

A disciplina 1 é composta por 20 módulos semanais. No presente estudo a predição iniciou na 5ª semana e se estendeu até a 15ª semana. Foram analisadas duas turmas da disciplina.

Os resultados apresentados no gráfico da Figura 28 evidenciam que foi possível prever em média 75,7% dos casos de evasão durante todo o período analisado. Apesar dessa taxa de predição da evasão, a taxa de indicação de alunos que permaneceriam no curso foi de 59,5% de acerto. A análise dos logs realizada no final do semestre revela que as notas de atividades foram registradas no sistema EAD após algumas semanas da entrega, fazendo com que muitos alunos, apesar do seu comportamento de acesso, fossem indicados como potenciais à evasão por não



**Figura 28:** Acerto de predição da disciplina 1

Fonte: Elaborado pelo autor.

apresentarem notas cadastradas. Além disso, é importante destacar que os dados dessa disciplina disponibilizados para a construção da solução apresentam uma quantidade considerada baixa de alunos, menos de 100.

### 6.3.2 Disciplina 2

A disciplina 2 é composta por 10 módulos semanais e o período observado foi a partir da terceira semana de aula até a sétima.

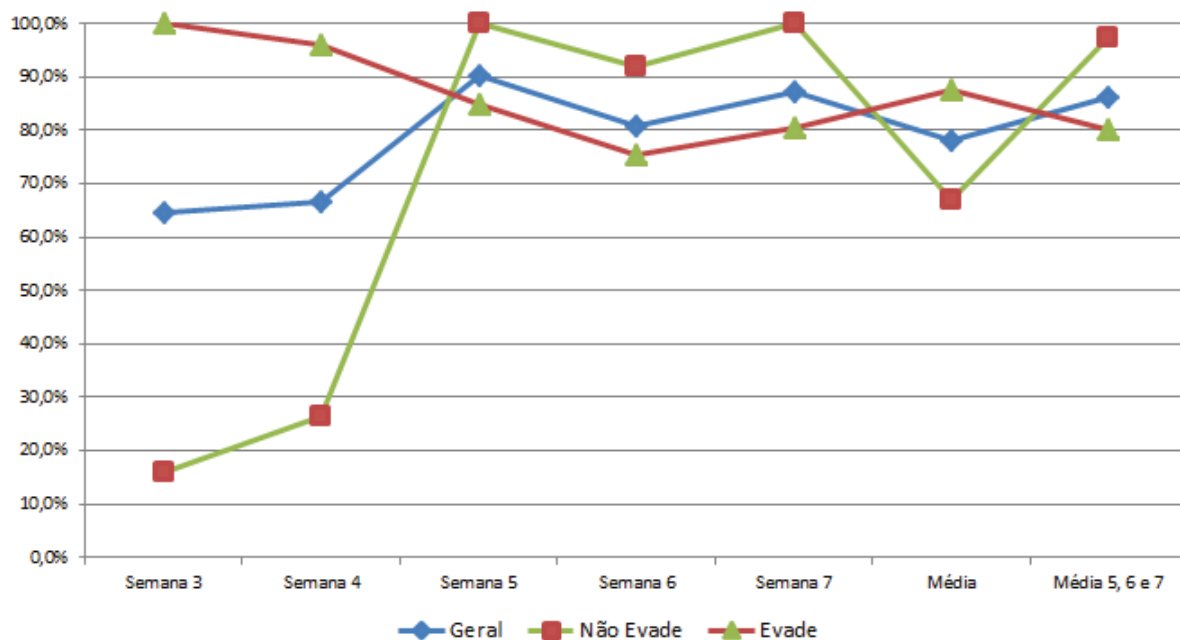
Neste período, conforme Figura 29 a solução desenvolvida conseguiu prever em média 91% dos alunos que realmente evadiram e 78,9% considerando alunos que evadiram e que não evadiram. O fato do lançamento das notas ocorrer após algumas semanas da entrega das atividades também teve influência nesta disciplina. Foi observado que as notas das primeiras tarefas foram lançadas somente na quarta semana do curso e então, consideradas na análise somente a partir da quinta. A média de acerto dos dois grupos de alunos considerando somente as três últimas semanas observadas (5, 6 e 7) foi de 90,1%, mais de 11% da média de todo o período observado.

### 6.3.3 Disciplina 3

A disciplina 3 é composta por 10 módulos semanais. No presente estudo a predição iniciou na 3ª semana e se estendeu até a 7ª semana. Foram analisadas três turmas da disciplina.

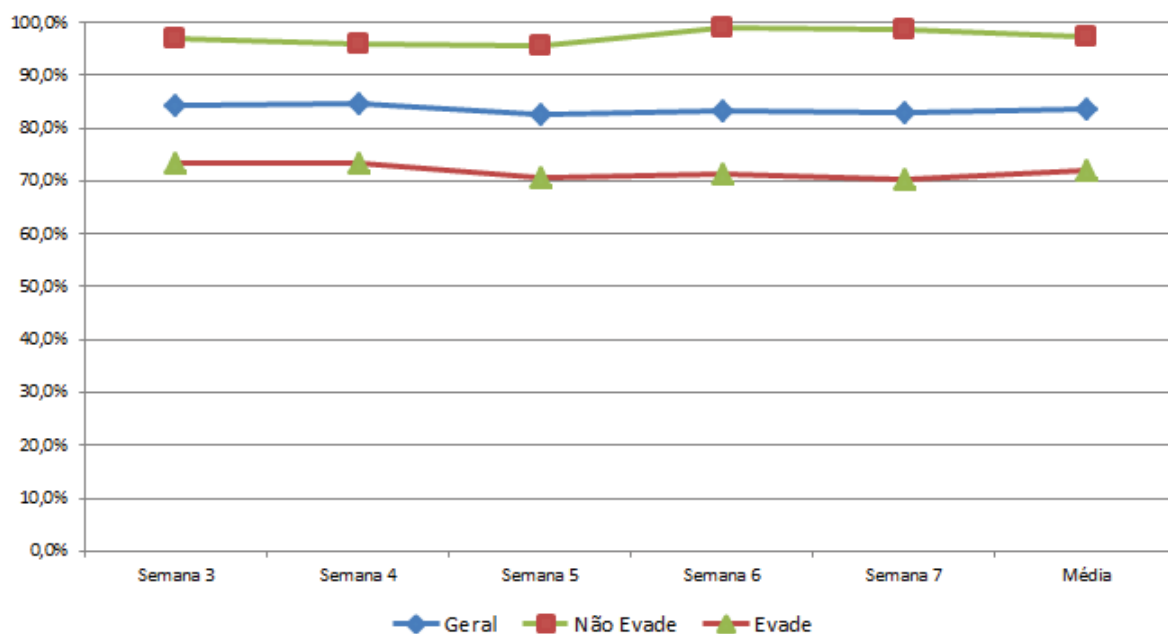
Para a disciplina 3 não foram consideradas as notas de tarefas ao longo do curso, sendo observado, conforme a Figura 30, somente o comportamento de acesso dos alunos, sendo que

**Figura 29:** Acerto de predição da disciplina 2



Fonte: Elaborado pelo autor.

**Figura 30:** Acerto de predição da disciplina 3



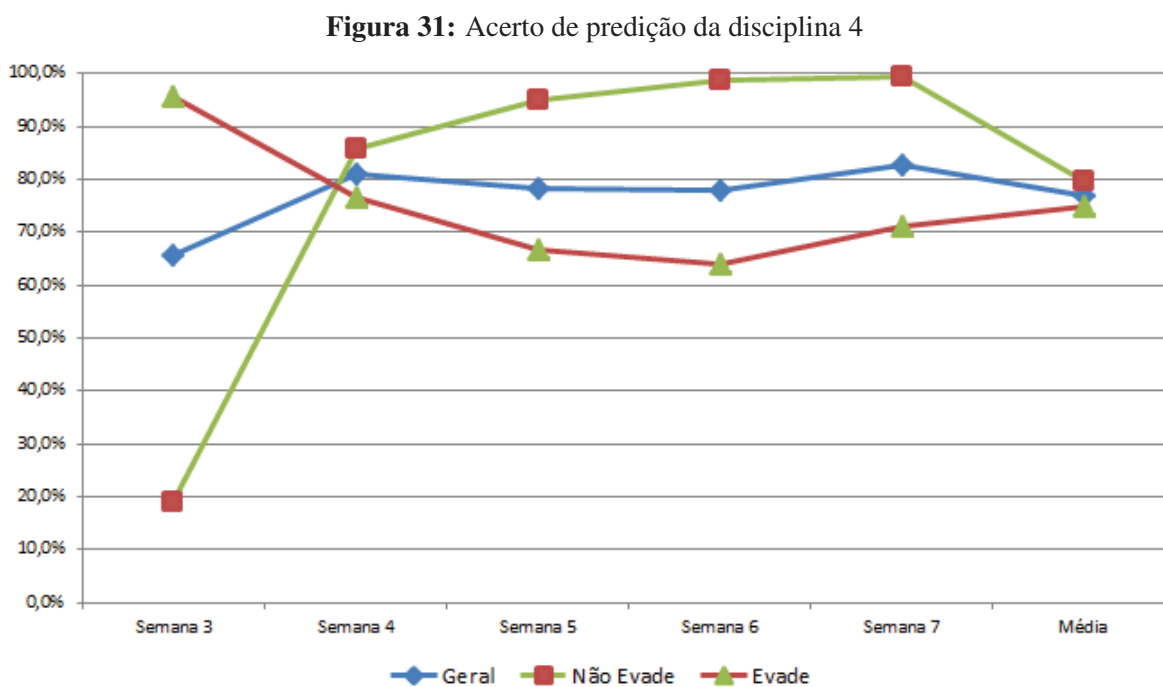
Fonte: Elaborado pelo autor.

apresentou uma média de 83,8% de acerto para os dois grupos de alunos, 72,6% de acerto dos alunos que realmente evadiram. A disciplina não apresentou em nenhuma semana para nenhum dos dois grupos de alunos uma taxa menor que 70%. Desta forma, esta disciplina foi a que apresentou o comportamento mais uniforme durante o período observado dentro todas as

turmas acompanhadas no estudo de caso.

#### 6.3.4 Disciplina 4

A disciplina 4 é composta por 10 módulos semanais. No presente estudo a predição iniciou na 3ª semana e se estendeu até a 7ª semana. Foram analisadas três turmas da disciplina.



Fonte: Elaborado pelo autor.

Na disciplina 4, as notas das tarefas também não foram consideradas. A solução apresentou, conforme a Figura 31 uma taxa de acerto média de 76,5% considerando alunos evadidos e não evadidos e de 74,2% para os alunos evadidos. Porém, os resultados da primeira semana de avaliação foram consideravelmente tendenciosos para um dos grupos de alunos, o que indica que para essa disciplina a predição seja viável somente a partir da quarta semana de aula.

#### 6.3.5 Considerações

De maneira geral, os resultados obtidos nesse estudo de caso são considerados bastante satisfatórios frente à baixa quantidade de dados disponíveis para cada disciplina, considerando tanto o número de alunos quanto a quantidade de informações de cada um deles. Neste estudo de caso foram acompanhados 693 alunos, destes o número de 337 (ou seja, 48%) realmente evadiram no decorrer da disciplina. Sendo assim, considerando a média dos resultados obtidos por disciplina, a utilização da solução possibilitaria a realização de ações pedagógicas em uma quantidade de 258 (76,5%) alunos evadidos a partir da terceira semana.

## 6.4 Reversão de Evasão

O quarto estudo de caso teve por objetivo a realização de ações nos estudantes com predição de evasão para verificar o potencial de reversão. Uma das bases motivacionais para a busca da reversão da evasão pode ser associada com a ampla interação com equipes interdisciplinares envolvidas no processo de implementação e manutenção dos cursos de EAD acompanhados. A partir da confirmação obtida quanto à possibilidade de obtenção de predições ligados à evasão, adotou-se um posicionamento de utilização desta informação a partir de ações pedagógicas com o objetivo de fomentar a reversão deste quadro.

De acordo com a interação efetivada entre os setores ligados ao projeto e desenvolvimento do software, pedagogos da instituição e grupo de professores, foram delimitadas etapas de um processo para garantir que os resultados dos relatórios de predição da evasão fossem utilizados como apoio na prática docente dos professores e tutores envolvidos com as turmas que participaram no estudo de caso. Isso envolveu um planejamento prévio, capacitação de participantes e posterior acompanhamento das ações pedagógicas. Deste modo, a cada semana os professores e tutores envolvidos com o estudo de caso possuíam acesso ao sistema proposto com a predição de evasão e a partir desta informação desencadeavam ações específicas de acolhimento e atenção aos alunos. Estas ações pedagógicas possuem caráter proativo e podem variar bastante, de acordo com as necessidades identificadas e com o momento de desenvolvimento dos conteúdos.

Para a realização do estudo foram analisadas doze disciplinas e estas tiveram a quantidade de alunos, percentual de evasão e percentual de aprovação registrados antes e depois da utilização da aplicação. A Tabela 2 possibilita que seja feita uma análise das taxas de evasão e aprovação de maneira comparativa entre as doze disciplinas que participaram dos estudos e nas quais ações proativas foram realizadas e as mesmas disciplinas antes da participação do estudo.

**Tabela 2:** Comparativo de taxas de evasão e aprovação antes e depois da utilização do modelo proposto

Disciplina	Antes			Depois		
	Alunos	Evasão	Aprovação	Alunos	Evasão	Aprovação
Disciplina 1	94	28,7%	68,1%	96	16,7%	72,9%
Disciplina 2	99	22,2%	76,8%	96	16,7%	79,2%
Disciplina 3	154	18,2%	59,1%	93	23,7%	62,4%
Disciplina 4	171	37,4%	59,1%	116	33,6%	63,8%
Disciplina 5	83	25,3%	65,1%	62	27,4%	66,1%
Disciplina 6	181	40,3%	48,6%	134	29,9%	56,7%
Disciplina 7	38	42,1%	57,9%	77	31,2%	63,6%
Disciplina 8	147	30,6%	63,3%	125	23,2%	72,0%
Disciplina 9	147	55,8%	27,2%	83	50,6%	28,9%
Disciplina 10	133	36,1%	38,3%	111	28,8%	36,9%
Disciplina 11	317	34,1%	49,5%	260	27,7%	50,8%
Disciplina 12	42	23,8%	76,2%	30	20,0%	80,0%

Fonte: Elaborado pelo autor.

**Tabela 3:** Resultado do Teste-T para Evasão

	<i>Evasão Antes</i>	<i>Evasão Depois</i>
Média	0,328880011	0,274477013
Variância	0,010811785	0,008328965
Observações	12	12
Correlação de Pearson	0,870024031	
Hipótese da diferença de média	<b>0</b>	
gl	11	
Stat t	3,675843172	
P(T<=t) uni-caudal	<b>0,001826103</b>	
t crítico uni-caudal	1,795884819	
P(T<=t) bi-caudal	0,003652206	
t crítico bi-caudal	2,20098516	

Fonte: Elaborado pelo autor.

Os dados evidenciam que após a utilização da aplicação houve uma redução de evasão nas disciplinas e um aumento da taxa de aprovação. Para validação dos resultados foi aplicado o Teste-T Student ao nível de significância de 5% com as seguintes hipóteses:

- Hipótese nula - Os resultados obtidos antes da utilização da aplicação **são iguais** aos resultados obtidos após a utilização da aplicação, ou seja, a melhoria foi obtida ao acaso.
- Hipótese alternativa - Os resultados obtidos antes da utilização da aplicação **são diferentes** aos resultados obtidos após a utilização da aplicação, ou seja, a melhoria não foi obtida ao acaso.

Conforme dados da Tabela 3, onde o valor-p uni-caudal é 0,18%, rejeita-se a hipótese nula e aceita-se a hipótese alternativa. Assim afirma-se que há relação entre as taxas de evasão obtidas e a utilização da aplicação proposta, pois o valor-p é menor do que 5%.

**Tabela 4:** Resultado do Teste-T para Aprovação

	<i>Aprovação Antes</i>	<i>Aprovação Depois</i>
Média	0,574267606	0,611121394
Variância	0,021407695	0,024802372
Observações	12	12
Correlação de Pearson	0,983747904	
Hipótese da diferença de média	<b>0</b>	
gl	11	
Stat t	-4,318738475	
P(T<=t) uni-caudal	<b>0,000608616</b>	
t crítico uni-caudal	1,795884819	
P(T<=t) bi-caudal	0,001217233	
t crítico bi-caudal	2,20098516	

Fonte: Elaborado pelo autor.

Observou-se também que a taxa de aprovação dos alunos melhorou. Conforme dados da Tabela 4, onde o valor-p uni-caudal é 0,06%, rejeita-se a hipótese nula e aceita-se a hipótese alternativa. Assim afirma-se que há relação entre as taxas de aprovação obtidas e a utilização da aplicação proposta, pois o valor-p é menor do que 5%.

Os estudos confirmaram que a realização de ações proativas podem evitar a evasão e melhorar o desempenho dos alunos. Observa-se que as taxas de evasão e aprovação das turmas que participaram do estudo são melhores do que as taxas de evasão e aprovação das turmas que não participaram do estudo. Apesar do fenômeno da evasão estar associado com múltiplos fatores, conforme descrito anteriormente, considera-se que estes resultados indicam um direcionamento positivo para o processo proposto.



## 7 CONCLUSÃO

Este trabalho apresentou um modelo de arquitetura para um sistema de LA com foco na redução de evasão em cursos de EAD, denominando GVwise. Também foram apresentados os conceitos da organização dos dados em múltiplas trilhas (CAMBRUZZI; RIGO; BARBOSA, 2012). Os modelos de referência para LA utilizados para definição das camadas foram: CHATTI et al. (2012) e GRELLER; DRACHSLER (2012). Também foram analisados aspectos do problema de evasão escolar, que constitui um elemento importante a ser tratado para contribuir com que um número maior de estudantes finalize com sucesso o ensino superior.

A MDE e o LA são áreas relativamente recentes de pesquisa, porém são áreas que encontram atualmente um amplo suporte nos processos de implantação de sistemas informatizados de gestão escolar, no uso de ambientes virtuais de aprendizagem e na tendência de ampliação de modalidades de ensino para a EAD. Todos estes ambientes possuem a capacidade de geração de dados relacionados com os diversos aspectos do processo de ensino aprendizagem. Além de permitirem a coleta e integração de dados originados em ambientes com funções complementares, enriquecendo as possibilidades de identificação de padrões ou informações de interesse.

Entretanto, destaca-se a importância da necessidade da utilização destas técnicas em conjunto com processo que envolva um amplo mapeamento de fatores associados com o evento da evasão escolar. Este processo deveria ser constituído de modo a envolver os diversos setores das instituições, dado que modelos teóricos acerca da evasão escolar apontam para múltiplas causas para este fenômeno, sendo que estas causas, em diversas medidas, são consideradas como inter-relacionadas. Esta análise aponta para possibilidades de soluções mais efetivas considerando-se a integração de dados históricos importantes, porém complementadas com conjuntos de dados mais dinâmicos, obtidos com o acompanhamento da interação de alunos e professores ao longo de períodos de semestres letivos. Também é importante destacar a necessidade de acompanhamento das ações pedagógicas desencadeadas, para aproveitamento de seus resultados em ações futuras, favorecendo assim o diagnóstico precoce e a realização de ações pedagógicas relevantes.

### 7.1 Contribuição

Como principal contribuição desse trabalho destaca-se o modelo de arquitetura para um sistema de LA com foco na redução de evasão em cursos de EAD, denominando GVwise e os conceitos da organização dos dados em múltiplas trilhas para ambientes educacionais.

Como contribuição adicional, foram publicados e apresentados à comunidade acadêmica os seguintes artigos científicos:

- Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades (RIGO; CAZELLA; CAMBRUZZI, 2012).



- Um Modelo para Gerenciamento de Múltiplas Trilhas Aplicado a Sistemas de Apoio à Educação (CAMBRUZZI; RIGO; BARBOSA, 2012). Artigo selecionado entre os 7 melhores trabalhos do 'XXIII Simpósio Brasileiro de Informática na Educação (SBIE)', entre 441 artigos submetidos ao evento e 88 aprovados., Sociedade Brasileira de Computação (SBC) - Comitê de Programa do SBIE 2012.
- Um Modelo para Gerenciamento de Múltiplas Trilhas Aplicado a Sistemas de Apoio à Educação (CAMBRUZZI et al., 2012).
- Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios (RIGO et al., 2014).

Além dos artigos foi apresentado um minicurso no Congresso Internacional ABED de Educação a Distância com o título Processos de Retenção da Evasão na EAD, integrando ao planejamento pedagógico os recursos de Mineração de Dados Educacionais, Learning Analytics e personalização de conteúdos no ano de 2013. No mesmo congresso foi realizada uma apresentação da mesa-redonda: Evasão - processos de retenção da evasão na EaD - planejamento pedagógico e mineração de dados educacionais - coleta de dados.

## 7.2 Trabalhos Futuros

Dada a diversidade dos dados envolvidos, existem possibilidades para utilização de conjuntos combinados de técnicas de MDE e LA, sendo que esta exploração de algoritmos, técnicas e mecanismos é destacada com um dos pontos cruciais para que sejam alcançados os resultados abrangentes e úteis que são esperados para a contribuição na mitigação de contextos de evasão escolar. A exploração e experimentação de alguns destes conjuntos de técnicas combinadas é uma linha de atuação de trabalhos futuros, ampliando as possibilidades do sistema atual.

Além da inclusão de aplicações de tratamento de mensagens textuais como componente adicional do modelo (dissertação sendo desenvolvida por aluno do PIPCA, Evandro Metz) e a inclusão de modelos de diagnóstico de comportamento de aprendizes de acordo com características de modelos de aprendizagem, como forma de apoiar a personalização das ações e alertas do modelo (dissertação sendo desenvolvida por aluno do PIPCA, Leonardo Heidrich).

## REFERÊNCIAS

- ADACHI, A. A. C. T. **Evasão e Evadidos nos Cursos de Graduação da Universidade Federal de Minas Gerais**. 2009.
- ALPAYDIN, E. **Introduction to Machine Learning**. 2nd. ed. [S.l.]: The MIT Press, 2010.
- ANANTHANARAYANAN, G.; HARIDASAN, M.; MOHAMED, I.; TERRY, D.; THEK-KATH, C. A. StarTrack: a framework for enabling track-based applications. In: **MOBILE SYSTEMS, APPLICATIONS, AND SERVICES**, 7., 2009, New York, NY, USA. **Proceedings...** ACM, 2009. p. 207–220. (MobiSys '09).
- ANDRIOLA, W. B.; ANDRIOLA, C. G.; MOURA, C. P. Opiniões de docentes e de coordenadores acerca do fenômeno da evasão discente dos cursos de graduação da Universidade Federal do Ceará (UFC). **Ensaio: Avaliação e Políticas Públicas em Educação**, [S.l.], v. 14, p. 365 – 382, 09 2006.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de Dados Educacionais: oportunidades para o brasil. **Revista Brasileira de Informática na Educação**, [S.l.], v. 19, n. 02, Aug. 2011.
- BIENKOWSKI, M.; FENG, M.; MEANS, B. **Enhancing teaching and learning through educational data mining and learning analytics: an issue brief**. [S.l.]: U.S. Department of Education, Office of Educational Technology, 2012.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. 2. ed.. ed. [S.l.]: Springer, 2007.
- BRA, P.; SMITS, D.; SLUIJS, K.; CRISTEA, A.; FOSS, J.; GLAHN, C.; STEINER, C. GRAPPLE: learning management systems meet adaptive learning environments. In: PEÑA-AYALA, A. (Ed.). **Intelligent and Adaptive Educational-Learning Systems**. [S.l.]: Springer Berlin Heidelberg, 2013. p. 133–160. (Smart Innovation, Systems and Technologies, v. 17).
- CAMBRUZZI, W. L.; MORAES, R. de; LEITHARDT, V. R. Q.; MENDES, C.; GEYER, C. F. R.; COSTA, C. A. da; BARBOSA, J. L. V.; RIGO, S. J. Um Modelo para Gerenciamento de Múltiplas Trilhas Aplicado a Sistemas de Apoio à Educação. In: **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**, 2012, Rio de Janeiro. **Anais...** [S.l.: s.n.], 2012.
- CAMBRUZZI, W. L.; RIGO, S. J.; BARBOSA, J. L. V. Uma proposta para Gerenciamento de Múltiplas Trilhas em Ambientes Educacionais. In: **SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA)**, 2012, São Paulo. **Anais...** [S.l.: s.n.], 2012.
- CARR, S. **As Distance Education Comes of Age, the Challenge Is Keeping the Students - Technology - The Chronicle of Higher Education**. Acessado em: 22/11/2012, Disponível em: <http://chronicle.com/article/As-Distance-Education-Comes-of/14334>.
- CENSOEAD.BR. **Relatório analítico da aprendizagem a distância no Brasil**. São Paulo: Pearson Education do Brasil, 2010.
- CENSOEAD.BR. **Relatório analítico da aprendizagem a distância no Brasil 2012**. Curitiba: Editora IBPEX, 2012.

CHATTI, M. A.; DYCKHOFF, A. L.; SCHROEDER, U.; THÜS, H. A Reference Model for Learning Analytics. **Int. J. Technol. Enhanc. Learn.**, Inderscience Publishers, Geneva, SWITZERLAND, v. 4, n. 5/6, p. 318–331, Jan. 2012.

DEY, A. K. Understanding and Using Context. **Personal Ubiquitous Computing**, London, UK, UK, v. 5, n. 1, p. 4–7, feb 2001.

DRIVER, C.; CLARKE, S. An application framework for mobile, context-aware trails. **Pervasive Mob. Comput.**, Amsterdam, The Netherlands, The Netherlands, v. 4, n. 5, p. 719–736, Oct. 2008.

DURAND, G.; LAPLANTE, F.; KOP, R. A Learning Design Recommendation System Based on Markov Decision Processes. **KDD-2011: 17th ACM SIGKDD**, [S.l.], 2011.

FAVERO, R. V. M.; FRANCO, S. R. Um estudo sobre a permanência e a evasão na Educação a Distância. **RENOTE - Revista Novas Tecnologias na Educação**, [S.l.], v. 4, n. 2, 2006.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. (Ed.). **Advances in Knowledge Discovery and Data Mining**. [S.l.]: AAAI/MIT Press, 1996.

GAMS, E.; REICH, S. The TrailTRECer Framework: applying open hypermedia concepts to trails. **Journal of Universal Computer Science**, [S.l.], v. 8, n. 10, p. 913–923, oct 2002. [http://www.jucs.org/jucs\\_8\\_10/the\\_trailtreecer\\_framework\\_applying](http://www.jucs.org/jucs_8_10/the_trailtreecer_framework_applying).

GRELLER, W.; DRACHSLER, H. Translating Learning into Numbers: a generic framework for learning analytics. **Educational Technology & Society**, [S.l.], v. 15, n. 3, p. 42–57, November 2012.

HAYKIN, S. **Redes neurais: principios e prática**. 2. ed. ed. Porto Alegre: Bookman, 2001.

JOHNSON, L.; ADAMS, S.; CUMMINS, M.; ESTRADA, V.; FREEMAN, A.; LUDGATE, H. **The NMC Horizon Report: 2013 higher education edition**. [S.l.]: NMC, 2013.

KAMPPFF, A. J. C. **Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente**. [S.l.: s.n.], 2009.

KEEGAN, D. **Foundations of distance education**. 3a. ed. London: Routledge, 1996.

KOEDINGER, K. R.; CUNNINGHAM, K.; SKOGSHOLM, A.; LEBER, B. An open repository and analysis tools for fine-grained, longitudinal learner data. In: EDUCATIONAL DATA MINING 2008: 1ST INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING, PROCEEDING, 2008. **Anais...** [S.l.: s.n.], 2008. p. 157–166.

KOTSIANTIS, S. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. **Artificial Intelligence Review**, [S.l.], v. 37, n. 4, p. 331–344, 2012.

LEVENE, M.; PETERSON, D. **Trail Records and Ampliative Learning**. [S.l.]: School of Computer Science and Information Systems, Birkbeck College, University of London, Research Report BBKCS-02-10, 2002.

LEVY, Y. Comparing dropouts and persistence in e-learning courses. **Computers & Education**, [S.l.], v. 48, n. 2, p. 185 – 204, 2007.

- LI, N.; MATSUDA, N. A machine learning approach for automatic student model discovery. **Proceedings of the 4th International Conference on Educational Data Mining**, [S.l.], 2011.
- LIU, B. Sentiment analysis and subjectivity. **Handbook of natural language processing**, [S.l.], v. 2, p. 627–666, 2010.
- LONGO, C. R. J. Educação a Distância: o estado da arte. In: \_\_\_\_\_. São Paulo: Pearson Education do Brasil, 2009.
- MACHADO, A. P.; FERREIRA, R.; BITTENCOURT, I. I.; ELIAS, E.; BRITO, P.; COSTA, E. Mineração de texto em Redes Sociais aplicada à Educação a Distância. **Colabor@ - A Revista Digital da CVA-RICESU**, [S.l.], v. 6, Julho 2010.
- MAIA, M. D. C.; MEIRELLES, F. D. S.; PELA, S. Análise dos Índices de Evasão nos Cursos Superiores a Distância do Brasil. , [S.l.], p. 1–11, 2004.
- MALIN, J.; MILLWARD, C.; SCHWARZ, H.; GOMEZ, F.; THROOP, D.; THRONESBERY, C. Linguistic text mining for problem reports. In: SYSTEMS, MAN AND CYBERNETICS, 2009. SMC 2009. IEEE INTERNATIONAL CONFERENCE ON, 2009. **Anais...** [S.l.: s.n.], 2009. p. 1578 –1583.
- MANHÃES, L. M. B.; CRUZ, S. M. S. da; COSTA, R. J. M.; ZAVALETA, J.; ZIMBRÃO, G. Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados. In: XXII SBIE - XVII WIE, 2011, Aracaju. **Anais...** [S.l.: s.n.], 2011. p. 150–159.
- MOORE, M.; KEARSLEY, G. **EDUCAÇÃO A DISTÂNCIA: uma visão integrada**. São Paulo: Thomson Learning, 2007.
- MOSTOW, J.; BECK, J. Some useful tactics to modify, map and mine data from intelligent tutors. **Natural Language Engineering**, [S.l.], v. 12, p. 195–208, 6 2006.
- OLADOKUN, V.; ADEBANJO, A.; CHARLES-OWABA, O. Predicting students' academic performance using artificial neural network: a case study of an engineering course. **The Pacific Journal of Science and Technology**, [S.l.], v. 9, n. 1, p. 72–79, 2008.
- RIGO, J. S.; CAZELLA, S.; CAMBRUZZI, W. Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2012, Curitiba. **Anais...** [S.l.: s.n.], 2012.
- RIGO, S. J.; CAMBRUZZI, W. L.; BARBOSA, J. L. V.; CAZELLA, S. Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação**, [S.l.], v. 22, p. 132–146, 2014.
- ROMERO, C.; VENTURA, S. Educational data mining: a review of the state of the art. **Trans. Sys. Man Cyber Part C**, Piscataway, NJ, USA, v. 40, n. 6, p. 601–618, 2010.
- ROMERO, C.; VENTURA, S. Data mining in education. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [S.l.], v. 3, n. 1, p. 12–27, 2013.
- SAP AG. **Standardized Technical Architecture Modeling**. 2007. Acessado em 14/01/2014 [http://www.fmc-modeling.org/download/fmc-and-tam/SAP-TAM\\_Standard.pdf](http://www.fmc-modeling.org/download/fmc-and-tam/SAP-TAM_Standard.pdf).

SCHOONENBOOM, J.; HELLER, K.; NEENOY, M.; LEVENE, M. Trails in Education: technologies that support navigational learning. In: \_\_\_\_\_. . Rotterdam, The Netherlands: Sense Publisher, 2007.

SCOTT, J. **Distance Education Report, California Community Colleges Chancellor's Office.** Acessado em: 22/11/2012, Disponível em: [http://californiacommunitycolleges.cccco.edu/Portals/0/reportsTB/DistanceEducation2011\\_final.pdf](http://californiacommunitycolleges.cccco.edu/Portals/0/reportsTB/DistanceEducation2011_final.pdf).

SIEMENS, G.; BAKER, R. S. J. d. Learning analytics and educational data mining: towards communication and collaboration. In: INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 2., 2012, New York, NY, USA. **Proceedings...** ACM, 2012. p. 252–254. (LAK '12).

SIEMENS, G.; LONG, P. Penetrating the fog: analytics in learning and education. **Educause Review**, [S.l.], v. 46, n. 5, p. 30–32, 2011.

SILVA FILHO, R. L. L. e.; MOTEJUNAS, P. R.; HIPÓLITO, O.; LOBO, M. B. d. C. M. A evasão no Ensino Superior Brasileiro. **Cadernos de Pesquisa**, [S.l.], v. 37, p. 641 – 659, 12 2007.

SILVA, J.; ROSA, J. a.; BARBOSA, J.; BARBOSA, D. N. F.; PALAZZO, L. A. M. Distribuição de Conteúdo em Ambientes Cientes de Trilhas. **WebMedia'09**, [S.l.], 2009.

SMITH, A. D. **Who Controls the Past Controls the Future - Life Annotation in Principle and Practice.** 2008. Tese (Doutorado em Ciência da Computação) — University of Southampton, 2008.

SMITH, A.; HALL, W.; GLASER, H.; CARR, L. Towards Truly Ubiquitous Life Annotation. , [S.l.], 2008.

TATTERSALL, C.; JANSSEN, J.; BERG, B. Van den; KOPER, R. Modelling routes towards learning goals. **Campus-Wide Information Systems**, [S.l.], v. 23, n. 5, p. 312–324, 2006.

TINTO, V. Dropout from Higher Education: a theoretical synthesis of recent research. **Review of Educational Research**, [S.l.], v. 45, n. 1, p. 89–125, Jan. 1975.

TOSCHER, A.; JÄHRER, M. Collaborative filtering applied to educational data mining. **KDD Cup**, [S.l.], p. 1–11, 2010.

WANG, A. I.; SØRENSEN, C. fredrik; BREDE, S.; SERVOLD, H.; GIMRE, S. **DEVELOPMENT OF LOCATION-AWARE APPLICATIONS The Nidaros framework.** 2008.