Aplicações de Mineração de Dados Educacionais e *Learning Analytics* com foco na evasão escolar: oportunidades e desafios

Educational data mining and Learning analytics applications in dropout: opportunities and challenges

Sandro J. Rigo
Wagner Cambruzzi
Jorge L. V. Barbosa
Universidade do Vale do Rio dos Sinos - Unisinos
Programa Interdisciplinar de Pós-Graduação em Computação Aplicada
São Leopoldo, Rio Grande do Sul, Brasil
rigo@unisinos.br, wagner@cambruzi.com.br,
jbarbosa@unisinos.br

Sílvio C. Cazella
Universidade Federal de Ciências da Saúde de Porto
Alegre, Departamento de Educação e Informação em
Saúde
Porto Alegre, Rio Grande do Sul, Brasil
silvioc@ufcspa.edu.br

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. A maioria dos trabalhos sobre a implementação destas aplicações prioriza 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 com taxas de acerto na ordem de 87% e a realização de intervenções pedagógicas, com obtenção de taxas médias de 11% na redução da evasão.

Palavras-Chave: Mineração de dados educacionais, Learning Analytics, evasão escolar.

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 with hit rates of predictions around 87% and the realization of pedagogical actions, to obtain average rates of 11% in reducing evasion.

Keywords: Educational Data Mining, Learning Analytics, drop out.

Recebido: 26 de Setembro de 2013 / Aceito: 8 de Março de 2014

DOI: 10.5753/RBIE.2014.22.01.132

1 Introdução

Este artigo aborda o problema relacionado com evasão escolar, dado que, segundo diversos estudos [1, 2, 6,
11], este é um desafio a ser superado na área da Educação. Associado com múltiplos fatores e também relacionado estreitamente com peculiaridades de áreas de conhecimento, níveis de ensino e metodologias específicas,
este assunto vem sendo abordado com a aplicação de
técnicas de Mineração de Dados Educacionais (MDE) [3,
10, 19] e com o uso de recursos de *Learning Analytics*(LA) [31, 32, 33]. Estas abordagens podem ser aplicadas
para gerar subsídios que apoiam a identificação precoce
de alunos dentro de um perfil associado com evasão escolar. De modo complementar, podem ser utilizadas para o
acompanhamento e visualização de diversos aspectos
fundamentais do processo de ensino e aprendizagem.

Este quadro de possibilidades vem se constituindo de forma bastante consistente, pois alguns recursos tecnológicos disponíveis atualmente podem tanto representar um elemento de apoio no processo de ensino-aprendizagem, bem como constituir fontes de informações para o contínuo acompanhamento de resultados obtidos pelos alunos e tendências a serem tratadas, alimentando assim sistemas de MDE e LA [42, 43, 44]. Alguns exemplos deste contexto podem ser encontrados em ambientes digitais de apoio ao processo de ensino aprendizagem ou nos diversos formatos de materiais instrucionais e objetos de aprendizagem, bem como o uso de dispositivos móveis e computação ubíqua [25, 26]. Estes recursos favorecem paradigmas associados com uma maior interação e colaboração entre estudantes e docentes, favorecendo a autonomia dos estudantes e ampliando a flexibilidade destes para o acesso, nos momentos de maior necessidade, aos materiais considerados mais adequados. Ao mesmo tempo, a sua utilização é passível de acompanhamento e integração com outras fontes de informação para a geração de modelos e indicadores que servem de base para ações de gestão e intervenções pedagógicas.

Em boa parte das iniciativas de sistemas de MDE e LA são observados resultados considerados satisfatórios, porém dentro de um escopo de avaliação restrito aos conjuntos de atributos tratados, que muitas vezes são escolhidos de acordo com critérios que privilegiam mais aspectos técnicos, tais como a disponibilidade de atributos em bases de dados ou o interesse em técnicas específicas [3, 14, 19]. Porém a sua utilização em casos de acompanhamento em ambientes mais amplos demanda uma análise mais abrangente, que envolva equipes multidisciplinares normalmente envolvidas no processo de ensino e aprendizagem. Desta forma, aspectos adicionais podem ser destacados, tais como as interações com colegas e docentes ou as especificidades de materiais instru-

cionais e ferramentas de apoio utilizadas nas aulas.

O objetivo deste trabalho constitui-se no estudo de fatores envolvidos no fenômeno de evasão escolar, descrição e consequente análise da utilização de um sistema para MDE e LA em um contexto de cursos de graduação na modalidade de Educação a Distância. O diferencial do sistema em questão pode ser localizado na ampliação de possibilidades tradicionalmente observadas de utilização dos dados gerados no processo de mediação digital destes cursos acompanhados. O trabalho descreve e justifica a necessidade de uma ampliação do processo de análise dos fatores tradicionalmente monitorados e utilizados em iniciativas de MDE e LA, com a consideração de elementos associados ao papel exercido pelos docentes e pelo conjunto de aspectos metodológicos selecionados em cada situação. Desta forma, constitui-se em um estudo que busca avaliar não apenas os aspectos técnicos e a viabilidade das ferramentas atualmente disponíveis, mas justamente a sua utilização de forma integrada ao contexto metodológico de uma instituição de ensino superior.

Em decorrência desta abordagem, foi desenvolvido um sistema que atua na facilitação de coleta e integração de dados de interação de alunos e que disponibiliza opções de visualizações destes dados e de identificação precoce e dinâmica de perfis de alunos com tendências para a evasão ou o baixo rendimento [30]. Estas opções são utilizados para destacar e orientar ações pedagógicas eficientes, as quais, por sua vez, devem ser monitoradas de forma que seus resultados sejam também incluídos como elementos fundamentais de um processo mais amplo e que possibilite a coleta e análise de informações de modo continuado.

Portanto, este trabalho apresenta o estudo de aspectos da evasão escolar e descreve a aplicação e os resultados de uso de sistemas de MDE e LA em uma experiência prática de aplicação de recursos tecnológicos em busca do tratamento do problema de evasão escolar. Um dos elementos principais desta experiência foi a necessidade de ampla interação com equipes multidisciplinares para a delimitação de melhores estratégias de utilização destas tecnologias para a mitigação do problema da evasão. Esta interação mostrou-se indispensável em todas as etapas do processo, sejam elas de análise e planejamento, bem como de execução e avaliação.

Este artigo está estruturado da forma descrita a seguir. Na seção 2 são identificados elementos do problema de evasão escolar, sendo descritas em maiores detalhes algumas das suas implicações na Educação. A seção 3 discute abordagens para solução do problema da evasão escolar utilizando Mineração de Dados Educacionais, bem como aspectos importantes a serem ampliados neste contexto. A seção 4 dedica-se a conceituar e exemplificar iniciativas de *Learning Analytics*. Na seção 5 são analisa-



dos trabalhos relacionados e na seção 6 está descrita a abordagem sugerida neste trabalho, bem como o sistema utilizado. Os resultados de um estudo de caso com a abordagem sugerida estão descritos na seção 7. No final do artigo a seção 8 apresenta uma conclusão do estudo e uma discussão sobre trabalhos futuros.

2. Contextualização da evasão escolar

A seguir, considerações acerca do significado comumente utilizado para o termo evasão escolar são descritas e também são apresentados alguns aspectos históricos que justificam a importância de seu estudo e de realizações de ações de mitigação.

2.1 Definições e índices de acompanhamento

No Brasil, já em 1995 registra-se o início do desenvolvimento de estudo amplo, fomentado pelo MEC [16] e conduzido por Comissão Especial de Estudos de Evasão para reunir um conjunto de dados sobre o desempenho das universidades públicas brasileiras em relação aos índices de diplomação, retenção e evasão dos estudantes nos seus cursos de graduação. Em 2007, dentro do âmbito do REUNI (Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais), relata-se a existência de índices de evasão importantes. Como consequência da verificação destes índices elevados de evasão, fazem parte do documento gerado nesta ocasião [17] destaques específicos estabelecendo e recomendando metas para redução da evasão.

Informações relacionadas com a modalidade de Educação a Distância, geradas anualmente, também corroboram a existência de ocorrências importantes de evasão, tanto no Brasil [4] como em outros países [20, 11]. Dentre outras organizações, a *Organisation for Economic Cooperation and Development* (OECD) mantém dados históricos que permitem verificar que o problema da evasão apresenta-se de forma importante em grande número de países. Portanto, pode ser depreendido destas ações e relatos que o fenômeno da evasão escolar não está delimitado a uma determinada modalidade de educação apenas.

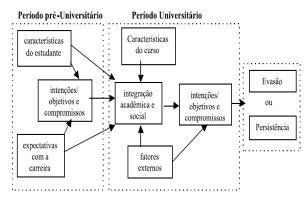
O termo evasão escolar permite diversas interpretações e é utilizado em diferentes contextos com significados ligeiramente diversos. Também são diversificadas as iniciativas para a definição e aplicação de índices que permitem realizar comparações e avaliações a seu respeito. Em alguns casos considera-se como evasão a desistência do curso pelo estudante, independentemente da quantidade de participações efetuadas [6]. Esta noção é corroborada em [37], considerando-se evasão o simples cancelamento do vínculo entre o aluno e a instituição, antes do término de todo o processo previsto em um de-

terminado curso para a sua conclusão.

Em outras situações diferencia-se evasão de acordo com períodos médios para conclusão de curso e períodos anuais [12]. Porém também são consideradas como evasão as ocorrências de simples interrupção do ciclo de estudos, independentemente do nível em que estes estudos estejam localizados, bem como do tempo em que esta interrupção dure [39]. Alguns autores identificam como evasão as situações de desistência definitiva após determinado contato com o curso [38], sendo que em muitos casos termos como perda ou fuga são considerados como sinônimos do termo evasão.

A busca de referenciais teóricos que expliquem as questões relacionadas a evasão pode ser identificada em trabalhos diversos [1, 22] 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. A Figura 1 resume este modelo, que destaca deste modo tanto a influência de aspectos pessoais e sociais existentes antes do ingresso no curso universitário, como aspectos relacionados com o contato acadêmico, metodologia de aprendizagem e integração institucional.



Contexto social e institucional

Figura 1: Modelo da relação aprendizagem/permanência, adaptado pelo autor, com base em Tinto [22]

No modelo apresentado na Figura 1 são relacionados aspectos que são independentes do contexto educacional e das ações realizadas em sala de aula, tais como as expectativas do estudante para com a sua carreira ou mesmo características socioeconômicas. Porém são também iden-

tificadas outras peculiaridades, tais como as características do curso sendo realizado e a interação entre as particularidades do aluno em relação a este contexto educacional. Estes modelos gerais permitem que sejam expressas algumas características gerais associadas com o fenômeno da evasão, porém dentro de uma necessidade de relativização constante.

Uma forma de acompanhamento da evasão, independente de modelos gerais de análise, é a manutenção de índices de ocorrência. A utilização de índices para medição de evasão [2], 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 identificação do número de matriculados entre semestres subsequentes, também denominada de evasão imediata, não permite medir as situações em que temporariamente o aluno suspende sua participação no curso. Como exemplo, pode ser considerado o índice utilizado para cálculo da evasão pelo MEC, capaz de medir a evasão verificada em um curso, entre semestres subsequentes (denominada de evasão imediata). Este índice está descrito na Equação 1 e relaciona a quantidade de alunos com matrícula, ingressantes e reingressantes em um período semestral, com a quantidade de alunos matriculados e com a quantidade de formados neste período anterior. Na equação 1 os termos M_n e M_{n-1} representam os alunos com matrículas no período e no período anterior. Já o termo I_n representa os ingressantes no período e R_n indica o número de reingressantes no período. O termo F_{n-1} indica a quantidade de formandos no período anterior.

$$I_{MEC} = \left(1 - \left(\frac{M_n - I_n - R_n}{M_{n-1} - F_{n-1}}\right)\right) x 100 \quad (1)$$

Deve ser observado que os resultados indicados por este índice da Equação 1 sempre descrevem uma situação já bem delimitada, sendo que sua aplicação pode ser importante em casos de acompanhamento de séries históricas de eventos. Entretanto, os resultados deste tipo de índice não são amplamente aplicáveis em contextos preventivos para o fenômeno da evasão.

2.2 Estudos de evasão no Ensino superior

Os estudos realizados periodicamente por autarquias como o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) ou o Ministério de Educação e Cultura (MEC), além de estudos realizados eventualmente dentro do âmbito de instituições, permitem o acompanhamento de indicadores gerais, bem como a identificação de fatores regionais e sociais ligados à evasão. Cabe salientar que estes estudos já são realizados e interpretados de modo relativo, compondo análises que levam em conta os indicadores e demandas sociais, ou então aspectos econômicos e tendências associadas com algumas áreas específicas [2].

Alguns destes estudos relacionam também aspectos mais específicos, ligados às características pessoais e institucionais. Estes estudos destacam adicionalmente 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. Aliam-se a estes fatores aspectos pessoais e de envolvimento e desempenho dependentes do aluno ingressante.

No estudo apresentado por Vitelli, Fritsch e Rocha [40] alguns destes aspectos são destacados, sendo associados com categorias de fatores mais gerais, tais como fatores sociais, fatores econômicos, fatores de desempenho e fatores de escolha. Com relação a fatores sociais foram identificados no estudo aspectos relevantes associados à idade do aluno ingressante nos cursos, bem como o formato de ingresso. Já em relação aos aspectos financeiros o estudo mostra que a quantidade média de disciplinas cursadas e a existência ou não de algum tipo de apoio financeiro, tal como bolsas de estudou gratuidades, estão associados de forma importante as ocorrências de evasão. Já em relação aos fatores de desempenho são descritas várias situações em que observa-se uma correlação importante entre a sua ocorrência e a ocorrência de evasão, tal como nos casos de baixa média no vestibular, baixa taxa de aprovação em disciplinas matriculadas (menos de 40%), média de desempenho inferior a três nas disciplinas matriculadas e uma taxa de cancelamento igual ou acima de 20% em disciplinas.

Em alguns contextos são realizados periodicamente estudos e coletas de dados para manutenção de índices de ocorrência de evasão. Por exemplo, no Brasil a Associação Brasileira de Educação a Distância (ABED) mantém um censo anual para a coleta destes dados entre instituições nacionais. Nesta coleta, nos últimos anos observase que na maioria das instituições constata-se que houve aumento no volume de matrículas.



Entretanto a evasão média nos cursos de EAD, informada pelas instituições para este censo, localiza-se em torno de 20%, sendo menor em disciplinas obrigatórias, com 17,6%, e maior nos cursos livres, com 23,6% [4]. Em 2004 no estudo apresentado por Maia, Meireles e Pela [35], onde foram analisadas 22 instituições brasileiras e 51 mil alunos, os índices de evasão foram em torno de 30%. Desta forma, os altos índices de evasão apresentados na EAD são uma preocupação presente no Brasil. Entretanto esta preocupação é pertinente a outros países também [20, 34]. O trabalho de Longo [36] apresenta índices internacionais com valores acima de 65% de evasão.

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 de tendências pode ser feita com a utilização de informações que apresentam, em geral, pouca variação ao longo de períodos de tempo, sendo compostas, por exemplo, pelo histórico social, motivacional e educacional do aluno. Já em outras situações as informações apresentam-se com características bastante dinâmicas, tal como observa-se em relação à competências docentes ou interação e colaboração discente ao longo de um semestre letivo.

Os recursos de MDE e LA 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.

3. Mineração de Dados Educacionais

O avanço das Tecnologias de Informação e Comunicação possibilitou um aumento substancial na quantidade de dados gerados e disponibilizados. Atualmente a capacidade de geração de conjuntos de dados é muito maior do que a capacidade dos pesquisadores e analistas de fazer análises sobre os dados armazenados, tendo em vista limitações de processamento ou outras restrições associadas com aspectos destes conjuntos de dados, que em boa parte são dados não estruturados. Estes aspectos que apresentam por si só diversos desafios ao tratamento destes conjuntos de dados.

O mesmo cenário de grande disponibilidade de conjuntos de dados diversos é observado também na área da educação [3]. Com a ampla difusão do uso de sistemas

informatizados nas escolas e universidades, cresce a cada dia o volume de dados gerados e armazenados em bases de dados. Alguns exemplos destes fatores são as Tecnologias de Informação e Comunicação, o suporte computacional para Ambientes de Aprendizagem, a consolidação da modalidade de Educação a Distância e do *Blended Learning*, a incorporação de Sistemas Integrados de Gestão em instituições educacionais, as comunidades virtuais e suas ferramentas de relacionamento, compartilhamento e publicação *on-line* [30].

Este grande volume de dados tem fomentado o interesse na sua utilização, junto com técnicas de Mineração de Dados, na busca de respostas para perguntas específicas da Educação, relacionadas com processos de aprendizagem, desenvolvimento de materiais instrucionais, acompanhamento e previsões, entre outros [3, 14]. Estas iniciativas de utilização em geral são observadas a partir da estruturação de processos para a obtenção de informações e para posterior geração de padrões de comportamento que possam apresentar aspectos importantes para apoiar determinadas práticas pedagógicas.

A avaliação da importância de sua utilização, em conjunto com recursos de LA, vem crescendo continuamente, o que pode ser constatado pelo interesse expresso não apenas por universidades, mas por instituições [42] e governos [41], além de projetos de pesquisa e pela criação de sociedades destinadas a fomentar o desenvolvimento e pesquisa neste campo, tais como a International **Educational** Data Mining (http://www.educationaldatamining.org), ou então a Socifor Learning Analytics Research (http://www.solaresearch.org/).

Quanto a Mineração de Dados, esta pode ser entendida como uma das etapas do processo de Descoberta de Conhecimento em Base de Dados (DCBD), o qual se utiliza de algoritmos específicos para a extração de padrões destas bases de dados [24]. O objetivo principal da etapa de Mineração de Dados constitui-se em extrair conhecimento implícito em bases de dados [18], através de diversas tarefas e com com objetivos, em geral, de descrever ou então de prever a ocorrência de padrões úteis para atividades associadas.

O processo de DCBD pode ser definido como sendo o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis presentes nos dados sendo analisados e tratados [24]. A Figura 2 apresenta resumidamente as etapas envolvidas no processo de DCBD:

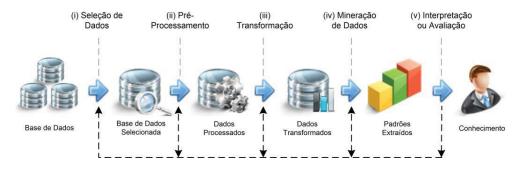


Figura 2: Etapas do processo de DCBD adaptado pelo autor, a partir de [7]

- i) Seleção de Dados: A partir do entendimento do domínio da aplicação, é realizada a seleção de um conjunto de dados de acordo com os objetivos do processo;
- ii) Pré-Processamento: Os dados disponíveis normalmente não estão em um formato preparado para a análise final, podem apresentar inconformidades tais como: duplicidade, falta de consistência, ruídos, entre outros. Dessa forma torna-se necessária a aplicação de métodos para tratamento, limpeza e redução de volume de dados, para que os dados selecionados estejam em perfeitas condições para as próximas etapas de análise;
- iii) Transformação: Nesta etapa, os dados já préprocessados, serão transformados para que possam ser utilizados nos algoritmos de extração de padrões, conforme a tarefa e objetivo da mineração de dados;
- iv) Mineração de Dados: Esta etapa consiste (uma vez que o objetivo da mineração e tarefa tenham sido definidos) na aplicação do algoritmo de mineração de dados, que é responsável pela extração do conhecimento implícito no conjunto de dados selecionado. Esta etapa caracteriza-se pela descoberta de conhecimento e padrões propriamente dito;
- v) Interpretação ou Avaliação: Nesta etapa o conhecimento extraído pela etapa de mineração de dados é analisado quanto a conhecimento útil e novo, e caso caracterize-se como tal, poderá ser utilizado no suporte ao processo de tomada de decisão na área de domínio da aplicação.

As tarefas de mineração de dados, segundo a literatura [7], são as tarefas de associação, classificação, agrupamento e a sumarização. Quanto ao objetivo da tarefa de mineração de dados, estas podem ser:

- 1) Descritiva: Objetiva gerar padrões descritivos através da avaliação do comportamento dos dados [7];
- 2) Preditiva: Utiliza algumas variáveis da base de dados como atributos, para prever valores desconhecidos ou futuros de outras variáveis de interesse [7].

A Mineração de Dados Educacionais trata da aplicação das 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, [19]. Alguns exemplos de aplicação destas técnicas são a geração de alertas [8], o apoio a sistemas de recomendação [5, 23] ou a captura de perfis de estudantes [9].

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 educacionais 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 educativas. 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.

4. Learning Analytics

Mesmo possuindo uma grande quantidade de dados sobre os estudantes, as instituições de ensino superior têm sido tradicionalmente ineficientes no uso desses dados, muitas vezes realizando análises com atrasos substanciais, retardando ações e perdendo oportunidades de intervenções [31]. Isso se deve ao fato que o crescimento contí-

nuo 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 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 cenário define implicações positivas para a educação em geral, sendo enfaticamente destacado em diversos estudos na área, tais como o Horizon Report (New Media Consortium) [42], ou relatórios do Office of Educational *Technology (US Department of Education)* [41, 43].

Esse contexto tem promovido um crescente interesse na análise automática dos dados que recentemente foi referenciada como Learning Analytics (LA) [32, 33]. Cabe ressaltar que LA 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 do processo de ensino e aprendizagem. A relação entre LA e as áreas de pesquisa correlatas foram referenciadas no trabalho de Chatti [32] e incluem Learning Analytics, action research, Mineração de Dados Educacionais, sistemas de recomendação e aprendizagem personalizada ou 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. O seu crescente interesse pode ser observado em diversos projetos e eventos destacados a partir da recentemente criada Society for Learning Analytics Research (http://www.solaresearch.org).

Chatti et al. [32] propuseram um modelo de referência para LA baseado em quatro dimensões, com o objetivo de identificar o conjunto completo de elementos necessários para construir sistemas a partir da abordagem de análise em LA. Estas quatro dimensões seriam o tipo de dados coletados, o público-alvo da análise, o objetivo da análise dos dados e , por fim, a técnica utilizada. Já no trabalho de Greller e Drachsler [33] são analisadas questões que devem ser consideradas para minimizar desafios em potencial e permitir uma exploração benéfica de dados edu-

cacionais. Neste trabalho [33] é proposto 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 melhorar a eficiência e eficácia do professor. Segundo Ferguson [44], o desenvolvimento e adoção dos recursos de LA (e também dos recursos de MDE), passa por alguns desafios importantes, descritos em quatro grandes grupos, que seriam: a ampliação da conexão com ciências do aprendizado, tais como a cognição, a meta-cognição e a pedagogia; a necessidade de desenvolver métodos para tratamento de maiores conjuntos de dados e dados mais diversos, tais como dados de computação móvel, dados biométricos e dados descrevendo emoção ou humor; a atuação na perspectiva do aluno e do aprendiz; o desenvolvimento de um conjunto ético de premissas para guiara a utilização destes recursos de forma adequada.

5. Trabalhos relacionados

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 aspectos motivadores para ações de evasão escolar são bastante diversas e envolvem fatores tanto externos como internos, aspectos pessoais e institucionais. Alguns destes fatores podem ser identificados com apoio de sistemas de MDE e LA, proporcionando informações para a realização de ações de prevenção e de atendimento.

Diversos trabalhos vêm sendo realizados com o objetivo de identificação de tendências de evasão [5, 8, 9, 14, 15, 23]. Entretanto, o que se observa, de modo geral, é que este conjunto de trabalhos possui em comum o objetivo da experimentação de técnicas para identificação de aspectos que possam ser monitorados e que permitam gerar informações relevantes para a seleção dos alunos com probabilidade de evasão ou para tarefas auxiliares neste sentido. Alguns dos exemplos acima tratam do uso de MDE e posterior geração de alertas [8], ou o uso de recursos para o apoio a sistemas de recomendação [5, 23] ou ainda o tratamento de dados diversos para a captura de perfis de estudantes [9] e posterior relacionamento com evasão. Alguns aspectos podem ser considerados ainda incipientes e ainda sendo investigados, tais como a integração e relacionamento de dados, podem ser tratados com recursos semânticos integrados às técnicas de mineração de dados [10]. 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 [13], ou então, dados textuais em redes sociais também podem ser utilizados a partir de técnicas de mineração [15].

Alguns trabalhos resumem aspectos gerais das áreas de MDE e LA, apresentando uma discussão sobre suas características e aplicações [3, 19, 45, 47, 48, 49]. Nestes trabalhos podem ser encontradas análises específicas sobre problemas técnicos associados com o uso de recursos de MDE ou LA, bem como a evolução das formas de utilização destes recursos no apoio à educação. Alguns trabalhos vem sendo apresentados com a proposição de modelos gerais e frameworks, ou a discussão de desafios destas áreas [32, 33, 44, 46]. Nestes trabalhos são apresentadas discussões demandando uma análise mais abrangente sobre os conjuntos de dados a serem utilizados e sua finalidade. Entretanto não são observados estudos mais amplos e sistemáticos sobre a utilização destes recursos de acordo com políticas institucionais e relacionados com escolhas de gestão ou intervenções pedagógicas.

Portanto, identifica-se a necessidade de uma abordagem que integre de forma mais ampla o estudo prévio dos fatores a serem monitorados por técnicas de MDE e LA, 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 MDE pode ser tanto mais efetiva quanto maior for o envolvimento de uma equipe multidisciplinar, que garanta a melhor utilização do conhecimento existente acerca deste fenômeno. Embora a MDE tenha o potencial de identificar padrões relevantes, é 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 [19].

O mesmo é válido para o caso de uso de recursos de LA, que podem perder sua efetividade quando utilizados sem o adequado conhecimento dos aspectos gerais associados com a evasão escolar. Grande parte dos trabalhos aqui citados está descrevendo uma etapa onde o foco principal de atenção está relacionado com a geração adequada e correta dos resultados pelos sistemas de MDE e LA. Esta etapa pode ser considerada como uma etapa inicial importante, porém a partir do momento em que estes resultados são disponibilizados, entra em questão a utilização destas informações, que idealmente devem estar associadas com ações afirmativas institucionais, tanto no aspecto de gestão educacional como no aspecto de iniciativas pedagógicas. Um dos exemplos conhecidos e que surge com grande frequência nos trabalhos citados é o uso destas informações para a geração de alertas a serem utilizados pelos professores, que permitem que o professor deixe de ser reativo, respondendo ao aluno

apenas quando solicitado, e passe a ser proativo, ampliando as possibilidades de apoio e instigação no processo de aprendizagem do aluno.

Um dos diferenciais do trabalho aqui apresentado para com os demais trabalhos citados está relacionado com a implementação de um sistema que apresenta grande flexibilidade nas etapas de aquisição, de tratamento e de utilização dos dados originados na mediação digital. Esta característica possibilita a utilização de dados de fontes e dispositivos diversos, compondo um conjunto rico de dados sobre os alunos. Além disso, este aspecto do sistema torna viável a disponibilização de diferentes serviços de interesse para os professores, tais como serviços de visualização de dados, relatórios ou predições. Outro diferencial deste trabalho foi a aproximação com as diversas equipes institucionais envolvidas com o fenômeno da evasão e com a ampla utilização do sistema desenvolvido em um estudo de caso, durante um período de 3 semestres, para o apoio à implementação de ações de mitigação da evasão. A flexibilidade na utilização dos dados proporcionou a implementação de um conjunto de opções de interesse para as equipes de educadores envolvidos com o uso do sistema.

6. Sistema desenvolvido

Esse artigo descreve os resultados de um projeto de pesquisa em parceria com uma empresa de tecnologia educacional, culminando no desenvolvimento de um sistema que possibilite a descoberta de conhecimento em bases de dados a partir de técnicas de MDE e utilize esse conhecimento para identificar tendências de enquadramento de perfis de alunos, juntamente com recursos de LA para possibilitar ao corpo docente e gestor o acesso à informações de acompanhamento sobre o processo de ensino-aprendizagem. Essas tendências são informadas aos envolvidos para que possam ser tratadas e terem seu impacto minimizado, principalmente em situações que podem levar os estudantes a contextos de evasão imediata.

O sistema contempla a implementação de um processo com as seguintes etapas: descoberta do conhecimento, registro de padrões de interesse, identificação de tendências conforme os padrões descobertos, aviso aos envolvidos, registro das ações realizadas e resultados obtidos, visualização de aspectos de interação e acompanhamento de informações textuais. A arquitetura e os pressupostos do sistema estão descritos em trabalho relacionado [30]. Entretanto cabem ser ressaltados aqui alguns destes pressupostos. O modelo utilizado visa obter duas vantagens relacionadas com o uso de dados educacionais. 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 descre-

vendo aspectos específicos. A segunda é a flexibilidade para utilização dos dados, já que aplicações com necessidades distintas podem ter apenas os conjuntos de dados de interesse consultados. 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. A implementação deste modelo foi realizada com a linguagem ASP.NET C# para os componentes dedicados à coleta, manipulação dos dados, visualização de dados e de mineração. Para a persistência dos dados foi utilizado o banco de dados SQL Server 2008.

A arquitetura do sistema desenvolvido foi organizada em uma camada denominada camada servidor, uma camada denominada cliente e outra dedicada aos dispositivos de coleta ou fontes de coleta de dados. De forma resumida, a última camada está destinada à acomodar a diversidade de dispositivos de interação ou fontes de dados de interesse. Tipicamente esta camada permite a coleta de dados de Ambientes Virtuais de Ensino e Aprendizagem (AVEA) ou a coleta de dados gerados pela interação de um aluno através de um dispositivo móvel. Cada diferente fonte de dados ou dispositivo de interação possui um componente na segunda camada (a camada cliente) que é responsável pelo atendimento de solicitações originadas pela interação do usuário ou pela consulta periódica à fontes de dados cadastradas, tal como no caso de acessos a um AVEA. Por fim, a camada servidor possui como objetivos armazenar os dados obtidos das diferentes fontes e disponibilizar diferentes tipos de serviços, para que seja bastante ampla a possibilidade de uso dos dados neste sistema.

Alguns exemplos destas possibilidades são os serviços de visualização de dados, que permitem o relacionamento de diferentes aspectos, tais como as interações em ambientes de discussão e os resultados acadêmicos. Outro exemplo são os serviços de tratamento dos dados textuais, para identificação de necessidades expressas em mensagens individuais. O tratamento de dados para a identificação de perfis cognitivos e a personalização de mensagens e conteúdos instrucionais de acordo com dispositivos sendo utilizados pelos alunos completam este quadro de possibilidades, juntamente com os serviços de predição de evasão com base em aspectos de interação do aluno no AVEA. Alguns destes exemplos de possibilidades serão descritos a seguir.

A Figura 3 descreve um exemplo de utilização das informações geradas pela análise integrada de mensagens de fóruns de discussão em um AVEA e resultados obtidos por alunos em disciplinas de graduação. Este tipo de

visualização integra portanto duas dimensões presentes no processo de ensino-aprendizagem do aluno, que são os resultados obtidos com as notas de avaliações e as interações realizadas dentro do contexto do ambiente de relacionamento, neste caso, em especial, dentro do fórum e discussão. A Figura 3 não inclui a legenda das cores utilizadas, mas as cores com tom marrom representam as notas de valor mais baixo, enquanto que as cores de tom verde apresentam as melhores notas. A espessura dos arcos deste grafo representa a quantidade de interações realizadas. Esta ferramenta permite que os professores possam acessar dados detalhados de cada estudante, representados estes pelos nodos do grafo. Também pode ser utilizada esta ferramenta para a realização de avaliações gerais sobre o andamento de turmas.

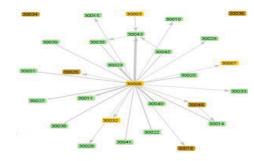


Figura 3: Visualização de interações em um fórum de discussões

A Figura 4 ilustra outra opção de acompanhamento geral de turmas, que consiste em uma ferramenta de análise textual de mensagens postadas em diferentes ferramentas de AVEAs, tais como fórum, chat ou diário, entre outras. Cada uma destas mensagens postadas pelos alunos pode ser analisada para a identificação de frases contendo a expressão de dúvidas ou dificuldades percebidas pelo aluno ao longo do processo. Além disso são detectados os tópicos gerais destas dúvidas, que são então exibidos de acordo coma estrutura de conhecimento tratadas em uma determinada disciplinas. Na Figura 4 abaixo, cada elemento representa um tópico sendo tratado na disciplinas e o valor em vermelho descreve a quantidade de ocorrências de menções à dúvidas para este tópico.



Figura 4: Visualização de tópicos com dúvidas de alunos

Outras opções disponibilizadas no sistema estão voltadas para a predição de comportamentos associados com evasão ou baixo rendimento. Neste caso existe uma integração ampla com a metodologia adotada institucional-

mente e com expectativas de gestão e pedagógicas, que são destacadas aqui como um aspecto diferenciado deste sistema e do estudo de caso que será descrito na seção 7, para com outros sistemas de MDE e LA conhecidos.

Para a aplicação de técnicas de Mineração de Dados para a predição de evasão e baixo rendimento, a técnica de Redes Neurais Artificias (RNAs) foi escolhida por ser uma técnica clássica na área e apresentar um desempenho satisfatório em outros problemas similares de mineração de dados [27, 28]. Este artigo não objetiva aprofundar análises específicas sobre este aspecto, mas são comentados a seguir algumas informações relacionadas.

Por ser adotada uma técnica de aprendizado de máquina supervisionado [29], 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. Para o estudo de caso relatado na seção 7 foi utilizada a técnica de RNA *Multilayer Perceptron*, com o algoritmo *Backpropagation* [50].

As variáveis utilizadas para a RNA são todas relacionadas com a interação e resultados obtidos no AVEA, tais como a quantidade de participações em fóruns e chats, o tempo de acesso, a quantidade de acessos total e resultados de questionários e avaliações. Como os dados são originados no AVEA utilizado, não foram realizadas etapas significativas de pré-processamento ou transformações.

A Figura 5 apresenta um quadro de controle no qual estão representadas as opções de acompanhamento e de realização de ações para o caso específico de predição de evasão e de baixo rendimento. Este quadro é disponibilizado para os professores de disciplinas em EAD e contém todos os alunos de uma turma, com as seguintes informações: primeiro acesso ao AVEA, último acesso ao AVEA, notas parciais obtidas e a predição semanal de evasão e baixo rendimento. De acordo com estas informações, o professor pode visualizar com antecedência a probabilidade de ocorrência de situações indesejadas, tais como a evasão ou o baixo rendimento. Estas informações podem ser utilizadas para a realização de ações de gestão e de iniciativas de intervenção pedagógicas.

A oportunidade de realização de ações com antecedência é um dos diferenciais obtidos neste contexto de integração de recursos de MDE e LA com metodologias institucionais. A partir das políticas adotadas institucionalmente os professores possuem ao seu alcance a possibilidade de realizar contato com alunos que podem estar em situações de dificuldades. A expectativa existente é de que estas ações possam contribuir para com a mitigação de ocorrência de evasão.

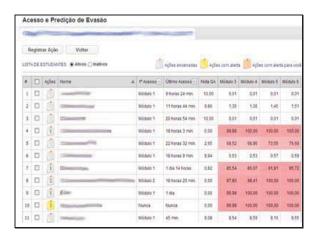


Figura 5: Quadro com predição e opções de ações sobre evasão

A Figura 6 apresenta uma janela para que seja formalizado o registro de ações realizadas com base nas indicações de predição de evasão. Deste modo, cada conjunto de ações realizadas é também parte das variáveis analisadas em etapas futuras, permitindo que sejam relacionadas situações, ações realizadas e os resultados obtidos. Nesta janela são registrados os indícios que motivaram a ação, seguidos de agrupamentos que permitem aproximar indicações de ações escolhidas, além de existir a possibilidade de detalhamento complementar o que pode representar uma fonte de informações importantes, tendo em vista que diversas das ações mais comumente realizadas envolvem algum tipo de retorno dos alunos, o que pode trazer elementos adicionais que auxiliem no entendimento da situações individual de cada aluno atendido. Portanto este conjunto de opções relacionadas com a predição da evasão e com o registro de ações permite o acompanhamento periódico sobre os resultados e a efetividade das ações.



Figura 6: Janela com registro de ações pedagógicas sobre evasão

Alguns diferenciais do sistema são destacados. Em primeiro lugar, destaca-se a ampla interação prevista entre a equipe técnica do projeto e os setores institucionais relacionados. Este aspecto está sendo vivenciado em um estudo de caso em andamento por três semestres letivos, envolvendo cursos de graduação e utilização de ambiente virtual de aprendizagem como uma das origens de dados para a detecção de perfis de evasão imediata. A colaboração institucional amplia a possibilidade de detec-

ção e encaminhamento de situações, sendo prevista uma etapa futura de integração de dados históricos e socioeconômicos. Um dos objetivos adicionais deste projeto está relacionado com a exploração de recursos considerados como fundamentais para ampliar as possibilidades de reconhecimento e comunicação de padrões significativos para o apoio a práticas pedagógicas, tais como recursos semânticos descritos a partir de ontologias, que permitem utilização mais ampla dos materiais coletados. Além disso, investiga-se também a utilização de informações linguísticas em consonância com recursos de mineração, como forma de melhor aproveitamento dos dados não estruturados textuais disponibilizados nos ambientes de estudo, destacando-se seu grande potencial de expressão de necessidades ou interesses dos alunos.

A partir dos resultados obtidos com a aplicação do conhecimento em turmas em andamento, deve ser tratada a questão de entrega e visualização destes dados, pois o público ao qual o sistema é destinado necessita receber informações de forma clara e sem sobrecarga, caso contrário tornaria o sistema pouco adequado. Os dispositivos usados para a visualização também devem ser estudados, para permitir aos usuários uma maior mobilidade e flexibilidade. Neste sentido surge a possibilidade de aplicação de técnicas da Computação Móvel e Ubíqua, principalmente o uso de estratégias baseadas na sensibilidade ao contexto. Assim, informações poderão ser disponibilizadas para os usuários na forma mais adequada e no momento mais apropriado.

7. Estudo de Caso

A validação da viabilidade do uso de MDE e de LA com base no sistema desenvolvido foi realizada em um estudo de caso objetivando a identificação de alunos com potencial risco de evasão escolar ou de baixo desempenho em cursos de graduação na modalidade de EAD. A seguir são descritas, na seção 7.1, as ações realizadas neste estudo de caso para a geração da Predição de evasão e baixo desempenho. Logo em seguida, na seção 7.2, são comentadas as ações de reversão de evasão e seus resultados.

7.1 Predição de Evasão e baixo desempenho

Este estudo consiste em ações de acompanhamento de dados prioritariamente ligados à interação de alunos e suas notas de avaliações gerados em um AVEA e sistema acadêmico utilizados por cursos de graduação em EAD realizados em uma Universidade do sul do Brasil. O período total compreende três semestres letivos, entre os anos de 2012 e 2013. O número total de alunos envolvidos no estudo de caso é de 2491.

A dinâmica adotada no estudo de caso considerou ini-

cialmente dois períodos em que foram realizadas ações de coleta e tratamento de dados, com a geração de predições de evasão e baixo rendimento. Isso foi realizado de forma que, ao final de cada semana fosse gerado um relatório indicando quais os alunos que foram classificados com um alto risco de evasão ou de baixo rendimento. Estes dois períodos serviram para validar os resultados obtidos e possibilitaram a realização de uma segunda ação, no terceiro período, que consiste em incluir os professores no contexto do uso da ferramenta, para que efetivamente fossem realizadas ações de mitigação de evasão e prevenção de baixo rendimento. A seguir são detalhados os resultados obtidos com a predição de evasão e baixo rendimento, para cada o período considerado no estudo de caso. Os resultados da reversão da evasão são observados e analisados na seção 7.2. Deste modo, no estudo de caso o objetivo principal adotado foi a identificação precoce de alunos que com possível comportamento de evasão ou baixo desempenho. Esta identificação precoce possibilita a realização de ações de prevenção de forma personalizada, a partir de intervenções pedagógicas e ações de gestão.

Na primeira etapa, realizada no primeiro semestre de 2012, foram acompanhadas 603 alunos em dez turmas de EAD nos cursos Administração, Sistemas de Informação e Gestão de Tecnologia em TI. Estas turmas correspondem à quarto disciplinas diferentes, sendo estas as disciplinas de Matemática para Computação, Matemática para Administração, Lógica, Laboratório I e Programação I. A média de acertos da predição de evasão e baixo rendimento foi de 76,45%, tendo sido considerado como critério para esta qualificação o fato do aluno não ter realizado o exame escrito final ou o exame escrito de recuperação, ou terem recebido nota zero nestes exames. A seguir são apresentados os gráficos detalhando a predição e sua média, ao longo das semanas de aula monitoradas. A Figura 7 descreve os resultados para a taxa de acerto da predição de evasão em uma das disciplinas, com acerto de uma média de 75,7% dos casos de evasão durante todo o período analisado.

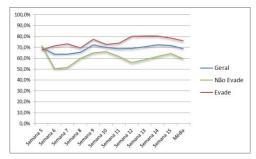


Figura 7: Resultados de predição de evasão – disciplina 1

A Figura 8 ilustra os resultados para a taxa de acerto da predição de evasão na segunda disciplina, onde se

verificou um acerto médio 81% dos casos de evasão.

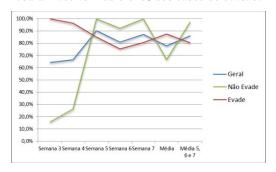


Figura 8: Resultados de predição de evasão – disciplina 2

A Figura 9 apresenta os resultados para a taxa de acerto da predição de evasão na terceira disciplina, com a obtenção de uma média de 72,6% de acerto dos casos de evasão durante todo o período analisado

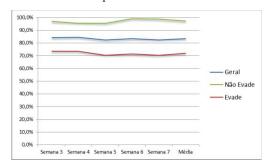


Figura 9: Resultados de predição de evasão - disciplina 3

A Figura 10 expõe os resultados para a taxa de acerto da predição de evasão na quarta disciplina, com média 76,5% dos casos de evasão.

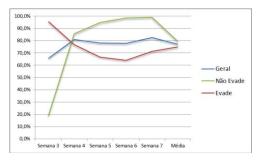


Figura 10: Resultados de predição de evasão - disciplina 4

Na segunda etapa foram acompanhadas quatro disciplinas, para um total de 250 alunos, com a obtenção de uma média de acerto na predição da evasão na ordem de 68,4%. A terceira etapa consistiu de experimentos onde foram acompanhados 925 alunos pertencentes a dezoito turmas e duas disciplinas de cursos de graduação em Educação a Distância. Neste experimento foi possível prever 83% dos alunos que realmente evadiram das disciplinas analisadas a partir da terceira semana de aula. Na quarta etapa foram acompanhadas 13 turmas, em um total

de 713 alunos. Neste último experimento a taxa de acerto da predição de evasão foi de 83,6%. A Figura 11 apresenta estes resultados de forma integrada.

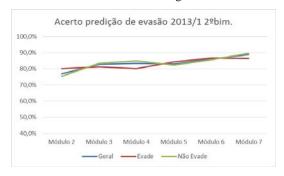


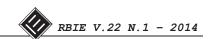
Figura 11: Resultados de predição de evasão quarta etapa

De forma a complementar o estudo de caso com um acompanhamento do potencial de uso dos relatórios gerados semanalmente, foi também realizado um trabalho de avaliação de possibilidades de reversão da evasão. Ele será descrito no item a seguir.

7.2 Reversão de Evasão

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 previsões relevantes para os comportamentos ligados à evasão e baixo rendimento, adotou-se um posicionamento de utilização desta informação a partir de ações pedagógicas com o objetivo de fomentar a reversão deste quadro. Este trabalho foi realizado com o uso de um componente da solução descrita no item anterior que complementa o módulo de predição de evasão e baixo desempenho. Este módulo adicional possui como objetivo suportar a implementação e monitoramento das ações pedagógicas a serem desencadeadas com o intuito de atender aos alunos com a classificação associada a um risco alto de evasão.

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 tinham acesso ao relatório com a predição de evasão e a partir desta informação desencadeavam ações específicas de acolhimento e atenção aos alunos. Estas ações peda-



gógicas possuem caráter proativo e podem variar bastante, de acordo com as necessidades identificadas e com o momento de desenvolvimento dos conteúdos.

Os estudos já realizados confirmaram que a realização de ações proativas podem evitar a evasão e melhorar o desempenho dos alunos. O gráfico mostrado na Figura 12 possibilita que seja feita uma análise das taxas de evasão de maneira comparativa entre turmas que participaram dos estudos e nas quais ações proativas foram realizadas e outras turmas que não participaram do estudo, pertencentes à mesma disciplina. Este estudo corresponde à terceira etapa descrita no item 7.1, com 925 alunos em duas disciplinas.

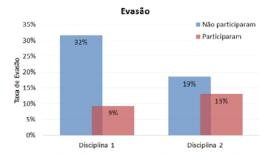


Figura 12: Resultados comparativos da reversão da evasão

Observa-se na Figura 12 que as taxas de evasão das turmas que participaram do estudo são mais baixas do que as taxas de evasão 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.

8. Considerações Finais

Neste artigo 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. Verifica-se que este problema está associado com um conjunto complexo de causas e situações. 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 Educação a distância e o Blended-learning. 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.

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 com a investigação dos ganhos possíveis com a integração de aspectos da análise textual com os demais aspectos já tratados pelo sistema.

De acordo com Ferguson [44], alguns desafios para as áreas de MDE e LA passam pela aproximação com as áreas de estudo da cognição e pedagogia, a definição de parâmetros éticos para uso ampliado destes recursos. Com este trabalho acredita-se ter sido reforçada a necessidade de ampliação de colaborações multidisciplinares, para que os recursos de predição, acompanhamento e visualização possam ser utilizados de forma integrada à metodologia e cultura institucional.

Referências

- [1] W. B. Andriola, C. G. Andriola, C. P. Moura. 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. Rio de Janeiro, volume 14, número 52, páginas 365-382, 2006.
- [2] A. A. C. T. Adachi. Evasão e evadidos nos cur-

- sos de graduação da Universidade Federal de Minas Gerais, Dissertação de Mestrado. Faculdade de Educação, UFMG, 2009.
- [3] R. S. J. Baker, A., M. J. Carvalho, S. Isotani. Mineração de Dados Educacionais: Oportunidades para o Brasil, Revista Brasileira de Informática na Educação, Volume 19, Número 2, 2011
- [4] Censo EAD.BR. Relatório analítico da aprendizagem a distância no Brasil, Associação Brasileira de Educação a Distância (ABED), Editora Pearson Education do Brasil, São Paulo, 2011.
- [5] G. Durand, F. Laplante, R. Kop. A Learning Design Recommendation System Based on Markov Decision Processes, KDD 2011 Workshop: Knowledge Discovery in Educational Data, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2011) in San Diego, CA. August 21-24. 2011.
- [6] R. V. M. Favero. Dialogar ou evadir: Eis a questão!: Um estudo sobre a permanência e a evasão na Educação a Distância, no Estado do Rio Grande do Sul. CINTED-UFRGS. Novas Tecnologias na Educação. Porto Alegre: V. 4 N° 2, Dezembro, 2006.
- [7] U. M. Fayyad, G. P. Shapiro, P. Smyth, R. Uthurusamy. Advances in knowledge discovery and data mining. Menlo Park: MIT, 611 p., 1996.
- [8] A. Kampff. Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente, Tese de Doutorado, Universidade Federal do Rio Grande do Sul, Porto Alegre –RS, 2009.
- [9] N. LI, cohen, W., K.R. koedinger, N. Matsuda. A Machine Learning Approach for Automatic Student Model Discovery. EDM 2011: 31-40. Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, The Netherlands, July 6-8, 2011.
- [10] H. Liu. Towards Semantic Data Mining. In Proc. of the 9th International Semantic Web Conf. (ISWC2010), November 2010.
- [11] Y. Levy. Comparing dropouts and persistence in e-learning courses, Computers & Education, 48, 185–204, 2007.
- [12] R. L. Lobo e Silva Filho, P. R. Motejunas, O. Hipólito, M. B. C. M. Lobo. A evasão no Ensino Superior Brasileiro. Cadernos de Pesquisa, v. 37, n. 132, p. 641-659, set./dez. 2007.
- [13] J. T. Malin, C. Millward, H. A. Schwarz, F. Gomez, D. Throop, C. Thronesbery. Linguistic text mining for problem reports. Systems, Man and Cybernetics, 2009. IEEE International Conf. San Antonio, 2009.
- [14] L. M. B. Manhães, S. M. S. Cruz, R. J. M. Costa, J. Zavaleta, G. Zimbrão. Previsão de Estudantes

- com Risco de Evasão Utilizando Técnicas de Mineração de Dados. XVII WIE, 2011.
- [15] A. P. Machado, R. Ferreira, I. I. Bittencourt, E. Elias, P. Brito, E. Costa. Mineração de texto em redes sociais aplicada à educação a distância, Colabor@ Revista Digital da CVA Ricesu, ISSN 1519-8529. V.6(23), Julho de 2010.
- [16] MEC Ministério da Educação e Cultura (1997) "Diplomação, Retenção e Evasão nos cursos de Graduação em Instituições de Ensino Superior Públicas", http://www.udesc.br/arquivos/ id_submenu/102/diplomacao.pdf, Acesso em abril de 2013.
- [17] MEC Ministério da Educação (2007) "Diretrizes Gerais do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais REUNI", http://portal.mec.gov.br/sesu/arquivos/pdf/diretrizesreuni.pdf, Acesso em Abril de 2013.
- [18] S. Rezende. Mineração de Dados. XXV Congresso da Sociedade Brasileira de Computação, Minicurso, Universidade do Vale do Rio dos Sinos Unisinos, São Leopoldo, Julho de 2005.
- [19] C. Romero, S. Ventura, M. Pechenizkiy, R. S. J. BAKER, Handbook of Educational Data Mining, Ed. C R C, 535p, 2010.
- [20] J. Scott. Distance Education Report, California Community Colleges Chancellor's Office, 2011.
- [21] E. M. Santos, J. D. O. Neto. Evasão na Educação a Distância: identificando causas e propondo estratégias de prevenção. Universidade Metropolitana de Santos (Unimes). Revista Científica de Educação a Distância. Vol 2 N°2 Dez. 2009.
- [22] V. Tinto. Dropout from higher education: a theoretical synthesis of recent research. Washington, Review of Educational Research v. 45, n. 1, p. 89-125, 1975.
- [23] A. Toscher, M. Jahrer. Collaborative filtering applied to educational data mining. KDD Cup 2010: Improving Cognitive Models with Educational Data Mining, 2010.
- [24] I. H. WITTEN, E. FRANK. Data mining: practical machine learning tools and techniques with java implementations. San Francisco: Morgan Kaufmann, 369 p., 2000.
- [25] J. L. V. Barbosa, R. Hahn, D. N. BARBOSA, A. Saccol. A Ubiquitous Learning Model Focused on Learner Integration. International Journal of Learning Technology, v. 6, p. 62-83, 2011.
- [26] J. L. V. Barbosa, D. N. F. Barbosa, A. Wagner. Learning in Ubiquitous Computing Environments. International journal of information and communication technology education, v. 8, p. 64-77, 2012.



- [27] C. M. Bishop. Pattern Recognition and Machine Learning. 2. ed.. ed. [S.l.]: Springer, 2007.
- [28] S. Haykin. Redes neurais : princípios e prática. 2. ed.. ed. Porto Alegre: Bookman, 2001.
- [29] E. Alpaydin. Introduction to Machine Learning. 2nd. ed. [S.l.]: The MIT Press, 2010.
- [30] W. Cambruzzi, et al. 2012. Um Modelo para Gerenciamento de Múltiplas Trilhas Aplicado a Sistemas de Apoio à Educação. In: Simpósio Brasileiro de Informática na Educação. SBIE 2012.
- [31] G. Siemens, P. Long. Penetrating the fog: analytics in learning and education. Educause Review, [S.l.], v. 46, n. 5, p. 30–32, 2011.
- [32] M. A. Chatti., Al. L. Dyckoff, U.Schroeder, H. Thüs. A reference model for learning analytics. Int. J. Technol. Enhanc. Learn., Inderscience Publishers, Geneva, Switzerland. v. 4, n. 5/6, p. 318–331, January 2012.
- [33] W. Greller, H. Drachsler. Translating Learning into Numbers: a generic framework for learning analytics. Educational Technology & Society, [S.l.], v. 15, n. 3, p. 42–57, Nov. 2012.
- [34] Y. Levy. Comparing dropouts and persistence in e-learning courses. Computers & Education, [S.l.], v. 48, n. 2, p. 185 204, 2007.
- [35] M. D. C. Maia, F. D. S. Meirelles, S. Pela. Análise dos Índices de Evasão nos Cursos Superiores a Distância do Brasil. Congresso Internacional de Educação a Distância. ABED, [S.l.], p. 1–11, 2004.
- [36] C. R. J. Longo, Educação a Distância: o estado da arte. São Paulo: Pearson Education do Brasil, 2009
- [37] C. A. dos S. Baggi, D. A. Lopes. (2011). Evasão e Avaliação Institucional no Ensino Superior: uma discussão bibliográfica. Avaliação, Campinas, Sorocaba, SP. 16, 2, 355-374, Jul. 2011..
- [38] L. P. Kira. A evasão no ensino Superior: o caso do curso de pedagogia da Universidade estadual de Maringá(2002). Dissertação de mestrado. Universidade Metodista de Piracicaba. Piracicaba, São Paulo, 2002.
- [39] N. P, L. Gaioso. (2005) A evasão discente na educação superior no Brasil: perspectiva de alunos e dirigentes. Dissertação de mestrado, Universidade Católica de Brasília, Brasília, 2005
- [40] R. F. Vitelli, R. Fritsch, C. S. Rocha. Evasão em disciplinas de cursos de graduação: fatores intervenientes no fenômeno. In: Gestão e Políticas Públicas de Educação: desafios atuais, 2012, Pelotas. Seminário ANPAE região Sul, 2012. v.
- [41] A. Dunca, K. Cator. (2013) U.S. Department of Education, Office of Educational Technology,

- Expanding Evidence Approaches for Learning in a Digital World, Washington, D.C., 2013. Disponível em http://www.ed.gov/technology. Acesso em setembro de 2013.
- [42] L. Johnson, A Becker, M. Cummins, V. Estrada, A. Freeman, H. Ludgate. (2013). NMC Horizon Report: 2013 Higher Education Edition. Austin, Texas: The New Media Consortium.
- [43] M. Bienkowski, F. Mingyu, B. Means. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. U.S. Department of Education Office of Educational Technology. Center for Technology in Learning SRI International . October, 2012.
- [44] R. Ferguson. (2012). Learning analitycs: drivers, developments and challenges. International Journal of Technology Enhanced Learning (IJTEL), 4(5/6), 304-317. 2012.
- [45] B. Dietz-Uhler, J. Hurn. (2013) Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. Journal of Interactive Online Learning. V.12, N. 1, 2013.
- [46] G. Siemens, R. S. J. Baker, 2012. Learning analytics and educational data mining: towards communication and collaboration Proceeding LAK '12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge Pages 252-254 ACM New York, NY, USA 2012.
- [47] Y. A. ALMazroui. 2013 A survey of Data mining in the context of E-learning. International Journal of Information Technology & Computer Science (IJITCS) V.7: No: 3: January, 2013.
- [48] R. S.J.D Baker, K. Yacef, The State of Educational Data Mining in 2009: A Review and Future Visions, JEDM Journal of Educational Data Mining. V.1, I.1, Pages 3-17. 2009.
- [49] R. Jindal,M. D. Borah. 2013. A SURVEY ON EDUCATIONAL DATA MINING AND RESEARCH TRENDS. International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2013.
- [50] M. Paliwala, U. A. Kumar. 2009. Neural networks and statistical techniques: a review of applications. Expert Systems and Applications, 36(1):2–17. 2009.