

Identificação de Perfis de Evasão e Mau Desempenho para Geração de Alertas num Contexto de Educação a Distância

Identifying Evasion and Poor Performance Profiles for the Generation of Alerts in a Distance Learning Context

Adriana Justin Cerveira Kampff, Vinicius Hartmann Ferreira, Eliseo Reategui y José Valdeni de Lima

Programa de Pós-Graduação em Informática na Educação. Universidade Federal do Rio Grande do Sul. Av. Paulo Gama, 110 - prédio 12105 - 3º andar sala 332 90040-060 - Porto Alegre (RS) - Brasil

E-mail: akampff@gmail.com; vinihf@gmail.com; eliseoreategui@gmail.com; valdeni@inf.ufrgs.br

Información del artículo

Recibido 23 de Junio de 2014. Revisado 9 de Diciembre de 2014.
Aceptado 14 de Diciembre de 2014.

Palabras-chave:

Educación A Distância,
Informática Educativa,
Inteligência Artificial,
Abandono Escolar,
Estatísticas Educacionais

Keywords:

Distance Learning,
Computers In Education,
Artificial Intelligence,
Evasion, Educational
Statistics

Resumo

Este artigo apresenta um estudo sobre o emprego de um sistema que utiliza técnicas de mineração de dados para identificar perfis de evasão e mau desempenho de alunos em um contexto de educação a distância, com o objetivo de alertar professores sobre a situação destes estudantes. Os alertas dão suporte à atuação do professor no acompanhamento dos processos de aprendizagem, com o monitoramento de perfis de alunos e de eventos no ambiente, notificando o professor sobre grupos de alunos que compartilham necessidades específicas. Os resultados de experimentos envolvendo 1780 alunos permitiram concluir que o sistema de alerta proposto pode contribuir com o aumento dos índices de aprovação e redução dos índices de evasão de disciplinas na modalidade à distância. Durante a pesquisa, também foi possível observar que o sistema de alertas contribuiu de maneira significativa na gestão do ambiente virtual de aprendizagem. Muitas das práticas docentes para engajamento dos alunos em seus estudos, bem como de resgate dos estudantes em processo de evasão, foram realizadas a partir do encaminhamento de mensagens aos alunos. A pesquisa mostrou ainda um alto nível de satisfação dos estudantes com relação a estrutura do curso e diálogo/comunicação com o professor. Também foram positivos os resultados quanto à percepção dos estudantes com relação a sua autonomia na realização das atividades do curso.

Abstract

This article presents a study about the use of a system that employs data mining techniques to identify student evasion and poor performance profiles in a distance learning context, with the goal to alert teachers about the situation of these students. The alert system supports the role of the teacher in monitoring the learning process, keeping track of events and notifying teachers about groups of students who share specific needs. Experimental results involving 1780 students showed that the proposed alert system contributed to increase approval rates and reduce dropout rates in distance learning courses. It was also possible to observe that the alert system contributed significantly to the management of the virtual learning environment. Many of the teaching practices to engage students in their studies and rescue them from evading were based on email message exchanges. The research also showed a high level of student satisfaction regarding the course structure and their communication with the teachers. Results were also positive regarding the perception of students with respect to their autonomy in carrying out course activities.

1. Introdução

O ensino na modalidade a distância (EAD) se encontra em franca expansão no mundo todo. Porém, ao mesmo tempo em que a demanda é crescente por cursos em EAD, gerada pela necessidade de formação e democratização da educação, os índices de evasão e reprovação nesta modalidade de ensino mostram-se elevados. O termo evasão diz respeito aos alunos desistentes, que abandonam o curso antes de sua conclusão, não completando o percurso didático estabelecido. Embora a modalidade EAD esteja relacionada à flexibilidade de tempo e local de estudo para o aluno, a falta de tempo para estudar ou participar do curso, a falta de adaptação à metodologia e o aumento de tarefas no trabalho são apontados como a principal causa para a evasão dos alunos (Censo EAD, 2012).

Com o crescimento do número de alunos, acompanhar seus processos de aprendizagem visando ao êxito desses sujeitos nos cursos que realizam, exige cada vez mais dos professores. São volumosos os dados gerados nas interações entre alunos, entre professores e alunos, bem como as informações de acesso aos recursos disponibilizados. Reuni-los e interpretá-los é um processo complexo e exaustivo. Este processo pode ser auxiliado pela Mineração de Dados, definida por Han, Kamber e Pei (2012) como o processo de descobrir padrões relevantes e conhecimento a partir de um grande volume de dados. Baker, Isotani e Carvalho (2011) destacam que, devido a esta grande quantidade de dados produzida pelos Ambientes Virtuais de Aprendizagem (AVAs) ou sistemas ligados ao ensino, surgiu uma nova área de pesquisa denominada Mineração de Dados Educacionais (MDE). Nesta área de pesquisa utilizam-se métodos de Mineração de dados sobre dados educacionais com o objetivo de compreender melhor os alunos e seus processos de aprendizagem (Baker e Yacef, 2009; Romero e Ventura, 2013).

Neste contexto, através da análise e visualização de dados de alunos e seus contextos, com o objetivo de compreender e aprimorar o aprendizado e os ambientes em que ele ocorre, desenvolveu-se também a área denominada Analítica de Aprendizagem, do inglês Learning Analytics (Greller e Drachsler, 2012). Embora áreas de MDE e Analítica de Aprendizagem possuam objetivos similares, ambas diferenciam-se principalmente pelo tipo de descoberta que priorizam. Enquanto a MDE tem um grande foco em automatizar a descoberta de informações, a Analítica de Aprendizagem tem maior foco em auxiliar a tomada de decisão por meio da visualização ou análise detalhada de dados (Siemens e Baker, 2012).

Frente a este cenário, esta pesquisa busca, por meio da técnica de Mineração de Dados, identificar perfis de alunos com risco de evasão ou reprovação, visando à geração de alertas para sensibilizar o professor sobre possíveis problemas. Tais alertas são gerados a partir de relações identificadas pelos processos de mineração de dados e são direcionados a um ou mais alunos com necessidades similares, para que o professor possa dirigir uma comunicação ou atividade personalizada e contextualizada para esses sujeitos, tornando sua mediação mais eficaz e colaborando para reduzir os índices de insucesso em cursos da modalidade EAD.

2. Analítica de Aprendizagem

Através da ampliação da utilização de sistemas computacionais para dar suporte ao ensino ou mesmo na administração das instituições, um grande volume de dados é gerado. Porém, as instituições de ensino têm sido ineficientes quanto à utilização destes dados (Siemens e Long 2011; Cambuzzi, 2014). As oportunidades de intervenção nem sempre são aproveitadas, e quando o são, ocorrem muitas vezes de forma tardia ou ineficaz. Acrescenta-se a isso o fato de que os professores e gestores sem conhecimento técnico específico, apresentam dificuldades em manipular a grande quantidade de dados disponibilizada.

Neste contexto, cresce o interesse pela análise automática de dados educacionais, denominada Analítica de Aprendizagem (Greller e Drachsler, 2012). Cambruzzi (2014) destaca que não se trata de uma nova área de pesquisa, mas sim de um campo para o qual convergem diversos trabalhos sobre o emprego de técnicas de mineração e análise de dados para apoio a processos educacionais. Em um Painel, realizado no 2º LAK (International Conference on Learning Analytics & Knowledge) (Baker, Duval, Stamper, Wiley e Shum, 2012), Duval afirma que a área de Analítica de Aprendizagem busca rastrear dados registrados e percursos feitos pelos alunos ao utilizar os sistemas computacionais, empregando tais informações para otimizar processos educacionais. Além disso, de acordo com Duval, enquanto a MDE processa estes dados de forma automática para classificar ou descobrir padrões, a Analítica de Aprendizagem pode ser utilizada para fornecer informações aos professores de forma a auxiliá-los em suas tarefas docentes. No mesmo Painel, Wiley afirma que a Analítica de Aprendizagem combina, internamente, conhecimento e técnicas da mineração de dados para trabalhar com dados comportamentais e dados sobre o desempenho dos estudantes para apoiar processos de ensino e aprendizagem. Em suas aplicações finais, a Analítica de Aprendizagem integra conhecimento e técnicas de visualização de dados com o objetivo de possibilitar que professores e/ou alunos sem qualquer conhecimento técnico específico possam fazer uso dos dados para aperfeiçoar suas práticas (Baker *et al.*, 2012). Neste contexto, D' Aquim e Jay (2013), afirmam que a Analítica de Aprendizagem está relacionada à interpretação, ou visualização, dos resultados obtidos nos processos de MDE.

Pesquisas como as de McNelly, Gestwicki, Hill, Parli-Horne e Johnson (2012), Wolff, Zdrahal, Nikolov e Pantucek (2013), Cambruzzi (2014) e Jayaprakash, Moody, Lauría, Regan e Baron (2014), são uma amostra do crescente interesse em Analítica de Aprendizagem, bem como de suas múltiplas aplicações. No trabalho de McNelly *et al.* (2012) é apresentado o sistema *Uatu*, que integrado ao aplicativo de escrita colaborativa Google Docs, gera visualizações em tempo real de métricas sobre contribuição e edição de documentos colaborativos. Ao desenvolver o sistema, McNelly *et al.* (2012) tinham por objetivo responder duas questões: qual seria a contribuição da Analítica de Aprendizagem no processo de escrita colaborativa e se esta poderia apoiar o processo de intervenção do instrutor durante a produção textual. Os autores concluíram, após a realização de um experimento com o sistema *Uatu*, que este mostrou-se mais útil quando se trata da construção de um documento de maior complexidade por pessoas geograficamente distribuídas. Além disso, concluiu-se também que a Analítica de Aprendizagem aplicada à escrita colaborativa não substitui a intervenção e análise mais próxima do professor, porém serve como suporte nas intervenções. Acrescenta-se também que o sistema *Uatu*, e também a Analítica de Aprendizagem, tornam-se ferramentas relevantes em atividades de escrita desenvolvidas em cursos online.

O trabalho de Wolff *et al.* (2013) apresenta uma forma de prever o risco que alunos correm de reprovar em seus módulos de ensino, ou disciplinas. Utilizando dados registrados por um AVA sobre a interação dos alunos com os materiais e seu acesso ao AVA, é construído um modelo para o aluno. Este modelo, quando comparado a dados gerados no passado, também pelo AVA, torna possível prever e alertar situações em que os alunos têm comportamento similar àqueles com risco de reprovação. Apresentando resultados positivos, o trabalho destaca que é importante considerar que alunos aprendem de diferentes formas e a forma como cada curso é estruturado no AVA.

Similar ao trabalho de Wolff *et al.* (2013), Cambruzzi (2014) apresenta um modelo de arquitetura para um sistema Analítica de Aprendizagem com o objetivo de contribuir na redução da evasão em cursos EAD, intitulado GVwise. A arquitetura combina MDE e Analítica de

Aprendizagem, permitindo visualizar dados detalhados sobre a participação e desempenho dos alunos em diferentes componentes curriculares. Além disso, também facilita, através da mesma interface gráfica, a tomada de ações com o objetivo de estimular ou alertar o aluno sobre suas ações em um determinado AVA. Assim como no trabalho de Wolff *et al.* (2013), uma etapa de análise sobre a aplicação do sistema sobre turmas reais foi realizada, demonstrando o potencial da arquitetura proposta.

O trabalho apresentado por Jayaprakash *et al.* (2014) descreve uma iniciativa que tem objetivos e emprega técnicas similares às propostas por Wolff *et al.* (2013) e Cambuzzi (2014). A partir da combinação de técnicas de predição da MDE com alertas para os alunos em risco de reprovação (Analítica de Aprendizagem), concluiu-se que, sem os alertas, os alunos só se preocupariam com o risco de reprovação após notas baixas no intermeio do curso. Contudo, neste momento a recuperação das notas já seria difícil para muitos.

3. Mineração de Dados Educacionais

De acordo com Han *et al.* (2012), Mineração de Dados é o processo de descoberta de conhecimento e de padrões relevantes a partir de um grande volume de dados. Estes dados podem ser provenientes de Banco de Dados, da Web ou de quaisquer outros tipos de repositórios de informações.

Existem muitas funcionalidades para a Mineração de dados, como a caracterização e discriminação; descoberta de padrões frequentes, associações e correlações; classificação e regressão; agrupamentos; e análise de dados discrepantes. Cada uma dessas funcionalidades pode ser classificada em descritiva ou preditiva. Através da abordagem descritiva é possível caracterizar propriedades dos dados dentro de um conjunto, e através da abordagem preditiva é possível estimar uma previsão futura com base em dados similares do passado (Han *et al.*, 2012).

A Mineração de Dados incorpora tecnologias de diferentes domínios, como estatística, aprendizagem de máquina, banco de dados e recuperação de informações. Para Han *et al.* (2012), esta interdisciplinaridade contribui significativamente para o sucesso desta área de pesquisa. Além disso, de forma cada vez mais evidente, diferentes áreas têm buscado na Mineração de Dados uma forma de extrair informações a partir de uma enorme quantidade de dados. Dentre estas áreas destacam-se a Administração, a Saúde, a Bioinformática e também a Educação.

De acordo com Siemens e Baker (2012), a MDE pode ser definida como uma área emergente, que tem por objetivo desenvolver métodos para explorar dados provenientes de sistemas educacionais ou AVAs, utilizando estes métodos para compreender melhor os estudantes e suas características. De acordo com Baker *et al.* (2012), MDE e Analítica de Aprendizagem diferenciam-se principalmente por seu foco. Enquanto a Analítica de Aprendizagem se foca no aluno e em usar os dados educacionais para explorar como a interação do aluno com a tecnologia pode afetar sua aprendizagem, a MDE se foca na utilização dos dados para compreender os processos de aprendizagem.

Para Baker *et al.* (2011), a disponibilização de dados padronizados e bem estruturados pelas instituições de ensino é condição fundamental para o avanço da MDE. Com o crescimento rápido da modalidade EAD no Brasil, a quantidade de dados que podem ser utilizados em pesquisas em MDE tende a crescer. E, a partir do desenvolvimento da MDE, será possível desenvolver mecanismos e ferramentas educacionais mais eficientes, modelos para identificar alunos com dificuldades, aperfeiçoar os materiais didáticos e desenvolver métodos pedagógicos mais eficazes.

Dentre as muitas pesquisas envolvendo MDE, cabe destacar os trabalhos desenvolvidos por Durand, Laplante e Kop (2011), como apoio para sistemas de recomendação, Li e Matsuda (2011), para identificar e capturar perfis e Cambruzzi (2014), Wollf et al. (2013) e Jayaprakash et al. (2014), já citados anteriormente e que combinam técnicas de Analítica de Aprendizagem com MDE para auxiliar no combate à evasão escolar na modalidade EAD. Percebe-se, através destes trabalhos, que quando os resultados da MDE podem dar subsídios a diferentes ações pedagógicas ou de gestão. No contexto do trabalho aqui apresentado, através do conhecimento extraído dos dados são gerados alertas que sinalizam aos professores possíveis problemas com seus estudantes.

4. O Sistema de Alertas Proposto

Com base nos estudos realizados, foi proposto um sistema de alertas para AVA configurável pelo professor a partir de indicadores do ambiente virtual, e informações gerados a partir da mineração de dados educacionais.

Dois tipos de alertas são gerenciados pelo sistema, chamados aqui de alertas fixos e alertas baseados em padrões. Os alertas fixos estão relacionados a questões que o professor deseja acompanhar explicitamente, tais como acesso a materiais, avaliações, presença em sala de aula, acesso ao sistema, realização de tarefas. A Figura 1 exemplifica possibilidades de configuração de alertas relativos aos materiais disponíveis na ferramenta Biblioteca Virtual. Para cada material disponibilizado, o professor pode indicar se o acesso é obrigatório, bem como a data limite preferencial. Além disso, o professor deve indicar em que período deseja ser notificado sobre os alunos que não acessaram o material. Assim, o sistema gerará um alerta informando quais alunos não acessaram o recurso, por exemplo, um dia após a data limite indicada.

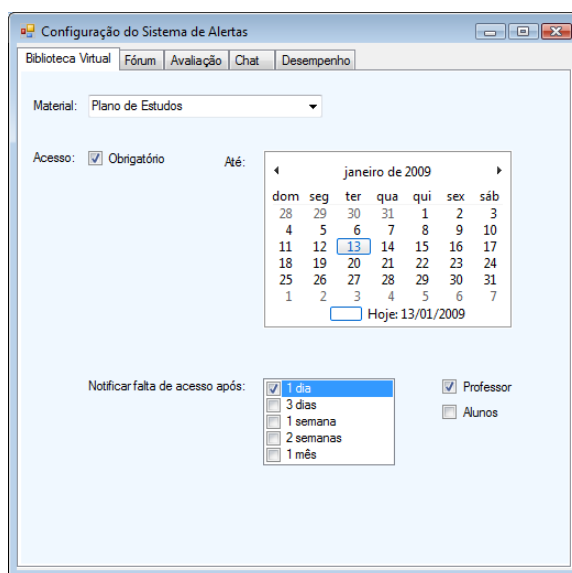


Figura 1 ± Exemplo de Configuração dos Alertas: Biblioteca Virtual

Os alertas baseados em padrões, obtidos por meio de processos de mineração de dados, são automaticamente gerados pelo sistema. O módulo de mineração de dados consulta os dados históricos das disciplinas ou cursos correspondentes, previamente organizados, e gera as regras de classificação

para o período correspondente, armazenando-as na base de regras de classificação. A base de dados com os dados históricos é atualizada a cada nova edição de um curso, o que leva à necessidade de atualização / geração de novas regras para aplicação em turmas subsequentes. Neste trabalho, para geração das regras de classificação, foram utilizados os algoritmos RuleLearner e DecisionTree da ferramenta de mineração RapidMiner¹.

O módulo de aplicação do sistema de alertas, aplica as regras sobre os dados dos alunos e os classifica como evadido, aprovado, reprovado ou sem acesso. A Figura 2 mostra o ambiente, com regras e dados extraídos do experimento realizado para o trabalho, para exemplificar seu funcionamento.

Figura 2 ± Sistema de Alertas: classificação

A aba Visualização de Regras permite que o professor veja, para cada regra específica, quais são os alunos classificados por ela.

5. Experimento

A partir dos estudos realizados e da arquitetura criada para sistemas de alertas em AVA, realizou-se um experimento em disciplinas ofertadas na modalidade a distância em uma universidade no sul do país. Nesta instituição, cada turma de até 80 alunos é acompanhada por um professor. Com calendário comum, as atividades de abertura e encerramento das salas virtuais, bem como as datas das provas presenciais, são compartilhadas e compatibilizadas com o calendário acadêmico dos cursos de graduação presenciais.

Para validar a geração de alertas a partir de dados históricos de desempenho nas disciplinas e eventuais evasões nos cursos, optou-se pelo emprego de dados gerados por turmas ofertadas a distância

¹ <http://rapidminer.com/>

da disciplina de Instrumentalização Científica (IC), que conta com uma estrutura bem estabelecida, tanto na organização de materiais quanto nas atividades de apoio e diálogo com os alunos. Foram considerados dados históricos de acesso ao AVA por alunos distribuídos em 20 turmas, em dois semestres diferentes. As turmas eram constituídas por alunos provenientes de diferentes cursos e de 8 unidades presenciais distintas, totalizando 1550 alunos.

O fato da disciplina ter estrutura com conteúdos e atividades relativamente estável foi um fator facilitador para esta pesquisa. O conjunto de tarefas propostas e recebidas em semanas específicas, conforme plano de ensino da disciplina, era seguido por todos os professores. Cada uma das tarefas avaliativas buscava avaliar aspectos definidos da disciplina, não havendo variação no conjunto de atividades nos semestres considerados.

5.1. A mineração de dados

Para cada aluno, foram obtidos os dados como: unidade presencial, curso de graduação em que estava matriculado, sexo, idade, nota de cada avaliação da disciplina (data final de entrega estipulada pelo professor e data de entrega efetivada pelo aluno, perfazendo um total de cinco atividades), notas das avaliações. Para cada uma das 19 semanas do semestre letivo foram também obtidas informações sobre número de acessos ao ambiente virtual, à ferramenta aula (conteúdo on-line), à biblioteca virtual (materiais complementares), ao fórum, o número de postagens no fórum, o número de e-mails enviados e o número de e-mails recebidos. Para o processo de mineração de dados, foram selecionados 87 atributos.

A descoberta de conhecimento por meio da mineração, para a geração de alertas, foi organizada em cinco momentos determinados, para extração de regras para cada um dos cinco meses do semestre letivo. As informações demográficas, os dados referentes ao primeiro mês de interação no ambiente, bem como o resultado final alcançado, foram minerados para observar como as características dos alunos e seus comportamentos no ambiente poderiam servir de fatores preditivos dos resultados finais. Ou seja, buscou-se identificar padrões de dados/comportamentos que fossem indicadores de resultados finais, utilizando-se contudo informações apenas das primeiras semanas do semestre.

O processo de mineração foi realizado também ao final do mês dois (semana 8), mês três (semana 11), mês quatro (semana 15) e mês cinco (semana 19), de forma análoga ao descrito para o primeiro mês, sempre incorporando os dados gerados no espaço de tempo do início de curso até a semana indicada. Utilizou-se desta vez o algoritmo RuleLearner da ferramenta RapidMiner. A opção por este algoritmo deu-se pelo fato de que o volume de dados e atributos era muito grande para a obtenção de uma árvore de decisão que pudesse ser utilizada de forma prática na geração de alertas. Julgou-se que a extração de regras a partir dos dados traria resultados mais fáceis de serem aplicados na geração dos alertas.

A análise das regras geradas permite observar padrões interessantes a partir das características demográficas dos alunos e de seus primeiros comportamentos de acesso. Para exemplificar, destacam-se algumas regras, no Quadro 1. As regras são empregadas em seqüência para a classificação dos alunos, o que significa dizer que, uma vez classificado, o aluno não é submetido às regras posteriores.

Quadro 1 ± Exemplos de Regras obtidas pelo RapidMiner

1. if AF_TM1 > 6.500 and AG_TM1 > 25.500 and AF_TM1 <= 53 then Aprovado
4. if FREQ_M1 > 1.500 and AB_TM1 > 17.500 and AF_TM1 > 26.500 and AB_TM1 <= 93.500 and AF_TM1 <= 32.500 then Aprovado
6. if FREQ_M1 > 1.500 and AF_TM1 > 13.500 and Idade > 26.500 and Idade <= 35.500 then Aprovado
7. if FREQ_M1 > 1.500 and AF_TM1 > 13.500 and AF_FM1 > 2.500 and Curso = DIREITO then Aprovado
24. if AG_TM1 > 4.500 and AG_TM1 <= 12.500 and Pri_Ac > 1.500 and PF_TM1 <= 1.500 and AB_TM1 > 5.500 and AB_TM1 <= 19.500 and ER_TM1 > 1.500 and ER_TM1 <= 14.500 and Área = EducaçãoCiênciaseArtes then Aprovado

A regra 1, obtida pela mineração dos dados históricos, indica que se o acesso total à sala virtual no primeiro mês foi superior a 25 ($AG_TM1 > 25.500$) e o acesso total aos fóruns esteve entre 7 ($AF_TM1 > 6.500$) e 53 ($AF_TM1 \leq 53$) vezes, o aluno tende a se enquadrar na classe Aprovado. A regra 4 evidencia questões de frequência ao ambiente em no mínimo 2 das quatro semanas ($FREQ_M1 > 1.500$), com acessos às aulas e à biblioteca entre 18 ($AB_TM1 > 17.500$) e 93 ($AB_TM1 \leq 93$) e acessos aos fóruns entre 27 ($AF_TM1 > 26.500$) e 32 ($AF_TM2 \leq 32.500$) vezes como padrão para enquadramento na classe Aprovado.

As regras 6, 7 e 24 trazem a combinação de mais elementos, como faixa etária, curso, área do curso do aluno, frequência de acesso a ferramentas diversificadas do AVA, semana em que foi realizado o primeiro acesso, quantidade de postagens em fóruns, quantidade de e-mails recebidos. As regras baseiam-se em padrões identificados nas ocorrências anteriores da disciplina, relacionando os resultados obtidos pelos alunos aos dados disponíveis no período de tempo considerado.

Na pesquisa realizada, as regras de classificação obtidas na etapa de mineração foram utilizadas para auxiliar no acompanhamento de novas turmas, alertando o professor sobre alunos em risco e suas características, com o propósito de servir de apoio à ação docente.

5.2. Alertas gerados e exemplos de intervenção

A partir das regras de classificação obtidas dos históricos de acesso ao AVA e da definição de indicadores, acompanhou-se quatro turmas de IC durante um semestre, totalizando 230 alunos, gerando alertas e realizando intervenções junto a eles. Os dados das turmas de acompanhamento foram obtidos para cada período de tempo estipulado (semana 4, 8, 11, 16 e 19).

Para o experimento realizado, foram definidos como indicadores de acompanhamento:

- Atividades avaliativas pendentes ± alerta fixo, a partir do prazo/evento definido pelo professor. Com dois dias de antecedência do prazo para entrega de cada tarefa, o professor era notificado sobre quem eram os alunos que ainda estavam com a tarefa pendente;
- Regras de classificação ± alertas do sistema, a partir da aplicação das regras resultantes da mineração de dados. No final do mês 2, 3 e 4, o professor foi notificado sobre grupos de alunos com tendência à aprovação, à reprovação ou à evasão, indicando a regra obtida pelo *RapidMiner* que permitiu classificar cada grupo de alunos.

Cada alerta, com base no evento ou regra específica, apontava a regra e o grupo de alunos, notificando o professor e sugerindo o contato com os estudantes. A partir da análise do professor, outras ações puderam ser tomadas, como, por exemplo, a ampliação de prazos ou a oferta de materiais complementares.

5.3. Alertas sobre atividades avaliativas pendentes

Durante o semestre no qual o experimento foi realizado, para cada uma das atividades avaliativas (A1 a A5 na Tabela 1), com dois dias de antecedência para o prazo final de entrega, o professor foi alertado sobre o grupo de alunos que ainda não havia enviado sua atividade. Com base no alerta, nas características da atividade e do grupo de alunos, o professor pôde enviar uma comunicação mais adequada para estes alunos. A tabela 1 mostra, para um total de 230 alunos, o número de alunos que já havia entregue a tarefa com dois ou mais dias de antecedência, o número de alunos que recebeu o alerta sobre o prazo da atividade, a quantidade de alunos que realizou a tarefa neste prazo final, o total de alunos que entregou as atividades e a mensagem enviada.

Tabela 1 ± Alertas fixos: atividades avaliativas pendentes

	Alunos com trabalho entregue com 2 ou mais dias de antecedência	Alunos que receberam o alerta sobre prazo da atividade	Alunos que entregaram atividade no prazo	Total de alunos que entregou a atividade, para o total de 230
A1	132	98	85	217
A2	127	103	82	209
A3	184	46	18	202
A4	138	92	75	213
A5	143	87	70	213

O número de alunos que deixou para enviar as tarefas avaliativas no prazo final é bastante elevado. Não é possível, em um primeiro olhar, afirmar que os alertas lembrando prazos e atividades pendentes tenham levado o aluno a fazê-las ou que teriam deixado de realizá-las caso não fossem contatados. Mas, os índices de aprovação foram superiores e os índices de evasão menores em comparação àqueles obtidos nas turmas anteriores, sem os alertas. Além disso, quando questionados sobre sua percepção dos alertas, os alunos afirmaram que as mensagens os auxiliavam na lembrança de atividades pendentes, como a realização de atividades avaliativas. Entre o grupo de participantes, 85,45% concordaram plenamente com esta afirmação, seguidos de 7,27% que também concordaram.

5.4. Alertas baseados em regras de classificação

Após o final dos meses 2, 3 e 4, foram empregadas as regras extraídas pelo RapidMiner para alertar o professor sobre a possibilidade de seus alunos se enquadrarem em um perfil de reprovação ou evasão. A Tabela 2 apresenta o número de alunos classificados pelo sistema em cada classe (aprovação, evasão e reprovação), as regras que os classificaram e observações sobre as regras e sobre os resultados obtidos ao final do semestre, após a conclusão do experimento.

O professor, ao ser alertado sobre os grupos de alunos em cada uma das três classificações, poderia, por exemplo, emitir uma mensagem única para o grupo desejado, ou analisar os subgrupos vinculados a cada regra de uma classificação específica, de forma a observar mais detalhadamente as especificidades de cada grupo. Dos 46 alunos acompanhados de maneira mais próxima (44 classificados como possíveis evasores e 2 como possíveis aprovados), 38 obtiveram aprovação ao final do curso.

Tabela 2 ± Classificação Mês 2

Classificação: Tendência a	No. de Alunos classificados
<i>Aprovação</i>	184
<i>Evasão</i>	44
<i>Reprovação</i>	2

Basicamente, havia dois grandes indicadores nas regras que classificaram os alunos como possíveis evasores ou reprovados: baixa frequência de acesso aos fóruns e baixo rendimento nas primeiras atividades. Para envolvê-los em seus estudos, de forma a participarem das discussões da turma, bem como retomar os estudos relacionados, foram enviadas mensagens especificamente para esses alunos, colocando a importância das leituras complementares e da participação e leitura dos fóruns de discussão. O Quadro 2 mostra um exemplo de mensagem enviada nestas situações.

Quadro 2 ± Exemplo de mensagem enviada para alunos após mês 2

Prezado(a) NOME_DO_ALUNO,

Após o estudo dos capítulos 6, 7, 8 e leitura do texto "O que é científico?" - participe do fórum sobre esses capítulos, o que pode não ter uma implicação direta em sua avaliação, mas pode auxiliá-lo na sistematização do conhecimento e demonstrar seu engajamento na disciplina. Além disso, leia a contribuição dos colegas! Imprima ritmo de estudo e organização, sabendo que os professores estão aqui para apoiá-lo nos seus percursos de aprendizagem!

Para o mês três, novamente as regras obtidas pelo RapidMiner foram utilizadas para identificar alunos com perfil de evasão ou de reprovação. A Tabela 3 apresenta o número de alunos classificados em cada classe (aprovação, evasão e reprovação), as regras que os classificaram e observações sobre as regras e sobre os resultados obtidos ao final do semestre, após a conclusão do experimento.

Tabela 3 ± Classificação Mês 3

Classificação: Tendência a	No. de Alunos classificados
<i>Aprovação</i>	197
<i>Evasão</i>	25
<i>Reprovação</i>	8

O acesso ao AVA tardiamente mostrou-se aqui como um dos principais indicadores de mau desempenho ou de evasão da disciplina. Além disso, o baixo número de acessos ao ambiente, baixa nota de G1 e a ausência de postagens nos fóruns estavam entre os antecedentes de muitas das regras de mau prognóstico. O Quadro 3 mostra um exemplo de mensagem enviada nestas situações.

Quadro 3 ± Exemplo de mensagem enviada para alunos após mês 3

Prezado(a) NOME_DO_ALUNO,

A nossa discussão no fórum sobre Direitos Autorais é bastante relevante. Compreender a temática e conhecer a lei que trata do assunto em nosso país nos auxiliará em nossos trabalhos acadêmicos e mesmo em situações do nosso cotidiano, em que tantos não respeitam os direitos dos autores de produção intelectual. Estamos em fase de debates no fórum e gostaríamos de ouvir sua opinião. Participe!

Dica importante: Leia as contribuições já postadas no fórum, pois as reflexões dos colegas podem auxiliar em uma compreensão mais ampla sobre os Direitos Autorais!

Bons estudos e aguardo a sua contribuição!

Para o mês quatro, mais uma vez as regras obtidas pelo RapidMiner foram utilizadas para identificar alunos com perfil de evasão ou de reprovação. A Tabela 4 apresenta os resultados respectivos.

Tabela 4 ± Classificação Mês 4

Classificação: Tendência a	No. de Alunos classificados
<i>Aprovação</i>	170
<i>Evasão</i>	41
<i>Reprovação</i>	19

Dos 60 alunos em evidência (41 classificados como possíveis evasores e 19 com perfil de possível reprovação), 43 obtiveram êxito na conclusão da disciplina. O mês 4 foi concluído na semana 16 e a prova presencial ocorreu na semana 17, sendo a substituição de grau, para os alunos que necessitaram, realizada na semana 19. O Quadro 4 mostra um exemplo de mensagem enviada.

Quadro 4 ± Exemplo de mensagem enviada para alunos após mês 4

Prezado NOME_DO_ALUNO,

No próximo sábado teremos a nossa prova de substituição de grau de Instrumentalização Científica! Está aí a oportunidade para, retomando os estudos, alcançar os objetivos propostos na disciplina e obter aprovação. Você pode estudar um grupo de conteúdos por dia, ou então dois ou três grupos por dia, dependendo da sua disponibilidade. É interessante, também, retomar os fóruns temáticos e as avaliações realizadas no NetAula.

Desejo um bom estudo e me coloco à disposição, mais uma vez, para auxiliar na resolução de dúvidas durante seus estudos.

Para fins de fechamento do experimento, foram aplicadas as regras pós mês 5 para os alunos das turmas acompanhadas. Todos os alunos foram classificados corretamente, ou seja, o resultado obtido por cada aluno acompanhado foi igual à classificação indicada pela aplicação das regras geradas na mineração dos dados históricos.

6. Resultados

A Tabela 5 apresenta os números absolutos e percentuais de cada tipo de resultado alcançado pelos alunos, na Amostra 1 (dados históricos dos dois semestres inicialmente considerados) e na Amostra 2 (dados das turmas acompanhadas no semestre corrente).

Tabela 5 ± Amostra H (Dados Históricos) x População 2 (Dados Acompanhados)

	Amostra 1 (Histórico)		Amostra 2 (Acompanhamento)	
	Número de Alunos	Percentual	Número de Alunos	Percentual
<i>Aprovado</i>	1286	82,97	199	86,52
<i>Reprovado</i>	52	3,35	7	3,04
<i>Evadido</i>	212	13,68	24	10,43
<i>Total</i>	1550	100,00	230	100,00

Para validação dos resultados, utilizou-se o teste de cálculo de Z para comparação dos percentuais de aprovação, reprovação e evasão. Na equação, p_1 e p_2 representam cada uma das proporções (Amostra 1 e Amostra 2, respectivamente); q é seu complemento ($1 - p$); e n é o tamanho da amostra.

$$(1) \quad Z_{\text{CALC}} = \frac{(p_1 - p_2)}{\sqrt{\frac{p_1 \times q_1}{n_1} + \frac{p_2 \times q_2}{n_2}}}$$

Tomando por base um nível de significância α de 0,1, o aumento ou a redução dos índices (teste unilateral), foi obtido o valor para Z_{TAB} de -1,28 (teste lateral à esquerda) e de 1,28 (teste lateral à direita).

Para a hipótese de que o percentual de aprovados nas turmas sem alertas (Amostra 1) foi inferior ao percentual de aprovados nas turmas com alertas (Amostra 2), calculou-se o valor de Z_{CALC} aplicando-se o teste unilateral à esquerda. O valor resultante foi -1,45141, inferior ao Z_{TAB} de -1,28, com nível de significância $\alpha=0,1$. Há evidências, portanto, de que o índice de aprovação nas turmas para as quais foram gerados os alertas foi significativamente superior ao índice de aprovação observado na amostra dos dados históricos.

Com relação à hipótese de que o percentual de reprovados nas turmas sem alertas (Amostra 1) foi superior ao percentual de reprovados nas turmas com alertas (Amostra 2), aplicado o teste unilateral à direita, obtendo-se um valor para Z_{CALC} de 0,253924, inferior ao Z_{TAB} de 1,28, com nível de significância $\alpha=0,1$. Neste caso, fica rejeitada a hipótese de que o índice de reprovação nas turmas com os alertas tenha sido significativamente inferior ao índice de reprovação observado na amostra dos dados históricos.

Por fim, para a hipótese de que o percentual de evadidos nas turmas sem alertas (Amostra 1) foi superior ao percentual de evadidos nas turmas com alertas (Amostra 2), aplicou-se o teste unilateral à direita. Z_{CALC} resultou no valor 1,479774, superior ao Z_{TAB} de 1,28, com nível de significância $\alpha=0,1$. Há evidências, portanto, de que o índice de evasão nas turmas com os alertas tenha sido significativamente inferior ao índice de evasão observado na amostra dos dados históricos.

Além da análise dos dados de desempenho e evasão nas disciplinas durante o semestre de acompanhamento dos alunos, foi realizada uma pesquisa junto a estes para melhor conhecer sua percepção com relação ao desenvolvimento das disciplinas na modalidade a distância, bem como com relação ao sistema de alertas utilizado. Sem caráter obrigatório e avaliativo, os alunos foram convidados a responder o questionário enviado, que apresentava objetivos da pesquisa e propunha um conjunto de questões que podiam ser respondidas de acordo com uma escala Likert de 5 pontos. Por se tratar de uma pesquisa com participação voluntária, um total de 55 questionários foi retornado (Tabela 6).

Tabela 6 ± Resultados dos questionários com os alunos

	Concordo Totalmente	Concordo	Indiferente	Discordo	Discordo Totalmente
1. O Ambiente Virtual de Aprendizagem (NetAula) foi adequado para as demandas de estudar a distância.	27 (49,09%)	26 (47,27%)	1 (1,82%)	1 (1,82%)	0,00
2. O Plano de Ensino e Aprendizagem, juntamente com o Cronograma de Atividades do semestre, disponíveis na Biblioteca Virtual, traziam detalhes suficientes para o bom andamento da disciplina.	32 (58,18%)	23 (41,82%)	0,00	0,00	0,00
3. O Plano de Atividades, na parte central da sala virtual, facilitou a organização e o acesso aos materiais e tarefas, bem como a compreensão do encadeamento das atividades da disciplina.	37 (64,45%)	17 (30,91%)	1 (1,82%)	0,00	0,00
4. Os critérios de avaliação e prazos de liberação, entrega e devolução de notas foram informados e cumpridos.	36 (65,45%)	16 (29,09%)	0,00	3 (5,45%)	0,00
5. Os avisos gerais foram relevantes e atualizados.	39 (70,91%)	13 (23,64%)	2 (3,64%)	1 (1,82%)	0,00
6. Os professores utilizaram uma linguagem clara e objetiva ao comunicar-se.	43 (78,18%)	10 (18,18%)	1 (1,82%)	1 (1,82%)	0,00
7. As mensagens enviadas pelos professores eram lidas, seja no e-mail externo ou no e-mail interno da NetAula.	37 (62,27%)	16 (29,09%)	2 (3,64%)	0,00	0,00
8. As mensagens enviadas por meio da NetAula eram recebidas no seu e-mail externo.	0,00	54 (98,18%)	0,00	1 (1,82%)	
9. O recebimento de uma mensagem no e-mail externo o levava a acessar o ambiente.	34 (61,82%)	14 (25,45%)	5 (9,09%)	2 (3,64%)	0,00
10. As mensagens auxiliavam na lembrança de atividades pendentes, tais como leitura de materiais, realização de atividades avaliativas e participação em fóruns.	47 (85,45%)	4 (7,27%)	3 (5,45%)	1 (1,82%)	0,00

	Concordo Totalmente	Concordo	Indiferente	Discordo	Discordo Totalmente
11. As mensagens foram importantes para a sua organização e engajamento na disciplina.	41 (74,55%)	9 (16,36%)	3 (5,45%)	2 (3,64%)	0,00
12. As mensagens auxiliaram no estabelecimento de um ritmo de estudos a distância.	35 (63,64%)	13 (23,64%)	5 (9,09%)	2 (3,64%)	0,00
13. As mensagens permitiram estabelecer um vínculo mais próximo com o professor, pois você percebeu que este estava atento às suas necessidades.	38 (69,09%)	12 (21,82%)	3 (5,45%)	2 (3,64%)	0,00
14. A quantidade e a frequência de mensagens foram adequadas ao momento da disciplina e à sua necessidade.	40 (72,73%)	12 (21,82%)	2 (3,64%)	0,00	1 (1,82%)
15. As mensagens auxiliaram na compreensão das relações entre os conteúdos e as atividades, bem como da importância de realizá-las para a aprendizagem dos assuntos trabalhados.	40 (72,73%)	11 (20,00%)	4 (7,27%)	0,00	0,00
16. Houve interesse da sua parte para realizar as leituras dos conteúdos liberados e materiais complementares sugeridos.	32 (58,18%)	13 (23,64%)	4 (7,27%)	6 (10,91%)	0,00
17. Houve organização da sua parte para realizar os trabalhos solicitados e cumprir os prazos.	27 (49,09%)	23 (41,82%)	2 (3,64%)	3 (5,45%)	0,00
18. Você manteve uma postura ativa de aprendizagem, acessando regularmente o ambiente NetAula.	29 (52,73%)	22 (40,00%)	0,00	4 (7,27%)	0,00
19. Você participou dos fóruns, visando aprofundamento de temas e esclarecimento de dúvidas.	31 (56,36%)	19 (34,55%)	3 (5,45%)	2 (3,64%)	0,00
20. Você leu as diversas postagens dos fóruns para aprender e contribuir com os colegas.	24 (43,64%)	21 (38,18%)	7 (12,73%)	3 (5,45%)	0,00

As questões de 1 a 4 referiram-se a aspectos da estrutura da disciplina. Os resultados mostram a satisfação dos alunos com relação a este aspecto, com a maioria das respostas «concordo» e «concordo totalmente». As questões 5 a 15 diziam respeito aos aspectos de diálogo e comunicação com o professor. As respostas dos estudantes também mostram sua satisfação quanto ao conteúdo, frequência e adequação das mensagens. A maioria dos alunos concordou que o recebimento de uma mensagem (alerta) os levava a acessar a sala de aula virtual, além de ser uma forma de lembrá-los sobre as atividades pendentes. Ainda sobre os alertas recebidos, a maioria dos concordou que as mensagens recebidas permitiram estabelecer um vínculo mais próximo com o professor, ou seja, minimizar a distância transacional, reduzindo a sensação de isolamento, percebendo que o professor acompanhava e

estava atento a suas necessidades. As questões 16 a 20 tratavam de aspectos relativos à autonomia do estudante. Os resultados mostram que os alunos tiveram uma percepção muito positiva com relação a sua organização e postura ativa na realização das atividades.

7. Discussão e Considerações Finais

Este artigo apresentou uma proposta de sistema de alertas em AVA a partir de dados gerados por meio de um processo de mineração de dados. Os alertas dão suporte à atuação do professor no acompanhamento dos processos aprendizagem, com o monitoramento de perfis de alunos e de eventos no ambiente, notificando o professor sobre grupos de alunos que compartilham necessidades específicas. Cada alerta, portanto, notifica o professor sobre a situação de um grupo de alunos, formado por um ou mais sujeitos, para que o professor possa dirigir comunicações ou atividades específicas. Os grupos identificados são dinâmicos, sendo gerados no momento oportuno para sinalizar ao professor situações específicas detectadas por conjuntos de regras oriundas do sistema de mineração de dados.

O sistema de alerta proposto diferencia-se de outras pesquisas por combinar uma proposta híbrida como fonte dos alertas gerados ao professor: indicadores pré-estabelecidos com parâmetros configuráveis, além da aplicação das regras geradas pela mineração de dados. Diferente dos trabalhos desenvolvidos por Cambruzzi (2014), Wolff et al. (2013) e Jayaprakash et al. (2014), que se focam somente em alertas gerados através de MDE, os alertas gerados através de indicadores pré-estabelecidos pelos professores também são relevantes. Através desta possibilidade de configuração, o sistema de alertas permite ao professor acompanhar a o andamento das atividades, possibilitando a este intervir em situações específicas.

Acrescenta-se ainda que, ao invés de gerar alertas automáticos para os alunos, o sistema proposto sinaliza ao professor situações possivelmente problemáticas. A decisão de intervir e de como fazê-lo fica ao encargo do professor. Esta forma de operação do sistema contribui para reforçar a comunicação entre professor e aluno. Esta abordagem reforça as conclusões de McNelly et al. (2012) sobre a relação do tutor/professor com a visualização dos dados (Analítica de Aprendizagem). Ambos os trabalhos concordam que o sistema não substitui o papel de um tutor ou professor, mas serve como mecanismo de auxílio na tomada de decisões.

A mineração de dados, por sua vez, traz um diferencial importante para a geração de alertas: mais do que eventos fixos, pré-programados, os alertas podem basear-se na descoberta de relações, de padrões de comportamento e características de alunos em edições anteriores do curso, analisando a combinação de vários aspectos para identificar perfis de alunos/comportamentos com tendência à evasão.

Os resultados de experimentos envolvendo 1780 alunos, incluindo histórico de cursos anteriores e alunos acompanhados durante um semestre letivo, permitiram concluir que o sistema de alerta proposto pode contribuir com o aumento dos índices de aprovação e redução dos índices de evasão de disciplinas na modalidade à distância.

Durante a pesquisa, também foi possível observar que o sistema de alertas contribuiu de maneira significativa na gestão do AVA. Muitas das práticas docentes de engajamento dos alunos em seus estudos, bem como de resgate dos estudantes em processo de evasão, foram realizadas a partir do encaminhamento de mensagens aos alunos. Além disso, através dos alertas, assim como no trabalho de Jayaprakash et al. (2014), verificou-se uma possibilidade de informar aos alunos, de forma antecipada, que suas estratégias de estudo não estavam levando-os a resultados positivos. Estes, tendo sido identificados por seu perfil/comportamento de acesso ao AVA, demandavam uma atenção maior, de forma que voltassem a participar ativamente das atividades da disciplina. Os alertas automatizados pelo

sistema, contribuíram para a tomada de decisão e intervenção junto aos alunos, visando a uma ação pedagógica de qualidade.

As áreas de MDE e Analítica de Aprendizagem têm muito a contribuir para a área da Educação e, em especial, para a EAD. São muitos os desafios da EAD para democratizar e qualificar os processos de ensino e aprendizagem. Ferramentas que integrem conhecimento dessas áreas, potencialmente, podem contribuir para a identificação de padrões em grupos normalmente bastante heterogêneos, apoiando professores, experientes ou não, em suas tarefas docentes. Também é importante ressaltar que cada estudante é um indivíduo único e desenvolve estratégias próprias de conduzir seu processo de aprendizagem, como destacado por Wolff et al. (2013). Consequentemente isto implica em formas diferentes de utilização de um AVA, sendo este um ponto importante com relação ao emprego de técnicas de MDE na predição de resultados.

8. Referências

- Baker, R. S. J. D., Isotani, S e Carvalho, A. M. J. B. (2011). Mineração de dados educacionais: oportunidades para o brasil. *Revista Brasileira de Informática na Educação*, 19 (2).
- Baker, R. S. J. D. e Yacef, K. (2009). The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining*, 1 (1).
- Baker, R. S. J. D., Duval, E, Stamper, J., Wiley, D e Shum, S. B. (2012). Panel: educational data mining meets learning analytics. *Proceedings Of International Conference On Learning Analytics & Knowledge*, 2.
- Cambrozzi, W. L. (2014). *GVwise: uma aplicação de learning analytics para a redução da evasão na educação a distância*. Dissertação de Mestrado em Computação Aplicada sem publicação, Universidade do Vale do Rio dos Sinos, Brasil.
- Censoead.Br (2012). *Relatório analítico da aprendizagem a distância no Brasil 2012*. Curitiba: IBPEX.
- D@quin, M. e Jay, N. (2013). Interpreting data mining results with linked data for learning analytics: motivation, case study and directions. *Proceedings Of International Conference On Learning Analytics & Knowledge*, 3.
- Durand, G., Laplante, F. e Kop, R. (2011). A learning design recommendation system based on markov decision processes. *Proceedings of Knowledge Discovery and Data Mining*, 17.
- Greller, W. e Drachsler, H. (2012). Translating learning into numbers: a generic framework for learning analytics. *Educational Technology & Society*, 15 (3), 42±57.
- Han, J., Kamber, M. e Pei, J. (2012). *Data mining concepts and techniques*. USA: Elsevier.
- Jayaprakash, S., Moody, E. W., Lauría, E. J. M., Regan, J. R. e Baron, J. D. (2014). Early alert of academically at-risk students: an open source analytics initiative. *Journal of Learning Analytics*, 1 (1).
- Li, N. e Matsuda, N. (2011). A machine learning approach for automatic student model discovery. *Proceedings of the International Conference on Educational Data Mining*, 4.
- Mcnelly, B. J., Gestwicki, P., Hill, J. H., Parli-Horne, P e Johnson, E. (2012). Learning analytics for collaborative writing: a prototype and case study. *Proceedings Of The International Conference On Learning Analytics And Knowledge*, 2.
- Romero, C. e Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3 (1), 12±27.
- Siemens, G. e Baker, R. S. J. D. (2012). Learning analytics and educational data mining: towards communication and collaboration. *Proceedings Of The International Conference On Learning Analytics And Knowledge*, 2.
- Siemens, G. e Long, P. (2011). Penetrating the fog: analytics in learning and education. *Educause Review*, 46 (5), 30±32.
- Wolff, A., Zdrahal, Z., Nikolov, A. e Pantucek, M. (2013). Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *Proceedings Of The International Conference On Learning Analytics And Knowledge*, 3.