

Análise do Desempenho Estudantil na Educação a Distância Aplicando Técnicas de Mineração de Dados

Hugo Guércio¹, Phillipe Marques¹, Victor Ströele¹,
Crystiam Kelle Pereira¹, Eduardo Barrere¹

¹Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora (UFJF)
CEP: 36036-900 – Juiz de Fora - MG – Brazil

{hugo.guercio,phillipeisrael,victor.stroele}@ice.ufjf.br

{crystiam.kelle,eduardo.barrere}@ice.ufjf.br

Abstract. *The availability of e-learning courses grows every year in Brazil and the use of e-learning systems supports increasingly this teaching modality. Due to the large amount of generated and stored data in e-learning systems, lots of information can be withdrawn from this large data set. Such information may assist educators and students in different learning aspects. This article aims to classify the students performance in earlier stages of discipline allowing decisions that might improve student performance. The experiments performed achieved an average accuracy of 73% in performance classification.*

Resumo. *A disponibilização de cursos na modalidade a distância no Brasil cresce a cada ano e o uso de um Ambiente Virtual de Aprendizado (AVA) auxilia cada vez mais essa modalidade de ensino. Devido a grande quantidade de dados gerados e armazenados em um AVA muitas informações podem ser retiradas desse grande conjunto de dados. Tais informações podem auxiliar educadores e alunos em diferentes aspectos do aprendizado. O presente artigo tem como objetivo classificar o desempenho dos estudantes em estágios prematuros da disciplina possibilitando decisões que possam melhorar o desempenho dos estudantes. A acurácia média alcançada nos experimentos realizados foi de 73% na classificação do desempenho.*

1. Introdução

A abrangência cada vez maior da Internet em todas as camadas da sociedade brasileira, em conjunto com programas do Governo Brasileiro que objetivam oferecer educação para todos tem proporcionado um grande avanço na modalidade de Educação a Distância (EAD). De acordo com Baker et al. [2011], o Brasil é um dos países que possui os maiores crescimentos no número de cursos ofertados na modalidade EAD.

O uso das tecnologias de informação e de comunicação tem sido um fator transformador na educação e, com isso, novos espaços virtuais de aprendizagem surgiram e um deles merece destaque: os Ambientes Virtuais de Aprendizagem – AVA.

O uso de AVA e outros *softwares* no contexto acadêmico produziu um grande repositório de *dados educacionais*. Esses repositórios possuem informações sobre alunos, professores, disciplinas, dentre outras, permitindo que sejam feitas diversas análises com o intuito de aprimorar ainda mais a Educação a Distância.

Como descrito por Han et al. [2006], Mineração de Dados (do inglês, *Data Mining*, ou DM) se refere ao processo de extrair conhecimento a partir de grandes bases de dados. Dentro deste contexto, surgiu uma nova área de pesquisa conhecida como Mineração de Dados Educacionais (MDE). A MDE é definida como a área de pesquisa que tem como principal foco o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais [Baker et al., 2011]. Assim, neste trabalho, serão utilizadas técnicas de MDE para definir e analisar o perfil dos alunos dos cursos de EAD.

Sistemas baseados em *e-learning*, ou *web-based education* (modelo para ensino a distância suportado por tecnologia) são sistemas que oferecem cursos *on-line*. Eles podem oferecer uma grande variedade de canais e espaços de trabalho, facilitando assim o compartilhamento de informação e comunicação entre os participantes de um curso [Castro et al., 2007]. Estes sistemas permitem aos educadores distribuir informações para os alunos, produzir material de conteúdo, preparar atividades e testes, participar de discussões, gerenciar classes a distância e permitir a aprendizagem colaborativa com fóruns, chats, áreas de armazenamento de arquivos, serviços de notícias, etc [Romero et al., 2008].

Em *e-learning*, o professor pode ter a necessidade de escolher o melhor conteúdo que será passado para cada tipo potencial de usuário, pois diferentes alunos possuem diferentes necessidades e esta decisão pode ser complexa. Para avaliar a real necessidade de determinado aluno, este trabalho propõe a criação de um modelo conceitual que auxilie o professor na análise do comportamento dos alunos em diferentes períodos de tempo de determinada disciplina, permitindo assim, avaliar se um aluno possui tendência a aprovação ou reprovação antes do término da disciplina.

O trabalho está organizado como segue: Na seção 2 são analisados trabalhos correlatos. A seção 3 tem o objetivo de contextualizar os experimentos. Na seção 4 são descritos os experimentos realizados. A seção 5 apresenta os resultados e suas respectivas análises. Conclusões e possíveis trabalhos futuros são apresentadas na seção 6.

2. Trabalhos Relacionados

A MDE pode auxiliar diversas tarefas com diferentes objetivos. Alguns trabalhos pretendem fornecer *feedback* para auxiliar a tomada de decisão por parte de professores e tutores [Romero et al., 2013] e [Feng et al., 2009]. Outra aplicação pode ser a de recomendações para estudantes, este conteúdo pode ser bem diversificado como por exemplo tarefas ou materiais de estudo. O trabalho realizado por Barcellos et al. [2007] tem como objetivo indicar artigos para o usuário com base no seu perfil.

Também podemos aplicar a MDE para identificação de comportamentos indesejados como falta de motivação [Hershkovitz and Nachmias, 2008], propensão a desistência [Dekker et al., 2009] ou reprovação [Bresfelean et al., 2008] ou tentativas de burlar processos de avaliação [Baker et al., 2006]. Agrupar estudantes de acordo com seus perfis e habilidades de maneira a adaptar-se ao ambiente de ensino e ao conteúdo para melhorar a aprendizagem do grupo é o objetivo descrito em [Ayers et al., 2009].

Como visto em [Chen et al., 2008] e [Chen and Sue, 2013], pode-se utilizar a MDE no auxílio da criação de mapas de conceitos ou na construção do material didático facilitando o reuso de recursos entre diferentes sistemas e usuários.

Segundo Witten and Frank [2005], técnicas de classificação tem por objetivo desenvolver modelos que possam inferir um aspecto particular dos dados através de algumas combinações de outros aspectos destes dados.

Romero et al. [2008] apresentam um estudo evidenciando a possibilidade de utilização de diversas técnicas de Mineração de Dados tais como, classificação, agrupamento e regras de associação para extração de informações relevantes sobre estudantes no contexto educacional. Neste estudo, demonstram que ferramentas de mineração de dados não comerciais podem ser utilizadas para observar o comportamento de estudantes, de maneira a permitir a determinado instrutor obter uma visão geral em relação ao uso do ambiente de educação *online*.

Técnicas de Mineração de Dados educacionais, segundo Baker et al. [2011], tem sido frequentemente utilizadas para: (1) fornecer suporte e mensagens de feedback a professores; (2) recomendar a estudantes; (3) identificar grupos de estudantes com características comuns; (4) avaliar precisão do desempenho ou risco de evasão.

Conti and Charão [2013] realizam um estudo no processo de descoberta de informação com foco em análise de prazos e datas efetivas de submissões de tarefas em um ambiente virtual de aprendizagem, utilizando algoritmos de mineração de dados da ferramenta *Weka*, uma ferramenta implementada em Java seguindo uma abordagem de *framework* [Hall et al., 2009]. Nove algoritmos foram executados e, dentre eles, o algoritmo J48 obteve maior acurácia.

Em [Gottardo et al., 2014], uma análise de técnicas de mineração de dados é realizada na formação de inferências relativas ao desempenho de estudantes em cursos a distância, baseando-se em dados obtidos de um Ambiente Virtual de Aprendizagem.

O trabalho realizado por Rigo et al. [2014] tem como principal objetivo estudar fatores relacionados com a evasão escolar através da aplicação de técnicas de MDE e *Learning Analytics*. Um sistema foi desenvolvido para facilitar a coleta e integração de dados além de disponibilizar diferentes visualizações dos mesmos.

Ifenthaler and Widanapathirana [2014] desenvolveram um *framework* que objetiva analisar o aprendizado de estudantes, identificando alternativas aproximadas para analisar os dados e determinando a validade dos perfis analisados, baseando-se em diversos tipos de informações educacionais. O *framework* permite assim, identificar três tipos de perfis: (1) perfil do estudante, (2) perfil de aprendizagem (3) perfil curricular. Dados estruturados e não estruturados de todos os sistemas são combinados e processados em uma base de dados de múltiplas camadas usando algoritmos adaptativos.

3. Descrição dos Experimentos

Com o objetivo de se conhecer o rendimento do aluno durante a realização de uma disciplina em determinado período de tempo, realizou-se o processo KDD (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento em Banco de Dados) em uma base de dados *Moodle* da Universidade Federal de Juiz de Fora (UFJF). A estrutura dessa base de dados é composta por tabelas, que armazenam informações sobre cursos, usuários, acessos, atividades, avaliações, etc.

Tabelas com dados sobre o tipo de usuário e suas postagens e acessos foram utilizadas. Além disso, foi necessário criar uma nova tabela com o intuito de associar os dados

contidos no *Moodle* com os dados do SIGA que é o sistema integrado de gerenciamento acadêmico utilizado na Universidade Federal de Juiz de Fora.

Visando a criação de um estudo de caso válido para ser utilizado em diferentes disciplinas, desenvolvemos um modelo conceitual utilizando as disciplinas de “Algoritmos”, “Inglês Instrumental” e “Língua Portuguesa” do curso a distância de Licenciatura em Computação. Como critérios para seleção de tais disciplinas foram analisados o número de estudantes que concluíram as disciplinas, a quantidade de ofertas para disciplinas e a disponibilidade dos recursos.

Foram criados três conjuntos de treinamento, divididos por tempo de acordo com o período de duração das disciplinas. No caso, os conjuntos foram divididos nos períodos de 6, 12 e 18 semanas após o início do curso, para avaliar o comportamento dos alunos durante esses intervalos de tempo.

Para cada um dos períodos de tempo descritos, foram analisados o total de acessos a plataforma e a interação ocorrida nos fóruns. A análise de acessos de alunos foi realizada selecionando o total de acessos de alunos na disciplina assim como o total de acessos a fóruns, recursos e atividades. Para a análise da interação entre alunos e tutores, foram selecionadas a quantidade de postagens criadas por cada aluno e o total dessas postagens que foram respondidas por alunos e tutores. A tabela 1 descreve de maneira simplificada as dimensões e os atributos utilizados para avaliar o desempenho dos alunos.

Tabela 1. Descrição de dimensões e atributos utilizados para avaliação do desempenho

| Dimensão | Atributo Analisado |
|----------------------|---|
| Acessos a plataforma | Total de acessos a atividades |
| | Total de acessos a fóruns |
| | Total de acessos a recursos |
| | Nota final obtida na disciplina |
| Interações no fórum | Total de postagens realizadas |
| | Total de postagens realizadas respondidas por alunos |
| | Total de postagens realizadas respondidas por tutores |
| | Nota final obtida na disciplina |

Em algoritmos de classificação é necessário que os valores contínuos sejam transformados em valores discretos (discretização). No processo de transformação dos dados, os valores das notas finais obtidas pelos alunos necessitavam ser transformados em valores discretos. Nesta transformação, foram criadas 3 classes, cada uma contendo um intervalo de notas, conforme a tabela 2.

Tabela 2. Classes utilizadas para discretização das notas

| Nota | Classe |
|-------------|--------|
| < 46 | 1 |
| ≥ 46 e < 60 | 2 |
| ≥ 60 | 3 |

As classes 1, 2 e 3 representam, respectivamente, alunos com maior tendência a

reprovação, alunos com tendência a reprovação que possuem potencial para serem aprovados e alunos com tendência a aprovação.

A ferramenta Weka [Hall et al., 2009] foi utilizada para a realização dos experimentos. Os classificadores escolhidos foram os da seção de árvores de decisão. Essa seção foi escolhida por possuir um modelo de representação intuitiva, facilitando a interpretação dos resultados obtidos. Os testes foram realizados com os algoritmos *Random Forest*, *Random Tree* e J48.

O algoritmo J48 gera árvores, podadas ou não, baseado-se no algoritmo C4.5 desenvolvido por Quinlan [1993] e utilizado por Bravo and Ortigosa [2009] para identificação de sintomas referentes ao baixo desempenho estudantil. O algoritmo J48 também foi utilizado por Bresfelean et al. [2008] na identificação de alunos com perfil de reprovação acadêmica.

O algoritmo *Random Tree* foi utilizado por apresentar resultados de fácil entendimento por educadores sem familiaridade com DM. Além disso, Srimani and Patil [2012] encontraram resultados expressivos com acurácias superiores a 90%. Segundo Yadav et al. [2012] o algoritmo é utilizado em apenas 4,5% dos estudos e, por isso, consideramos interessante aplicar o algoritmo para análise e comparação dos resultados.

Gottardo et al. [2014] utilizaram o algoritmo *Random Forest* para estimar o desempenho acadêmico dos estudantes em cursos EAD. O trabalho realizado encontrou resultados com acurácia entre 73% e 80%. Além disso, a acurácia encontrada pelo algoritmo *Random Forest* foi comparada a acurácia obtida pelo algoritmo *Multilayer Perceptron* que implementa a técnica de Redes Neurais Artificiais Diretas.

4. Experimentos

Analisar se um aluno está com tendência de aprovação ou reprovação em uma disciplina baseando-se na quantidade de acessos realizados e em como o aluno interage nos fóruns da disciplina, é o foco principal deste trabalho. Busca-se desta maneira, verificar como esses fatores, quando analisados no decorrer da disciplina, influenciam a nota final do aluno.

As disciplinas de Língua Portuguesa, Algoritmos e Inglês Instrumental, possuem, respectivamente, 281, 183 e 154 instâncias cada uma. Com os dados selecionados referentes a cada período e disciplina foi iniciado o processo de mineração de dados.

Para cada uma das disciplinas selecionadas, dois experimentos foram realizados. A intenção é analisar cada uma das etapas do curso (divididas em 6, 12 e 18 semanas) e estudar o comportamento do algoritmo em cada uma delas.

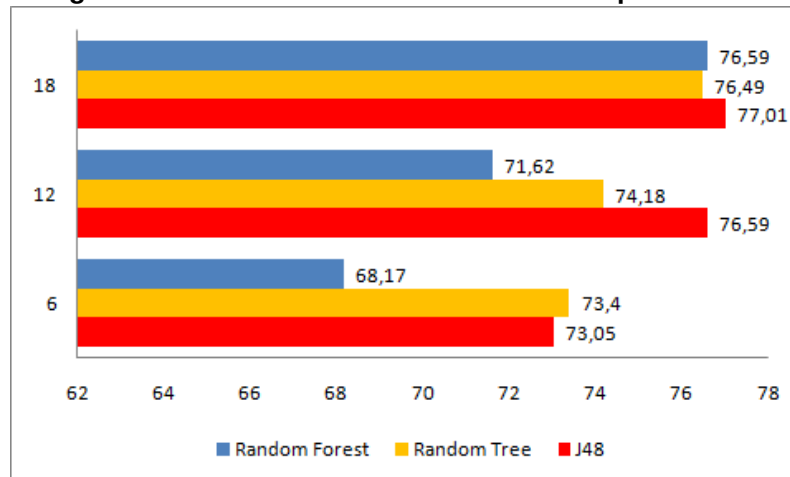
O primeiro experimento foi realizado para avaliar o comportamento do aluno no ambiente de aprendizagem, analisando quais fatores possuem maior influência no desempenho dos mesmos na disciplina. O segundo, por sua vez, foi realizado para observar como ocorrem as interações dentro dos fóruns e quais fatores são mais importantes para associar a interação com o desempenho final obtido pelo aluno. Os atributos utilizados podem ser encontrados na tabela 1 presente na seção anterior.

Para avaliar a precisão da predição realizada por cada um dos algoritmos, utilizamos o método *Cross-Validation*, descrito em [Witten and Frank, 2005]. Esse método

é uma técnica de avaliação, onde se define um número de *folds*, ou partições de dados. Utilizamos o valor 10 para partições, utilizando a evidência teórica de que este é o melhor valor para estimativa de erros. A escolha do número de partições dividiu cada conjunto de dados em 10 partições aproximadamente iguais, onde uma parte dos dados é utilizada para testes e outra é utilizada para treinamento.

O gráfico apresentado na figura 1 destaca a média dos resultados obtido no primeiro experimento, demonstrando a precisão dos algoritmos para predizer o desempenho final de um aluno em cada um dos períodos analisados. Os valores 6, 12 e 18, apresentados à esquerda da figura 1, representam as semanas avaliadas pelos algoritmos. O resultado apresentado, no eixo horizontal, corresponde a um valor médio obtido para as 3 disciplinas analisadas. Os valores demonstram a porcentagem de acertos obtidos para análise da quantidade de acessos a tarefas, fóruns e atividades, quando relacionadas com a nota final de determinado aluno.

Figura 1. Média dos resultados obtidos no experimento 1



Na figura 2, por sua vez, podemos analisar a precisão dos algoritmos no experimento 2. Os mesmos períodos de tempo foram analisados e são representados pelo eixo vertical do gráfico com escala pela quantidade de semanas. O valor médio da acurácia dos algoritmos, executados para as disciplinas escolhidas previamente, é explicitado pelo valor presente à direita de cada barra, sendo o eixo horizontal responsável pela apresentação dos valores.

Nas tabelas 3 e 4, são mostradas as médias obtidas nos experimentos realizados, além do valor de desvio padrão calculado. O desvio padrão é uma medida utilizada para medir a quantidade de variação (dispersão) dos dados dentro da amostra em relação a média. Um baixo valor para desvio padrão indica que a amostra tende a ser mais homogênea.

5. Análise

Os resultados obtidos demonstram que os algoritmos *Random Tree*, *Random Forest* e *J48* podem ser utilizados para realizar inferências em relação ao desempenho dos alunos. Comparando-se em uma determinada disciplina o total de acessos de um aluno e as interações entre alunos e tutores, em determinado período de tempo, com a nota final

Figura 2. Média dos resultados obtidos no experimento 2

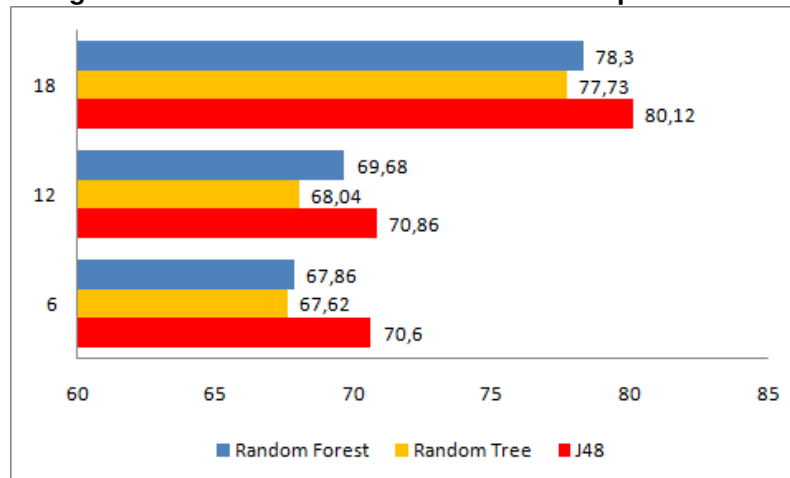


Tabela 3. Acurácia média e desvio padrão nas execuções do experimento 1

| Classificador | 6 semanas | 12 semanas | 18 semanas |
|---------------|--------------|--------------|--------------|
| J48 | 73.05 ± 7.16 | 76.59 ± 7.67 | 77.01 ± 3.88 |
| Random Tree | 73.40 ± 6.39 | 74.18 ± 6.50 | 76.49 ± 5.72 |
| Random Forest | 68.17 ± 6.00 | 71.62 ± 5.73 | 76.59 ± 8.12 |

obtida pelo aluno na disciplina. A capacidade de se realizar inferências é possível devido a acurácia obtida pelos algoritmos citados.

De acordo com os resultados analisados, a acurácia obtida pelo primeiro experimento foi ligeiramente superior se comparada ao segundo experimento, demonstrando assim a importância das informações relativas ao acesso do AVA na descoberta de padrões. A acurácia média dos algoritmos foi considerada significativa visto que os algoritmos receberam os mesmos parâmetros para as diferentes disciplinas estudadas.

Avaliamos que, para o primeiro experimento (análise de acessos), existe uma baixa dispersão dos dados em relação à média para todos os períodos analisados devido a maior homogeneidade dos dados.

No caso do segundo experimento, verificamos uma maior dispersão amostral em relação à média para o período correspondente a 18 semanas, isso foi causado devido ao grande número de postagens ocorridas nesta etapa do curso em uma das disciplinas analisadas. Com esse crescimento na quantidade de postagens, os algoritmos foram capazes de classificar de maneira mais eficiente as amostras e, com isso, obter melhores resultados se comparados aos resultados encontrados no mesmo período analisado nas outras disciplinas.

Observando as árvores geradas pelos algoritmos J48 e *Random Tree*, foi possível verificar também, quais fatores se mostram mais importantes para realizar inferências sobre o desempenho dos alunos.

Para a disciplina de Língua Portuguesa, quando analisada no experimento 1, ambos os algoritmos utilizaram o número de acessos aos recursos como principal fator de

Tabela 4. Acurácia média e desvio padrão nas execuções do experimento 2

| Classificador | 6 semanas | 12 semanas | 18 semanas |
|---------------|---------------|--------------|---------------|
| J48 | 70.6 ± 7.61 | 70.86 ± 7.53 | 80.12 ± 17.71 |
| Random Tree | 67.62 ± 9.84 | 68.04 ± 6.77 | 77.73 ± 22.74 |
| Random Forest | 67.86 ± 10.24 | 69.68 ± 7.44 | 78.3 ± 20.57 |

regras para a geração da árvore e, para o experimento 2, o principal fator verificado foi o número de postagens criadas pelo aluno.

A disciplina de Algoritmos, quando analisada no primeiro experimento, evidenciou que de acordo com os dois algoritmos, a quantidade de acesso às atividades se mostrou mais relevante. O algoritmo Random Tree também utilizou a quantidade de acesso aos recursos disponíveis como característica importante na classificação das amostras. O segundo experimento também demonstrou que a quantidade de postagens criadas pelo aluno foi fator determinante para a classificação em ambos os algoritmos. Nesta disciplina o algoritmo J48 também utilizou a quantidade de respostas realizadas por alunos como nó importante na árvore de decisão.

De acordo com os resultados obtidos na análise da disciplina de Língua Inglesa, assim como na disciplina de Algoritmos, o primeiro experimento revelou o acesso às atividades como característica mais relevante para a classificação em ambos os algoritmos. Assim como na disciplina de Algoritmos, a quantidade de acessos aos recursos disponíveis foi relevante na classificação. Observando os resultados obtidos no segundo experimento os dois algoritmos geraram árvores onde a característica mais relevante para classificação foi a quantidade de postagens criadas pelo aluno.

6. Conclusões e Trabalhos Futuros

Acompanhar individualmente os alunos nos cursos realizados em ambientes virtuais se mostra um fator de extrema importância, permitindo melhorar a qualidade dos cursos oferecidos nas instituições de ensino, além de ampliar a possibilidade de aumento do número de aprovações em decorrência de um melhor aproveitamento dos recursos disponíveis. Esta necessidade, nos motivou a buscar diferentes técnicas, capazes de serem aplicadas para auxiliar os professores no gerenciamento das disciplinas oferecidas na modalidade EAD.

A partir da análise do desempenho de um aluno em um período de tempo pré-determinado, um professor será capaz de acompanhar seu rendimento, verificando se ele possui potencial para ser aprovado ou não. Tendo esse conhecimento prévio, o professor poderá avaliar as necessidades individuais do aluno e agir de acordo com o que acreditar ser mais relevante para que o mesmo possa ser aprovado.

Neste trabalho observamos que é possível transformar os dados armazenados na base de dados da plataforma *Moodle* em conhecimento, umas das características do processo de KDD, gerando assim regras que são extremamente úteis para apoio a tomada de decisões.

Como observado a partir da análise dos resultados, de maneira geral, o atributo mais relevante para a classificação foi o acesso às atividades no caso de análise dos acessos

à plataforma. Na análise de interações no fórum o atributo mais relevante foi a quantidade de postagens realizadas pelo aluno.

Como Trabalhos futuros, pretendemos identificar novos conjuntos de atributos para o processo de mineração, assim como a validar dos resultados obtidos em disciplinas em curso. Também pretendemos identificar, com o auxílio de educadores e alunos, características importantes em cada disciplina visto que de acordo com Romero and Ventura [2010], algumas características pertinentes a MDE podem necessitar de ajustes individuais para melhores resultados.

Referências

- Elizabeth Ayers, Rebecca Nugent, and Nema Dean. A comparison of student skill knowledge estimates. *International Working Group on Educational Data Mining*, 2009.
- Ryan Shaun Joazeiro Baker, Seiji Isotani, and Adriana Maria Joazeiro Baker de Carvalho. Mineração de dados educacionais: Oportunidades para o Brasil. In *Revista Brasileira de Informática na Educação, Volume 19, Número 2*, 2011.
- Ryan SJ Baker, Albert T Corbett, Kenneth R Koedinger, Shelley Evenson, Ido Roll, Angela Z Wagner, Meghan Naim, Jay Raspat, Daniel J Baker, and Joseph E Beck. Adapting to when students game an intelligent tutoring system. In *Intelligent Tutoring Systems*, pages 392–401. Springer, 2006.
- Carla Duarte Barcellos, Daniela Leal Musa, André Luiz Brandão, and Mariusa Warpechowski. Sistema de recomendação acadêmico para apoio a aprendizagem. *RENOTE*, 5(2), 2007.
- Javier Bravo and Alvaro Ortigosa. Detecting symptoms of low performance using production rules. *International Working Group on Educational Data Mining*, 2009.
- Vasile Paul Bresflean, Mihaela Bresflean, Nicolae Ghisoiu, and C-A Comes. Determining students' academic failure profile founded on data mining methods. In *Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on*, pages 317–322. IEEE, 2008.
- Félix Castro, Alfredo Vellido, Àngela Nebot, and Francisco Mugica. Applying data mining techniques to e-learning problems. In *Evolution of teaching and learning paradigms in intelligent environment*, pages 183–221. Springer, 2007.
- Nian-Shing Chen, Chun-Wang Wei, Hong-Jhe Chen, et al. Mining e-learning domain concept map from academic articles. *Computers & Education*, 50(3):1009–1021, 2008.
- Shyi-Ming Chen and Po-Jui Sue. Constructing concept maps for adaptive learning systems based on data mining techniques. *Expert Systems with Applications*, 40(7): 2746 – 2755, 2013. ISSN 0957-4174. doi: <http://dx.doi.org/10.1016/j.eswa.2012.11.018>. URL <http://www.sciencedirect.com/science/article/pii/S0957417412012286>.
- Fabieli Conti and Andrea Schwertner Charão. Análise de prazos de entrega de atividades no moodle: um estudo de caso utilizando mineração de dados. In *Revista Renote, Volume 9, Número 2*, 2013.
- Gerben W Dekker, Mykola Pechenizkiy, and Jan M Vleeshouwers. Predicting students drop out: A case study. *International Working Group on Educational Data Mining*, 2009.
- Mingyu Feng, Joseph E Beck, and Neil T Heffernan. Using learning decomposition and bootstrapping with randomization to compare the impact of different educational

- interventions on learning. *International Working Group on Educational Data Mining*, 2009.
- Ernani Gottardo, Celso Antônio Alves Kaestner, and Robinson Vida Noronha. Estimativa de desempenho acadêmico de estudantes: Análise da aplicação de técnicas de mineração de dados em cursos a distância. *Revista Brasileira de Informática na Educação*, 22(01):45, 2014.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- Jiawei Han, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques*. Morgan kaufmann, 2006.
- Arnon HersHKovitz and Rafi Nachmias. Developing a log-based motivation measuring tool. In *EDM*, pages 226–233, 2008.
- Dirk Ifenthaler and Chathuranga Widanapathirana. Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning*, 19(1-2):221–240, 2014. ISSN 2211-1662. doi: 10.1007/s10758-014-9226-4. URL <http://dx.doi.org/10.1007/s10758-014-9226-4>.
- John Ross Quinlan. *C4. 5: programs for machine learning*, volume 1. Morgan kaufmann, 1993.
- Sandro José Rigo, Wagner Cambuzzi, Jorge LV Barbosa, and Sílvia C Cazella. Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. *Revista Brasileira de Informática na Educação*, 22(01):132, 2014.
- Cristóbal Romero and Sebastián Ventura. Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6):601–618, 2010.
- Cristóbal Romero, Sebastián Ventura, and Enrique García. Data mining in course management systems: Moodle case study and tutorial. *Comput. Educ.*, 51(1):368–384, August 2008. ISSN 0360-1315. doi: 10.1016/j.compedu.2007.05.016. URL <http://dx.doi.org/10.1016/j.compedu.2007.05.016>.
- Cristóbal Romero, Amelia Zafra, Jose María Luna, and Sebastián Ventura. Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data. *Expert Systems*, 30(2):162–172, 2013. ISSN 1468-0394. doi: 10.1111/j.1468-0394.2012.00627.x. URL <http://dx.doi.org/10.1111/j.1468-0394.2012.00627.x>.
- PK Srimani and Malini M Patil. A classification model for edu-mining. In *PSRC-ICICS Conference Proceedings*, pages 35–40, 2012.
- Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005. ISBN 0120884070.
- Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal. Data mining applications: A comparative study for predicting student’s performance. *arXiv preprint arXiv:1202.4815*, 2012.