

ANALISE E PREDIÇÃO DE DESEMPENHO DE DISCENTES A PARTIR DA INSERÇÃO DE MATERIAL DE APOIO EM AVAS

**Luciana da Conceição
Ferreira Mendes**

Universidade Estadual do
Maranhão
São Luis, Brasil
lucianamendes@nti.uema.br

Reinaldo de Jesus da Siva
Universidade Estadual do

Maranhão
São Luis, Brasil
reinaldo.silvarrrb@gmail.com

Luis Carlos Costa Fonseca
Universidade Estadual do

Maranhão
São Luis, Brasil
lccfonseca@gmail.com

RESUMO

A utilização Ambientes Virtuais de Aprendizagem (AVA) não se limita apenas a educação a distância, mas também auxilia na educação presencial, agindo como uma ferramenta a mais no processo de aprendizagem. No entanto, gerenciar o processo de aprendizagem nos AVAs com qualidade exige cada vez mais do professor que precisa utilizar materiais atualizados no aprendizado dos alunos. Neste sentido, este trabalho apresenta um modelo preditivo de Mineração de Dados (MD) em um AVA, a partir da inserção de material didático por parte dos professores no ensino presencial. O objetivo foi verificar o impacto que a não inserção de material tem no desempenho de alunos do ensino presencial que utilizam AVA como extensão da sala de aula. Para isso, foram realizados experimentos com conjuntos de dados distintos, onde as técnicas de mineração de dados Redes Bayesianas e Árvore de decisão foram aplicadas, sendo comparado ao final o desempenho de cada um.

Palavras chave

Baixo desempenho; mineração de dados; diagnóstico; AVA; material de apoio.

INTRODUÇÃO

As interações que os alunos têm entre si, com os professores e com recursos educacionais são indicadores valiosos da eficácia de uma experiência de aprendizagem. Mas com a crescente variedade de ferramentas que estão sendo usadas, o monitoramento do progresso dos alunos está se tornando um desafio [6].

Um fator importante que contribui para a eficácia de uma experiência de aprendizagem é a capacidade dos professores de monitorar o processo de aprendizagem e potencialmente atuar com base nos eventos observados [7]

Dessa forma, pensando no impacto negativo que o não acompanhamento do desempenho dos alunos e a consequente não atuação do professor diante dos eventos, o que pode acarretar não só para os estudantes, mas também para as Universidade e mesmo para o país, propõe-se neste artigo analisar o impacto da não inserção de material didático ou apoio ao aprendizado nos ambientes virtuais de aprendizagem (AVA) no desempenho dos alunos da graduação presencial usando as técnicas de árvore de decisão e redes bayesianas .

AMBIENTE VIRTUAL DE APRENDIZAGEM

Com o desenvolvimento e crescimento da Web, houve grande investimento na construção de plataformas, que possibilitassem transpor a sala de aula para o meio virtual. Dessa forma, surgem os AVAs espelhados em práticas tradicionais e com os seguintes fins: oferta de curso a distância e uso de mediação tecnológica para diminuir os encontros presencial [2].

Kalinke (2014, p.74), define AVA como: “novos espaços destinados à aprendizagem e nos quais ela pode ser favorecida. São espaços com características próprias e que permitem novas formas e encaminhamentos aos processos de ensino e aprendizagem”.

É importante que haja interações em AVAs, pois a colaboração entre professor e aluno ajuda a desenvolver soluções para possíveis problemas cognitivos implícitos na interação e na comunicação que aluno e professor teriam pessoalmente.

Sendo assim, a função que o professor desempenha no contexto educacional é de extrema importância, pois ele é o mediador entre os alunos e os objetos de conhecimento. O professor também tem papel de organizador do ambiente de aprendizagem, pois mesmo que os alunos sejam estimulados a buscarem por material de apoio, a intervenção do professor nesse processo é fundamental na avaliação da qualidade do conteúdo.

Dentre as ferramentas comumente encontradas em um ambiente virtual de aprendizagem (AVA), podem-se citar: fóruns, envio de mensagens instantâneas, compartilhamento de arquivos e aplicativos, agenda, gestão de turmas, gestão de grupos, gestão de usuários, sistema de avaliação, questionários e banco de questões. Não sendo necessário que um ambiente virtual tenha todas essas características para ser considerado um AVA.

Turma virtual

A Turma Virtual é uma ferramenta de ensino complementar colocada à disposição dos docentes e discentes. Ela é um espaço construído para ajudar no aprendizado dos discentes, criando uma extensão da sala de aula no Sistema Integrado de Gestão Acadêmica da Universidade Estadual do Maranhão (SigUema). Criada e desenvolvida pela Universidade Federal do Rio Grande do Norte, a ferramenta encontra-se nos Portais do Docente e do Discente,

permitindo o intercâmbio virtual de informações entre discentes e docentes de uma turma.

Na turma virtual o docente pode cadastrar o plano de curso, o cronograma de aulas, as avaliações, referências bibliográficas, conteúdo programático, criar fórum e chat para a turma, lançar a frequência dos discentes, gerenciar grupos, imprimir diário de turma e lista de presença, cadastrar materiais para disponibilizar para os discentes, cadastrar atividades e questionários para que os discentes resolvam, visualizar estatísticas de notas, alunos e acessos etc. Além dessas e de outras funcionalidades, o docente também pode, na Turma Virtual, efetuar o lançamento de notas ou conceitos e fechar a turma.

Trabalhos relacionados

Na literatura encontram-se algumas iniciativas de MD em AVA's, sendo, que visam analisar o desempenho de alunos e consequentemente descobrir padrões de comportamento a partir das interações dos alunos nesses ambientes. Dentre elas, as que mais se aproximam da metodologia proposta neste trabalho são as que seguem.

Silva (2015), apresenta o desenvolvimento de um modelo preditivo de MD em um AVA, a partir das interações de alunos em fóruns de discussão. O objetivo era realizar o diagnóstico de baixo desempenho de alunos, que é considerado um forte indicio para evasão, gerando relatórios que auxiliem as partes interessadas na tomada de decisão. A técnica de MD foi aplicada através de cinco algoritmos de classificação: J48, BFTree, SimpleCart, Bayesianos e BayesNet, sendo comparado o desempenho de cada um, a fim de que um modelo com melhor desempenho fosse obtido. O algoritmo J48 alcançou melhor desempenho (73,96%).

Através de técnicas de MD, Kampff et al. (2014) buscam identificar perfis de alunos com risco de evasão ou reprovação, visando à geração de alertas para sensibilizar o professor sobre possíveis problemas. Segundo os autores, nos resultados obtidos, há evidências de que o índice de evasão nas turmas com os alertas tenha sido significativamente inferior ao índice de evasão observado na amostra de dados históricos.

Santana, Maciel e Lins (2014) tiveram como objetivo realizar a avaliação da dimensão perfil de uso no ambiente Moodle. Foram utilizados dados de um curso ofertado na modalidade semipresencial extraídos do banco de dados do AVA Moodle. Foram utilizados sete algoritmos para analisar o desempenho do perfil, onde o J48 obteve o melhor desempenho, alcançando 74% de acurácia.

Guércio et al. (2014) propõem a criação de um modelo que auxilie o professor na análise do comportamento dos alunos no decorrer da oferta de determinada disciplina, no intuito de melhorar o desempenho dos estudantes.

Brito (2014) propõe a utilização de técnicas de MD para tentar prever o desempenho dos alunos no primeiro período

do curso de Ciência da Computação da UFPB, através das suas notas de ingresso no vestibular.

O estudo buscou encontrar relação entre a nota de ingresso de estudantes e o seu desempenho nas disciplinas de Cálculo Diferencial e Integral I, Física Aplicada à Computação I, Cálculo Vetorial e Geometria Analítica do primeiro período do curso de Ciência da Computação da UFPB. Através da ferramenta *Weka*, obteve-se precisão superior a 70%, utilizando um conjunto de três atributos de entrada: Média Geral, Média de Matemática e a Média de Física obtidas no processo seletivo para entrada na UFPB. O autor aponta que se tem conhecimento da existência de outras variáveis que podem influenciar o desempenho do aluno no primeiro período do curso, porém afirma que estas são muitas vezes subjetivas e difíceis de serem recuperadas, como motivação do aluno no curso, taxa de aprovação da turma, situação socioeconômica, entre outras.

Nesse sentido, o trabalho aqui proposto busca ampliar as contribuições dos trabalhos correlatos, uma vez que, insere no contexto uma variável nunca antes utilizada na análise do desempenho de discentes, buscando assim, elucidar o problema proposto.

Este trabalho faz a análise de duas variáveis, ou seja, professor e aluno. Além da abordagem ser do ponto de vista do professor analisando as atividades desenvolvidas por ele, não olhando só pelo ponto de vista das atividades de responsabilidade do aluno.

PROCEDIEMTOS METODOLÓGICOS PARA EXPERIMENTOS USANDO MINERAÇÃO DE DADOS

A seguir são apresentados os procedimentos metodológicos para os experimentos a serem realizados neste trabalho, visando gerar um modelo preditivo para diagnóstico de baixo desempenho a partir da utilização de MD.

Arquitetura do Modelo Preditivo

A metodologia deste trabalho tem foco na utilização de dados oriundos da base de dados dos cursos de graduação da Universidade Estadual do Maranhão (UEMA), para diagnosticar o perfil de alunos com mau desempenho, por meio das diversas interações realizadas por eles na turma virtual.

Foram analisados 10 cursos de graduação presencial, que continham turmas com postagens de material de apoio e posteriormente comparadas com turmas ministradas pelos mesmos docentes em períodos anteriores onde não houve postagem de material. Dessa forma, foram analisadas 100 turmas com média de 45 alunos cada.

A figura 1 mostra a arquitetura do modelo computacional proposto neste trabalho baseados em Silva (2015).

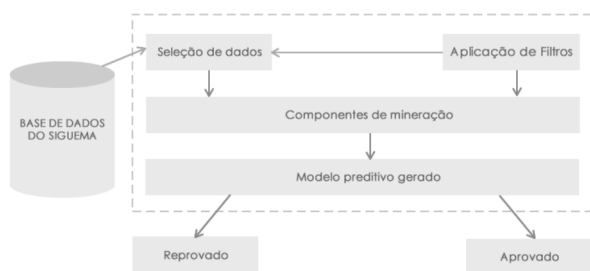


Figure 1. Arquitetura do modelo preditivo para diagnóstico de desempenho. Fonte: adaptado de Silva (2015)

O primeiro conjunto de dados, com 450 alunos, foi obtido pela seleção realizada a partir das tabelas e atributos da base de dados do Turma Virtual que registram as interações dos alunos no download de arquivos postados pelos docentes. Dessa forma, a tabela de sumarização apresenta os atributos selecionados.

Atributo	Descrição
nota_media_em_aval	nota média do aluno nas avaliações
qtd_frequencia	quantidade de frequência do aluno
qtd_down_turma	Quantidade downloads a arquivos postado pelo docente
total_post_prof	total de postagem do professor
resultado	resultado final do aluno na disciplina (atributo classe)

Tabela 1. Atributos de Sumarização

A distribuição das classes referentes ao resultado do primeiro conjunto de dados está assim apresentada: Aprovado Por Média = 250 (55,55%), Aprovado Por Final = 98 (21,17%), Reprovado Por Média = 80 (17,77%) e Reprovado Por Final = 22 (0,4%). Estas classes foram transformadas em apenas duas, sendo que foi mantida a classe Aprovado Por Média (anteriormente chamada de A.M), enquanto que as demais foram transformadas na classe Baixo Desempenho (anteriormente chamada de B.D). Com isso, uma nova distribuição pode ser observada, conforme mostrado na tabela 2.

Tabela 2 . Distribuição de dados do primeiro conjunto de dados

Resultado	Total de Alunos	Percentual
A.M	250	55,55%
B.D	200	44,44%
TOTAL	450	100%

O procedimento foi adotado porque a aprovação por média é o que se espera de um aluno em uma situação muito boa, mas quando isso não ocorre, significa que alguns fatores

influenciaram para que o mesmo tivesse desempenho abaixo do esperado. Desta forma, as três classes foram consideradas uma só por elas representarem os alunos nesta condição e porque a análise de desempenho é o que se busca neste trabalho.

O segundo conjunto de dados foi obtido a partir do primeiro através de um filtro por linha, que consistiu em manter apenas os alunos que estavam matriculados em turmas onde houve postagem de material de apoio, o que resultou em 273 alunos nessa condição, sendo mantidos os 5 atributos do conjunto de dados anterior e a distribuição das classes referentes ao resultado ficou assim: A.M = 273 (60,66%) e B.D = 177 (39,33%), conforme se observa na tabela 3

Tabela 3. Distribuição de dados do segundo conjunto de dados

Resultado	Total de Alunos	Percentual
A.M	273	60,66%
B.D	177	39,33%
TOTAL	450	100%

Para a mineração de dados deste trabalho fez-se uso da ferramenta Weka (*Waikato Environment for Knowledge Analysis*) que é uma coleção de algoritmos de aprendizagem de máquina e ferramentas de pré-processamento. É uma ferramenta de código aberto e foi desenvolvido na Universidade de Waikato na Nova Zelândia. Weka possui uma variedade de algoritmos de aprendizagem, que incluem ferramentas de pré-processamento. Além disso, oferece suporte a todo processo de mineração, que inclui suporte a preparação dos dados de entrada, avaliação estatística da aprendizagem, visualização dos dados de entrada e os resultados. Os modelos gerados a partir de árvores de decisão utilizaram o algoritmo **J48**, e os baseados em estatística utilizaram o algoritmo **NaiveBayes**.

Em seguida foi aplicada ainda uma técnica de validação de dados, chamada **Validação Cruzada de Dez Partições** (*cross validation 10-folds*), que consiste em dividir os dados em dez partições aleatórias, onde são retiradas nove dessas partições para serem utilizadas no conjunto de treinamento e uma partição para o conjunto de testes. Dessa forma, a primeira iteração será obtida a primeira precisão do modelo. Em seguida, para cada algoritmo de classificação aplicado, mais nove iterações percorrem todas as possibilidades de escolha, que resulta em mais nove valores de precisão. A precisão final do classificador é calculada, considerando a média das precisões das dez iterações. E finalmente, para a avaliação do desempenho dos algoritmos serão utilizadas as métricas Precision (percentual de amostras classificadas corretamente sobre o total de amostras classificadas como positivas), Recall (percentual de amostras positivas classificadas corretamente sobre o total de amostras positivas) e F-measure (média ponderada de Precision e

Recall que representa uma boa métrica para avaliar qual algoritmo utilizar) (SILVA, 2015).

ANÁLISE DOS RESULTADOS

O experimento inicial utilizou dados originais e a taxa de acerto dos algoritmos ficou acima de 70%, de forma que o maior índice de precisão no experimento foi registrado pelo algoritmo J48 (73,8092%), seguido pelo algoritmo NaiveBayes (73,1236%). Os resultados obtidos mostram que para dados originais, as duas técnicas tiveram desempenho muito semelhantes, com uma leve vantagem da técnica J48 em relação às baseadas em redes bayesianas. No segundo experimento, usando dados filtrados por linha, notou-se uma inversão em relação ao experimento anterior, visto que os melhores desempenhos foram obtidos pelas técnicas baseadas em redes bayesianas (todas acima de 72%), sendo que as técnicas de árvore de decisão ficaram acima de 73% e a maior taxa de acerto no experimento foi registrada pelo algoritmo NaiveBayes, que ficou abaixo de os 73%, como mostra a tabela 4.

Tabela 4 Precisão de acerto em cada conjunto de dados (% de acerto).

Método	Algoritmo	originais	filtrados por linha
Árvores de Decisão	J48	73,8092	73,5611
Bayesianos	NaiveBayes	73,1236	72,6998

CONCLUSÃO

Este trabalho analisou como os dados armazenados em um AVA, usado como segunda sala de aula de cursos presenciais, podem ser transformados em informações potencialmente úteis para apoiar o acompanhamento de alunos em cursos presenciais. Através dos experimentos realizados buscou-se obter um modelo preditivo com alta precisão que fosse capaz de prever quando um aluno apresenta características tendenciosas ao baixo desempenho a partir de das postagens de material de apoio.

A técnica baseada em árvores de decisão J48 é recomendada no contexto educacional, uma vez que ela gera um resultado mais compreensível e fácil de interpretar ao usuário que as utilizar para a tomada de decisão, de forma que nos experimentos deste trabalho observou-se que ela obteve os melhores desempenhos. Desta forma, é a mais indicada, dentre as testadas, para a geração de um diagnóstico mais preciso das tendências de baixo desempenho de alunos da graduação presencial.

Para trabalhos futuros, objetiva-se tornar a interpretação e compreensão dos resultados do modelo obtido de mais fácil entendimento. Para tanto, apresentam-se alguns desafios: criação de uma ferramenta para mostrar o diagnóstico ao público interessado para cada um dos resultados trabalhados neste estudo; apresentar a situação individual de cada aluno e integrar essa ferramenta como um módulo do SigUema.

Importante informar ainda que esta ferramenta já está em desenvolvimento e que utiliza as regras de classificação geradas por modelos baseados em árvores de decisão. O foco deste trabalho foi especificamente na postagem de material de apoio na turma virtual por parte dos docentes.

Por meio da mineração de dados educacional é possível analisar a relação entre uma abordagem pedagógica e o aprendizado do aluno, a fim de que o professor avalie se sua abordagem realmente está ajudando ou não o aluno a ter um bom desempenho em sala de aula

REFERÊNCIAS

1. De Brito, Daniel Miranda, et al. "Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina." Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). Vol. 25. No. 1. 2014.
2. Felcher, Carla Denize Ott; Pinto, Ana Cristina Medina; Ferreira, André Luis Andrejew. O uso do facebook como ambiente virtual de aprendizagem para o ensino dos números racionais. Revista Paranaense de Educação Matemática, v. 6, n. 10, 2017
3. Guércio, Hugo et al. Análise do Desempenho Estudantil na Educação a Distância Aplicando Técnicas de Mineração de Dados. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2014. p. 641.
4. Kalinke, Marco Aurélio. Tecnologias no ensino: a linguagem matemática na web. Curitiba: CRV, 2014.
5. Kampff, Adriana Justin Cerveira, et al. "Identificação de perfis de evasão e mau desempenho para geração de alertas num contexto de educação a distância." RELATEC: Revista Latinoamericana de Tecnología Educativa. Madri, Espanha. Vol. 13, n. 2 (2014), p. 61 76 (2014).
6. Romero, Cristóbal; Ventura, Sebastián; García, Enrique. Data mining in course management systems: Moodle case study and tutorial. Computers & Education, v. 51, n. 1, p. 368-384, 2008.
7. Romero-Zaldivar, Vicente-Arturo et al. Monitoring student progress using virtual appliances: A case study. Computers & Education, v. 58, n. 4, p. 1058-1067, 2012.
8. Santana, Leandro C.; Maciel, Alexandre MA; Rodrigues, Rodrigo L. Avaliação do Perfil de Uso no Ambiente Moodle Utilizando Técnicas de Mineração de Dados. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2014. p. 269.
9. Silva, Francisco et al. Um modelo preditivo para diagnóstico de evasão baseado nas interações de alunos em fóruns de discussão. In: Brazilian Symposium on