



**11ª Jornada Científica e
Tecnológica do IFSULDEMINAS**

**& 8º Simpósio de
Pós-Graduação**

MINERAÇÃO DE DADOS EDUCACIONAIS NO ENSINO MÉDIO INTEGRADO

André L. REIS¹; Marcelo M. AURELIANO²; Juliete A. R. COSTA³;

RESUMO

Este trabalho relata o projeto de iniciação científica do curso técnico em informática integrado ao ensino médio que tem como finalidade descobrir informações sobre dados educacionais utilizando a tarefa de mineração de dados denominada classificação. Os resultados iniciais mostraram potencial para explorar outras técnicas de mineração e/ou variáveis de interesse para auxiliar o processo de ensino aprendizagem da instituição.

Palavras-chave: Mineração de Dados; Dados Educacionais; Classificação

1. INTRODUÇÃO

Grandes volumes de dados têm sido gerados no ambiente educacional e estes dados juntamente com técnicas de mineração de dados (DM) podem ser úteis para auxiliar o processo ensino-aprendizagem. Os dados mais comumente analisados são *logs* de acesso a sistemas, avaliações feitas por alunos em ambientes virtuais de aprendizagem (AVA), histórico de navegação no ambiente educacional, dentre outros. Segundo RODRIGUES (2014), usar tarefas de Mineração de Dados Educacionais (MDE) pode ser útil para descobrir novos conhecimentos.

Essa linha de pesquisa objetiva desenvolver e/ou aplicar métodos de DM sobre conjuntos de dados educacionais. Assim, é possível compreender o desempenho dos alunos, como eles aprendem, perfis de alunos, além de outros fatores que influenciam o processo de ensino-aprendizagem. A maioria das pesquisas de MDE são derivadas diretamente da MD e, conseqüentemente, de suas tarefas. Em BAKER (2011) são apresentadas as principais subáreas de pesquisas em MDE e as tarefas relacionadas, tais como: classificação, agrupamentos, mineração de regras de associação e padrões sequenciais.

A tarefa de classificação, por exemplo, pode ser utilizada em MDE para auxiliar no desenvolvimento de atividades instrucionais, pois conseguem estimar os benefícios antes da aplicação. Essa técnica consiste em prever uma determinada classe para um objeto baseado em suas características. No campo educacional, por exemplo, a aplicação dessa técnica em um conjunto de dados poderia prever o desempenho do aluno baseado nas características pessoais deste e seus

¹ Graduando em Ciências Exatas, Universidade Federal de Juiz de Fora – andreluiz.dosreis.mg@gmail.com

² Graduando em História, Faculdade de Filosofia, Letras e Ciências Humanas, USP - marcelo.miguel12@usp.br

³ Professora EBTT, IFSULDEMINAS *Campus* Carmo de Minas - juliete.costa@ifsuldeminas.edu.br

logs de acesso em sistemas. O trabalho de GOTTARDO (2013) utiliza classificação para investigar o desempenho dos estudantes no cenário de atividades em grupos. Os trabalhos de RABELO (2017) e SILVA (2017), utilizam classificação para melhorar o processo de ensino a distância e identificar perfis estudantis em sala de aula, respectivamente.

2. MATERIAL E MÉTODOS

O banco de dados utilizado neste trabalho foi constituído por dados educacionais coletados da secretaria escolar e de questionários aplicados aos alunos. A Figura 1 ilustra o processo de coleta e normalização dos dados educacionais utilizados no projeto.

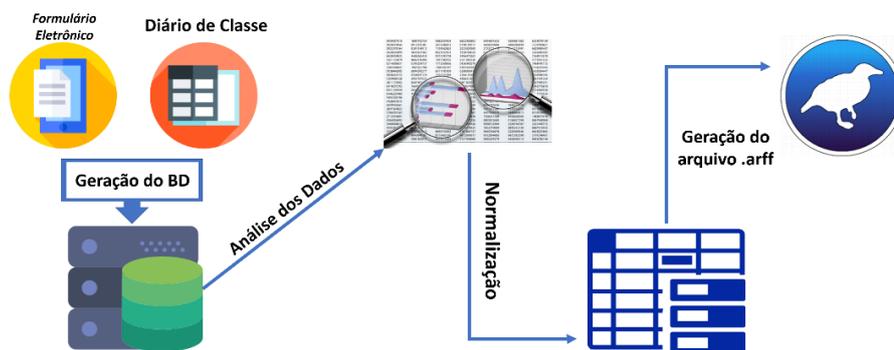


Figura 1 - Processo de descoberta da informação no cenário estudado

Inicialmente, foi desenvolvido um formulário eletrônico utilizando o *framework Bootstrap*, Linguagem de Programação PHP, Java Script e Banco de Dados (BD) MySQL. Após testes de usabilidade do formulário, este foi aplicado aos alunos participantes da pesquisa. As perguntas eram voltadas para descobrir o grau de afinidade destes para com as disciplinas cursadas e para com os professores que as lecionavam. Em paralelo, foram coletados dados dos diários de classe junto à secretaria escolar para compor a base de dados.

No segundo momento do projeto, realizou-se a integração dos dados coletados pelo formulário eletrônico e pelos diários eletrônicos. Tal integração foi feita a partir de conceitos da normalização de BD e integridade referencial entre as relações. Com o BD normalizado, efetuou-se uma análise geral selecionando os atributos de interesse para a etapa de DM. A Figura 2 ilustra uma amostra dos dados da etapa de normalização presente na Figura 1.

Serie	Sexo	Bimestre	Area	Disciplina	Afin_Disc	Afin_Prof	Plantao	Faltas	Conceito
1 Ano	M	3	CH	SOCIOLOGIA	4	5	1	0	A
1 Ano	M	3	CE	LINGUA PORTUGUESA	4	5	1	0	B
1 Ano	F	1	CH	BIOLOGIA	3	4	1	0	C
1 Ano	F	1	CH	INGLES	3	5	1	0	B
2 Ano	F	4	CE	LINGUAGEM DE PROGRAMACAO I	4	4	0	6	A
2 Ano	M	2	CH	FISICA	3	3	0	2	B

Figura 2 - Amostra dos dados analisados

O conjunto de dados normalizado consiste em uma relação com os atributos do aluno: série, sexo, bimestre, área que o aluno se identifica, disciplina, afinidade com a disciplina e com o professor que a leciona, participação no plantão de atendimento, quantidade de faltas e o conceito. O conceito foi obtido a partir da nota do aluno seguindo as seguintes regras: (1) Nota menor que 6, conceito C; (2) Nota no intervalo [6,8[, conceito B; (3) Nota maior ou igual a 8, conceito A. A partir do arquivo normalizado foi possível gerar o arquivo para testes com a ferramenta *Weka*⁴.

3. RESULTADOS E DISCUSSÕES

Para realização dos testes foi escolhida a técnica de MD classificação e, com o auxílio da ferramenta *Weka*, foram aplicados sete algoritmos diferentes. Todos os testes foram realizados em uma máquina com processador Athlon X2 B26 3.2Ghz Dual Core com 8GB de memória RAM rodando sob o sistema operacional Linux Ubuntu x64.

Escolheu-se duas medidas para avaliar os classificadores:

$$1. \text{Acurácia} = \frac{\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}}{\text{Total}}$$

que objetiva analisar quão bem o modelo treinado classifica cada instância na classe correta.

$$2. \text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}}$$

que analisa dentre os itens classificados em uma determinada classe, quantos efetivamente estão corretos.

Para calcular os valores de acurácia e precisão foram utilizadas as matrizes de confusão (veja Figura 3a fornecidas nos resultados da *Weka*. Os resultados obtidos pelos classificadores podem ser vistos na Figura 3b).

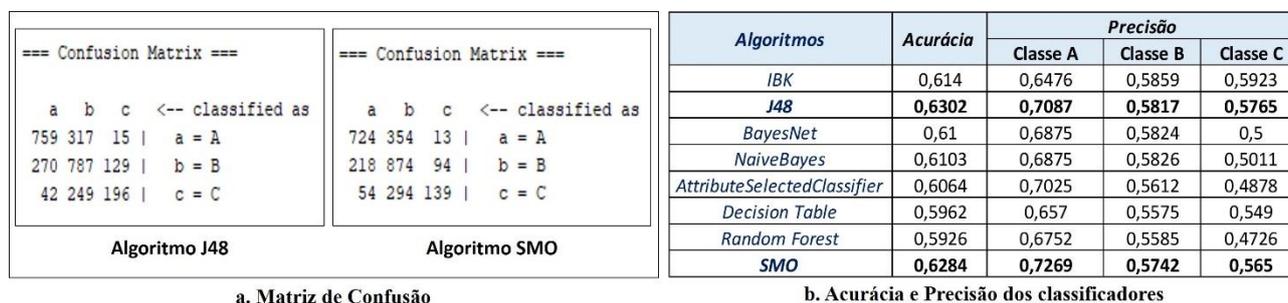


Figura 3 - Resultados dos Classificadores

Nota-se que os algoritmos J48 (Árvore de Decisão) e o SMO (*Suporte Vector Machine*) alcançam melhores resultados, com acurácia em torno de 63% e 62%, respectivamente. Para os resultados de precisão, quando o conceito=A, os classificadores conseguem melhores resultados,

⁴ <https://www.cs.waikato.ac.nz/ml/weka/>

uma vez que suas respostas ao questionário são similares. Novamente, os algoritmos se destacam nos resultados de precisão para a classe A. Portanto, eles conseguem prever o conceito do aluno com base nas características dispostas pelos valores dos atributos em aproximadamente 70% dos casos com classe=A.

4. CONCLUSÕES

A utilização de técnicas de MD no ambiente educacional tem se tornado cada vez mais comum. Com este trabalho foi possível iniciar os estudos em um projeto que visa a analisar o desempenho dos alunos no ensino técnico em informática integrado ao ensino médio. Além disso, o trabalho proporcionou aos bolsistas uma pioneira experiência na área de pesquisa acadêmica, fomentando o interesse na continuidade de novas pesquisas.

Como trabalho futuro, pretende-se melhorar as análises das variáveis que influenciam no desempenho dos alunos, buscando outras de interesse com foco nas disciplinas que os alunos têm maior dificuldade. Dessa forma, acredita-se que será possível auxiliar o processo de ensino-aprendizagem na instituição.

AGRADECIMENTOS

Agradecemos ao CNPq e ao IFSULDEMINAS pelo apoio destinado a este projeto.

REFERÊNCIAS

BAKER, Ryan; ISOTANI, Seiji; CARVALHO, Adriana. Mineração de dados educacionais: Oportunidades para o Brasil. **Brazilian Journal of Computers in Education**, v. 19, n. 02, p. 03, 2011.

GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. Aplicação de Técnicas de Mineração de Dados para Estimativa de Desempenho Acadêmico de estudantes em um AVA utilizando dados com classes desbalanceadas. In: **ICBL2013–International Conference on Interactive Computer aided Blended Learning**. 2013.

RABELO, Humberto et al. Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EaD em ambientes virtuais de aprendizagem. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2017. p. 1527.

RODRIGUES, Rodrigo Lins et al. A literatura brasileira sobre mineração de dados educacionais. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. 2014. p. 621.

SILVA, Viviane et al. Mineração de Dados Utilizando Análise de Redes Social para Identificar Tendências de Participação em Aulas Presenciais. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2017. p. 1467.