

10ª Jornada Científica e Tecnológica

7º Simpósio da Pós-Graduação

ISSN 2319-0124

PERFIL DISCENTE NO ENSINO TÉCNICO: uma análise baseada em *cluster*

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

RESUMO

Este trabalho objetiva relatar o projeto de iniciação científica desenvolvido por alunos do Curso Técnico em Informática Integrado ao Ensino Médio do campus Avançado Carmo de Minas. O projeto tem como finalidade a descoberta de informações sobre dados estudantis utilizando a tarefa de mineração de dados denominada agrupamento. A partir disso, foi possível descobrir perfis interessantes dos alunos do referido curso.

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

1. INTRODUÇÃO

O ambiente acadêmico hoje gera um grande volume de dados educacionais. Tais dados provindos de softwares educacionais e documentos de secretarias escolares podem fornecer valiosos conhecimentos às pessoas envolvidas no processo ensino-aprendizagem. De acordo com o autor Rodrigues (2014), o uso de técnicas de Mineração de Dados Educacionais (EDM) podem proporcionar a descoberta de novos conhecimentos para a comunidade escolar.

EDM é uma linha de pesquisa originada da Mineração de Dados (DM) e tem por objetivo a descoberta de informações relevantes a partir da análise de dados educacionais. A partir disso, é possível descobrir padrões sobre como ocorre o conhecimento dos alunos, perfis de grupos de alunos, variáveis que influenciam nos resultados, etc.

A maioria das técnicas de EDM são derivadas diretamente da DM e, conseqüentemente, suas tarefas. No trabalho apresentado por Backer (2011), são apresentadas as principais tarefas de mineração utilizadas em pesquisas em EDM, tais como: classificação, regressão, agrupamentos, mineração de regras de associação e padrões sequenciais.

Análise de *clusters* (agrupamentos) é uma tarefa de DM comumente aplicada no ambiente educacional (Ramos, 2016) e (Rabelo, 2017). O objetivo desta tarefa é agrupar dados de forma que as características das amostras do grupo (denominado *cluster*) sejam mais semelhantes entre si do que amostras pertencentes a outro grupo. Este trabalho tem como objetivo realizar a análise de *cluster* em dados coletados a partir de diários de classe e questionário aplicados à alunos do 1º e 2º do curso Técnico em Informática Integrado ao Ensino Médio.

¹ Bolsista PIBIC-EM/CNPq, IFSULDEMINAS Campus Carmo de Minas - andre.reis@alunos.ifsuldeminas.edu.br

² Bolsista PIBIC-EM/CNPq, IFSULDEMINAS Campus Carmo de Minas - marcelo.miguel@alunos.ifsuldeminas.edu.br

³ Orientadora, IFSULDEMINAS Campus Carmo de Minas - juliete.costa@ifsuldeminas.edu.br

2. MATERIAL E MÉTODOS

Para aplicação da técnica de DM nos dados educacionais coletados foi necessário desenvolver o processo de descoberta do conhecimento para a geração do arquivo .arff necessário.

O supracitado processo é ilustrado pela Figura 1.

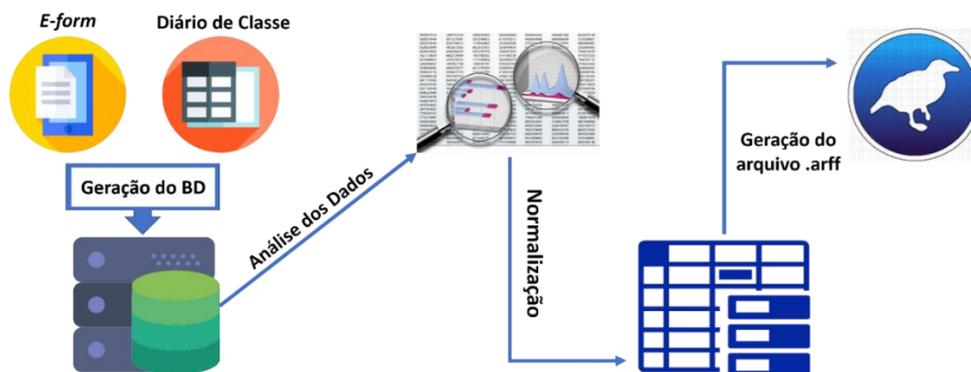


Figura 1- Processo de descoberta do conhecimento no cenário educacional trabalhado.

- **Aplicação de formulário online e coleta de registros acadêmicos:** foi desenvolvido um formulário eletrônico utilizando o framework *Bootstrap* e as linguagens de programação *Web Java Scrip* e *PHP*, onde foi avaliado numa escala de 1 a 5 o grau de afinidade do aluno com a disciplina e com o professor que a lecionava. O aluno respondeu um formulário a cada bimestre do ano de 2017. Além disso, foram coletados os diários de classe junto ao setor de registros acadêmicos para geração do banco de dados.
- **Geração do banco de dados:** com os dados coletados pelo formulário e diários de classe, foi criado um banco de dados utilizando o Sistema Gerenciador de Banco de Dados MySQL.
- **Análise dos dados coletados:** nesta etapa foi feita uma análise do novo banco de dados para selecionar os atributos relevantes à etapa de mineração de dados.
- **Normalização dos dados:** a partir das considerações feitas na etapa anterior, gerou-se um arquivo .CSV somente com os atributos de interesse.
- **Geração do arquivo .arff:** Com o arquivo .CSV foi possível gerar o arquivo necessário para mineração de dados utilizado pela ferramenta *Weka*⁴. A relação disponível no arquivo .arff possui os atributos: sexo, bimestre, área que o aluno se identifica, área a qual a disciplina pertence, afinidade do aluno para com a disciplina, afinidade do aluno para com o professor, influência do atendimento discente nos resultados escolares, número de faltas e a nota bimestral do aluno.

3. RESULTADOS E DISCUSSÕES

Para analisar os dados coletados, foi escolhida a técnica de agrupamentos (*cluster*). Com o auxílio da ferramenta Weka foram aplicados dois algoritmos: (1) o algoritmo baseado em densidade

⁴ <http://www.cs.waikato.ac.nz/ml/weka/index.html>

MakeDensityBasedClusterer e (2) o algoritmo baseado nos vizinhos mais próximos *SimpleKMeans*. 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.

Foram analisados os dados de duas turmas: 1º ano do Ensino Médio e 2º ano do Ensino Médio. As seções a seguir mostram os resultados de cada ano separadamente.

3.1. Resultados – 1º Ano

No primeiro teste fez-se a comparação entre os algoritmos (1) e (2) aplicados a um arquivo com 1624 instâncias referente aos alunos do 1º ano. A Figura 2.a mostra a comparação dos resultados obtidos, nota-se que o cluster 0, apesar de destacar bimestres diferentes há uma semelhança entre os valores dos atributos e ambos equivalem, aproximadamente, a 40% do quantitativo total. Com efeito, percebe-se que os resultados evidenciados no cluster 1, além de ponderarem o mesmo bimestre, possuem similaridade entre si nos valores dos demais atributos. Os grupos formados indicam que a maior parte de instâncias são do sexo feminino, e, embora a maioria se considere da área de Ciências Humanas (CH), obtiveram bons resultados em disciplinas da área de Tecnologia da Informação.

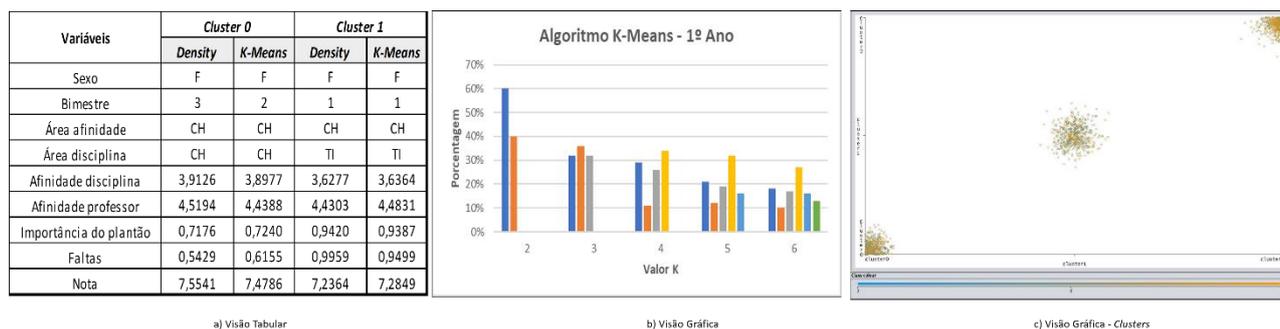


Figura 2 - Resultados 1º Ano

Outro teste realizado foi caracterizado pela variação do parâmetro K para o algoritmo *K-Means*. Com este teste foi possível analisar o comportamento do algoritmo frente a variável periódica do valor K. No gráfico ilustrado pela Figura 2.b nota-se que a distribuição mais homogênea dos clusters ocorreu quando K=3. Esta divisão pode ser vista na Figura 2.c que ilustra a segregação dos clusters e a similaridade intra-cluster. Além disso, percebe-se que o cluster 0 e cluster 2 (veja Figura 2.c) possuem mais instâncias de alunos com notas próximas ao valor 10.

3.2. Resultados – 2º Ano

Foram repetidos os mesmos experimentos para os dados referentes aos alunos do 2º ano. No primeiro teste, analisou-se os algoritmos (1) e (2) com a mesma quantidade de cluster, no qual notou-se similaridade mediana nos dados tabulares referentes ao cluster 0 e 1 apresentados na Figura 3.a. Ademais, a visão gráfica ilustrada pela Figura 3.b destaca novamente a variação de K para o algoritmo *K-Means*, onde o valor K=4 demonstra uma quantitativo similar de instâncias por

cluster. Tal situação é destacada na Figura 3.c que, além disso, mostra a separação dos clusters e a similaridade intra-cluster do valor K analisado. É possível analisar que os clusters 0 e 3 são compostos por alunos com notas entre 2,7 a 6,5, já os clusters 1 e 3 possuem mais alunos com notas mais elevadas (acima de 8).

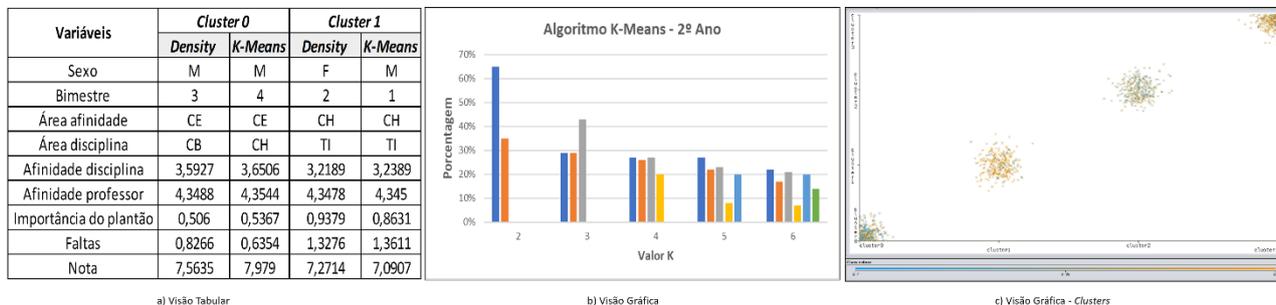


Figura 3 - Resultados 2º Ano

4. CONCLUSÕES

Com este trabalho percebeu-se a necessidade de estudos mais aprofundados na área de EDM, análises mais profundas geram novos conhecimentos, que não são visíveis apenas em registros acadêmicos, aos profissionais envolvidos no processo ensino-aprendizagem.

Além disso, o projeto desenvolvido proporcionou aos bolsistas uma pioneira experiência na área de pesquisa acadêmica durante o Ensino Médio, fomentando o interesse na continuidade de novas pesquisas e atuação na área de Tecnologia da Informação.

Como trabalho futuro, pretende-se analisar os dados das turmas de forma conjunta, aplicando, também, outras técnicas de DM neste novo conjunto de dados.

AGRADECIMENTOS

Agradecemos ao CNPq e ao *campus* Avançado Carmo de Minas pelo apoio destinado a este projeto.

REFERÊNCIAS

BAKER, R.S.J.; CARVALHO, A.M.J.B. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. *Revista Brasileira de Informática na Educação*. Volume 19, N. 2, p. 1-11, 2011.

RABELO, Humberto; BURLAMAQUI, A. M. Figueira; VALETIM, R. A. de Medeiros; RABELO, D. Silva de Souza; MEDEIROS, S. R. dos Santos. **Utilização de técnicas de Mineração de Dados Educacionais para predição de desempenho de alunos de EaD em Ambientes Virtuais de Aprendizagem**. In: *Simpósio Brasileiro de Informática na Educação*. p. 1527-1536, 2017.

RAMOS, J. L. Cavalcanti, SILVA, R. E. Dantas, SILVA, J. C. Sedraz, RODRIGUES, R. Lins and GOMES, A. Sandro, **A Comparative Study between Clustering Methods in Educational Data Mining**, in *IEEE Latin America Transactions*, vol. 14, no. 8, pp. 3755-3761, Aug. 2016.

RODRIGUES, Rodrigo Lins; RAMOS, Jorge Luis Calvacanti; SILVA, João Carlos Sedraz; GOMES, Alex Sandro. **A literatura brasileira sobre mineração de dados educacionais**. In: *Anais do Congresso Brasileiro de Informática na Educação*. 2014.

10ª Jornada Científica e Tecnológica e 7º Simpósio da Pós-Graduação do IFSULDEMINAS. ISSN: 2319-0124.