

# Uma análise do perfil e desempenho de estudantes de Computação através da Mineração de Dados

Jhonatan de Paula Candão, Eduardo Machado Real

Curso de Computação, Licenciatura – Universidade Estadual de Mato Grosso do Sul (UEMS) Nova Andradina – MS – Brasil

jpc\_suporte@hotmail.com, eduardomreal@uems.br

***Abstract.** The current educational system generates large amounts of data capable of being collected and analyzed. The Educational Data Mining can help to obtain different knowledge from information contained in this data. Thus, this paper presents and discusses the partial results of an investigation that proposes to use the Data Mining on student data. These results were obtained from experiments on two datasets, ran in three algorithms of Machine Learning, allowing the first analyzes and definition of the next steps of the proposed investigation.*

## 1. Introdução

Uma análise eficiente sobre dados de estudantes pode disponibilizar importantes informações de como está sendo conduzido o processo educacional em um curso ou mesmo em disciplinas. O fato é que tais dados podem aparecer em grande volume, a partir de, por exemplo, sistemas de informações escolares, sistemas de educação à distância ou, ainda, controles de atividades em disciplinas. Assim, uma das dificuldades é a de como obter conhecimentos a partir deles, sabendo-se que muitas das informações podem estar “ocultas” e apenas uma análise “manual” de um especialista tornaria este processo custoso e ainda poderia não ser totalmente eficiente.

A Mineração de Dados Educacionais (MDE) pode auxiliar gestores e professores, identificando diversas situações e compreendendo melhor o processo de ensino e aprendizagem. E assim, neste contexto, este trabalho apresenta os resultados parciais de um projeto de Trabalho de Conclusão de Curso que propõe utilizar a MDE em conjuntos de dados de estudantes originários de um curso de Computação: (1) de controles de disciplinas, fornecidos pelos professores e sistema acadêmico, e (2) do Exame Nacional de Desempenho de Estudantes (ENADE) [ENADE 2014].

## 2. A investigação proposta

Esta pesquisa propõe a investigação de três métodos de Aprendizado de Máquina (AM) para a MDE, a saber: (1) J48 (algoritmo C4.5, árvore de decisão) [Quinlan 1993], (2) Apriori (regras de associação) [Agrawal *et al.* 1993] e (3) K-means (agrupamento) [MacQueen 1967]. Para isso está sendo desenvolvido um sistema computacional com as implementações desses algoritmos, a partir das bibliotecas do software Weka (desenvolvido pelo *Department of CS of The University of Waikato*).

Até o momento já foram realizados os primeiros experimentos, a partir de dois conjuntos de dados: (1) de uma disciplina com “alta” taxa de evasão (30 dados),

descritos pelos atributos faltas, atividades, estudos extras, estudos em grupo e média final; (2) do ENADE (3200 dados de diversos cursos de Computação), descritos por 12 atributos selecionados do Questionário do Estudante. Em (2) foi necessária realizar a etapa de pré-processamento de dados, excluindo dados incompletos. E ainda, foi executado sobre dois ambientes, incluindo o atributo da: (1) nota bruta da prova objetiva - formação geral ou (2) nota bruta da prova objetiva – componente específico. Como resultado parcial, foi possível verificar que em boa parte das execuções dos algoritmos, dependendo do conjunto de dados e atributos selecionados, houve um bom desempenho de cada um. Por ora, a maior dificuldade ainda é a definição da quantidade suficiente de regras a serem geradas e a seleção de atributos a serem utilizados no algoritmo Apriori, bem como, em alguns casos, uma análise eficiente das regras de associação resultantes.

Como continuação deste trabalho, além da finalização do sistema proposto, serão realizados mais experimentos em outros conjuntos de dados – pelo menos em mais uma disciplina com “alta” taxa de evasão e em mais ambientes com os do ENADE. O objetivo geral então é analisar os perfis e desempenhos dos estudantes para: (1) detectar dificuldades e conter a evasão em disciplinas e no curso e (2) relacionar o perfil e desempenho de estudantes a partir do ENADE.

### **3. Considerações finais**

A análise sobre dados de estudantes pode ajudar a revelar informações que identificam questões importantes ao planejamento de ações que visem boas tomadas de decisões. Com a MDE busca-se, por exemplo, avaliar evasão, avaliar atividades realizadas, recomendar conteúdo, antever resultados, agrupar, levantar perfis, analisar desempenho, entre outras (tais como em [Paiva et al. 2012]).

### **Referências**

- Agrawal, R., Imieliński, T., e Swami, A. (1993) “Mining Association Rules between Sets of Items in Large Databases” In: Proc. Conf. on Management of Data, New York: ACM Press, p. 207–216.
- Exame Nacional de Desempenho de Estudantes (ENADE). (2014) “Microdados do Exame Nacional de Desempenho de Estudantes”, MEC/INEP, Disponível em: [http://download.inep.gov.br/microdados/Enade\\_Microdados/microdados\\_enade\\_2014.zip](http://download.inep.gov.br/microdados/Enade_Microdados/microdados_enade_2014.zip).
- Macqueen, J.B. (1967) “Some methods for classification and analysis of multivariate observations”, Proc. of the Symposium on Mathematical Statistics and Probability, p. 281–297.
- Quinlan, J. R. (1993), Programs for Machine Learning, Morgan Kaufmann Publishers, Inc. CA, USA.
- Paiva, R. Bittencourt, I. I. Pacheco, H., Silva, A. P., Jacques, P. E., Isotani, S. (2012) “Mineração de dados e a gestão inteligente da aprendizagem: desafios e direcionamentos”, Workshop de Desafios da Computação Aplicada à Educação.