

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Aplicação de algoritmos de Data Mining para a descoberta de conhecimento em bases de dados da Prova Brasil

Adriana da Rocha Silva¹

Instituto Politécnico, UERJ, Nova Friburgo, RJ

Stella Oggioni da Fonseca²

Instituto Politécnico, UERJ, Nova Friburgo, RJ

Anderson Amendoeira Namen³

Instituto Politécnico, UERJ, Nova Friburgo, RJ

1 Introdução

A Prova Brasil é uma avaliação que analisa alunos do 5º e 9º anos do ensino fundamental presentes nas escolas públicas. Esta avaliação é composta da aplicação de testes de Língua Portuguesa e Matemática, bem como questionários a alunos, professores e diretores [3]. Assim, as bases de dados que armazenam essas informações podem ser utilizadas para entender aspectos que possam influenciar no desempenho dos discentes.

O presente trabalho objetiva aplicar o processo de descoberta de conhecimento em bancos de dados (KDD - *Knowledge Discovery in Databases*) relacionados à Prova Brasil. Mais especificamente, procura-se descobrir relações entre as respostas dadas ao questionário pelos alunos e o desempenho obtido no teste de Matemática. A pesquisa focou nos dados do ano de 2013 referentes aos alunos do 9º ano residentes no Rio de Janeiro.

2 Metodologia

A metodologia consistiu em seguir o processo KDD, que constitui-se de três fases centrais: preparação dos dados, mineração de dados (*Data Mining*) e interpretação dos resultados [1]. Inicialmente, foi feito o *download* da tabela intitulada como *TS_ALUNO_9EF*, que contém as informações, em âmbito nacional, de 2720588 alunos [3]. Em seguida, por meio de códigos escritos em SQL (*Structured Query Language*), foi possível selecionar os alunos pertencentes ao Rio de Janeiro, que fizeram o teste e responderam, no mínimo, 70% das 57 perguntas presentes no questionário. Restaram para o estudo 113021 alunos. Posteriormente, a proficiência em Matemática, que é uma pontuação que varia de 0 a 500, foi categorizada conforme apresentado na Tabela 1.

¹arsilva@iprj.uerj.br²sfonseca@iprj.uerj.br³aanamen@iprj.uerj.br

Tabela 1: Categorias (ou classes) da proficiência em Matemática.

Condições	Categorias	Descrição das categorias	Número de alunos
$0 \leq proficiencia < 225$	Insuficiente	Alunos com pouquíssimo aprendizado	38938
$225 \leq proficiencia < 300$	Básico	Alunos que precisam melhorar	58859
$300 \leq proficiencia \leq 500$	Proficiente	Alunos preparados para continuar	15224

As condições presentes na Tabela 1 foram propostas pelo professor pesquisador Francisco Soares [2]. Foi, ainda, feito um processo de seleção de questões, com o intuito de identificar as que mais realçam as diferenças entre as categorias. A etapa de mineração, responsável por extrair informações, foi efetuada por intermédio do algoritmo *Naïve Bayes*, implementado dentro do *software* Weka [4].

3 Resultados e Trabalhos Futuros

Os resultados de duas questões são apresentados na Figura 1. Observa-se que foi possível corroborar o fato de que alunos com melhor desempenho gostam de estudar Matemática e pretendem continuar estudando.

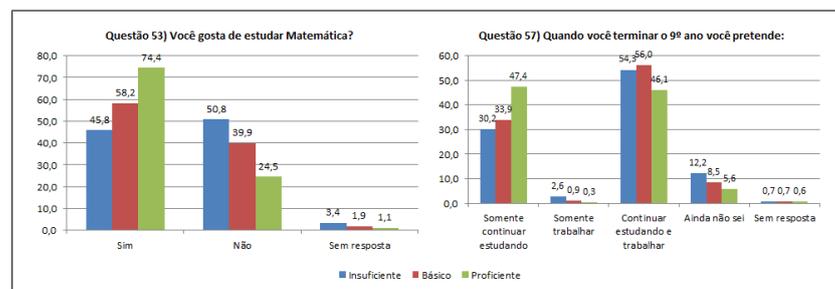


Figura 1: Percentual de alunos em cada categoria.

Outros aspectos como o aluno fazer o dever de casa e ler livros frequentemente também influenciam positivamente no desempenho. Pretende-se, como trabalhos futuros, aplicar outros algoritmos de mineração. Ademais, espera-se substituir a etapa de seleção de questões por análise fatorial (identifica variáveis inter-relacionadas).

Referências

[1] U. M. Fayyad, G. Piatetsky-Shapiro e P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*. USA, v. 17, n. 3, p. 37-54, 1996.

[2] Fundação Lemann e Meritt. *QEdu: aprendizado em foco*. 2015. Sítio na internet. Disponível em: <<http://qedu.org.br/>>. Acesso em: 20 Jun. 2016.

[3] INEP. *Microdados da Aneb e da Anresc 2013*. Brasília: Inep, 2015. Disponível em: <<http://portal.inep.gov.br/basicalevantamentos-acessar>>. Acesso em: 30 mai. 2015.

[4] I. H. Witten, E. Frank e M. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. USA: Morgan Kaufmann Publishers Inc., San Francisco, USA, 2011.