

Descoberta de Conhecimento sobre a Influência de Características Extra-Acadêmicas no Desempenho de Candidatos no Exame Nacional do Ensino Médio

Thales V. Maciel¹, Gustavo B. Seli¹

¹Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSul)
Campus Bagé – Av. Leonel de Moura Brizola, 2501 – 96408400 – Bagé – RS – Brasil

thalesmaciel@ifsul.edu.br, gustavo.b.seli98@gmail.com

***Abstract.** The present study aimed to discover a predictive model that is capable of helping education professionals in identifying the extra-academic characteristics that have impact on individual performance on the ENEM. Data mining experiments were performed with preprocessing tasks and classification with decision tree discovery. Partial results showed that, even though not the main reason, familiar income, when up to one minimum salary, on single candidates, may influence on poor performance.*

1. Introdução

Inicialmente, o Exame Nacional do Ensino Médio (ENEM) foi utilizado apenas como instrumento de critério de seleção de alunos que almejavam uma bolsa integral ou parcial no Programa Universidade para Todos (PROUNI) do Ministério da Educação (MEC) do Brasil, o que possibilitava que estudantes de baixa renda adquirissem diploma de curso superior, embora hoje seja usado como critério geral para o ingresso em instituições federais e privadas (Malusá et al., 2014).

Em Gomes (2015), foi abordada a descoberta de conhecimento com base em dados educacionais das edições de 2013 e 2014 do ENEM, embora sem a definição de um objetivo específico. O procedimento adotado foi o da mineração de dados pelas tarefas de associação, com o algoritmo Apriori, e classificação, com o algoritmo J48. Foram encontrados padrões nos dados, embora não tenha sido discutida a aplicabilidade do conhecimento descoberto.

Adeadato, Filho e Rodrigues (2014) objetivaram a avaliação de qualidade da educação secundária no Brasil. Para tal, os autores procederam à integração e seleção de dados do ENEM e censo escolar de 2011, posteriormente procedendo à tarefas de regressão, classificação e associação. Segundo os autores, assim foi possível prever quantitativamente o sucesso dos alunos egressos de cada escola do ensino secundário no exame, identificando a classe econômica da região geográfica de localização das escolas para predição do sucesso dos candidatos no ENEM.

O problema de pesquisa abordado no presente estudo é fundamentado em “considerando os aspectos socioeconômicos de cada indivíduo avaliado pelo ENEM, é possível identificar um perfil daqueles que logram sucesso na obtenção de vaga no ensino superior através do exame?”. O objetivo do trabalho é descobrir um modelo preditivo capaz de auxiliar os profissionais de educação na identificação de características extra-acadêmicas, dos candidatos, que tenham impacto no desempenho individual no ENEM, de forma a permitir que estes profissionais projetem políticas públicas que atenuem tais influências.

2. Solução Proposta

Procedeu-se à recuperação dos conjuntos de dados que viabilizaram o estudo, através de coleta nas bases de dados abertos do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), tendo composto um conjunto de dados de aplicações de prova do ENEM para 300 candidatos com os atributos sexo, estado civil, cor/raça, ensino médio público ou privado, escolarização do pai, escolarização da mãe, quantidade de moradores na mesma casa do candidato, renda familiar e o indicador de resultado satisfatório ou insatisfatório no exame, todos do tipo categórico.

Segundo Quinlan (1996), a árvore de decisão é um tipo de modelo de dados capaz de detalhar tomadas de decisão, quando elas ocorrem com base em dados, sendo comumente utilizado como resultado de tarefas de classificação. O autor provê, ainda, a especificação do C4.5, que é um algoritmo capaz de descobrir árvores de decisão com base em dados, de forma automatizada em ambiente computacional.

Este tipo de modelo é apreciado no contexto deste estudo pela simples interpretabilidade, além de que sua construção através da mineração de dados possibilita a explicação para a ocorrência de fenômenos com base em dados históricos. O modelo de classificação descoberto, apresentado na figura 1, foi testado por validação cruzada em 10 partições e obteve acurácia de 68% na predição de sucesso dos candidatos no exame.

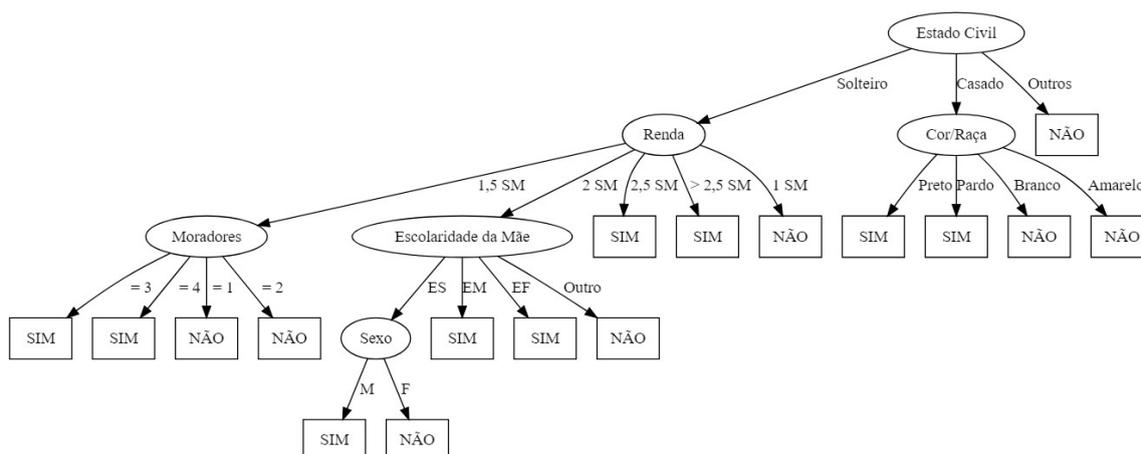


Figura 1. Árvore de decisão resultante do experimento realizado.

3. Considerações Finais

Resultados preliminares, com o experimento descrito, propõem que o estado civil dos candidatos, dentre os demais atributos analisados, é maior relevância na determinação de sucesso no exame. No segundo nível da árvore de decisão descoberta, foi possível observar os atributos referentes a renda familiar e cor/raça do candidato como determinantes no respectivo sucesso.

Trabalhos futuros envolvem maiores esforços na coleta de dados junto ao MEC, com vistas no aumento do conjunto de dados nos aspectos de quantidade de instâncias e de atributos, experimentações com diferentes algoritmos de classificação e respectivas parametrizações na realização dos experimentos e análise das matrizes de confusão e aplicabilidade prática dos modelos descobertos.

Referências

- Adeodato, P., Santos Filho, M. e Rodrigues, R. (2014). Predição de desempenho de escolas privadas usando o ENEM como indicador de qualidade escolar. In Simpósio Brasileiro de Informática na Educação 2014.
- Gomes, T. C. S. (2015). Descoberta de Conhecimento Utilizando Mineração de Dados Educacionais Abertos. UFRPE. Recife.
- Malusá, S., de Medeiros Ordones, L.L., & Ribeiro, E. Enem: pontos positivos para a educação brasileira. *Revista Educação e Políticas em Debate*, 3(2).
- Quinlan, J. R. (1996). Improved use of continuous attributes in C4. 5. *Journal of artificial intelligence research*, 4, 77-90. Chicago, IL.