

ANÁLISE DOS DADOS DO QUESTIONÁRIO SOCIOECONÔMICO E CULTURAL DO PROCESSO SELETIVO DO IFSULDEMINAS NO ANO DE 2017

Perses de VILHENA¹; Ramon G. T. M. da SILVA²; Francisco J. N. FILHO³

RESUMO

A tomada de decisão implica diretamente na eficiência da gestão escolar e conseqüentemente na obtenção de melhores resultados para instituições de ensino. Dentre as formas de fundamentar a tomada de decisão, a análise e visualização de dados auxiliam consideravelmente. Desta forma, este trabalho propõe a análise dos dados obtidos do questionário socioeconômico e cultural do processo seletivo do IFSULDEMINAS no ano de 2017, visando a identificação de perfis e padrões que possam auxiliar na tomada de decisão. A análise dos dados foi organizada em pré-processamento, extração de padrões com regras de associação, e avaliação dos resultados. Os resultados mostraram que por exemplo quem tem computador em casa utiliza a internet diariamente, e quem é de escola pública e de zona urbana avaliaram o processo seletivo ótimo, entre outras regras.

Palavras-chave:

Mineração de dados; KDD; Análise de dados; Regras de associação; Processo seletivo

1. INTRODUÇÃO

Atualmente, por meio das tecnologias e ferramentas, a extração e obtenção de uma grande quantidade de dados vem sendo facilitada e sendo feita com maior frequência. Essa grande quantidade de dados captados muita das vezes acabam não sendo exploradas por completo, ou até mesmo conservam-se de forma arquivada. Geralmente a compreensão e interpretação de grandes volumes de dados auxiliam a tomada de decisão, porém desafiam a capacidade humana. Em diversos segmentos as técnicas computacionais começaram a ser utilizadas para auxiliar nas tarefas de análises de dados para a tomada de decisão (regras de associação).

Na gestão educacional a análise de dados passou a ser utilizada como instrumento na tomada de decisões, no âmbito da gestão da diretoria e reitoria. Muitas vezes são utilizadas técnicas de descoberta de conhecimento em bases de dados, do inglês *Knowledge Discovery in Databases* (KDD), por meio da mineração de dados. A descoberta de informação útil pode ser realizado por algoritmos de aprendizado de máquina supervisionado ou não supervisionado. O método não

¹ IFSULDEMINAS – 12131001101@muz.ifsuldeminas.edu.br

² IFSULDEMINAS – ramon.silva@muz.ifsuldeminas.edu.br

³ IFSULDEMINAS – francisnardi@gmail.com

supervisionado é mais utilizado por não exigir que tenham padrões rotulados já identificados, sendo as regras de associação bastante comuns. E os métodos não supervisionados de aprendizado de máquina são chamados assim pois dispensam a necessidade de um supervisor para a realização do processo de aprendizagem, porém necessitam de amostras de conjuntos de dados rotulados para treinamento do algoritmo.

No IFSULDEMINAS, todo candidato ao processo seletivo preenche um questionário socioeconômico e cultural ao fazer a inscrição, o que gera uma quantidade significativa de dados que podem ser analisados visando a obtenção de informações úteis para o entendimento do perfil dos candidatos, o que pode colaborar para tomada de decisões. Em 2013, Madeira realizou um trabalho de Mineração de dados neste questionário, o que gerou informações relevantes para a instituição e uma metodologia para a análise de dados.

Neste sentido, este trabalho analisou os dados socioeconômicos e culturais dos candidatos ao vestibular do IFSULDEMINAS no ano de 2017, visando a obtenção de conhecimento acerca do perfil dos candidatos. O trabalho está organizado nas etapas de: pré-processamento, onde ocorre a busca, limpeza e preparação dos dados; extração de padrões com técnicas de aprendizado de máquina (regras de associação); validação e visualização das informações extraídas.

2. MATERIAL E MÉTODOS

2.1. Pré-processamento dos dados

Inicialmente os dados do questionário socioeconômico e cultural do vestibular do ano de 2017 foram obtidos por meio de uma planilha em formato CSV (*Common separated values*) junto a COPESE (Comissão permanente de seleção) do IFSULDEMINAS - Campus Muzambinho. Em seguida foi implementado um script em *PHP*, que realiza a leitura e processamento da planilha, e faz a inserção de forma normalizada e organizada em uma tabela no banco de dados *MySQL*. Em alguns registros houve falta ou inversão dos dados inseridos. Nestes casos houve a necessidade de limpeza de dados, onde dados faltosos ou invertidos foram tratados para não gerar ruídos nas análises. De acordo com a metodologia de Madeira (2013) os registros que possuíam mais de 2 atributos faltosos foram removidos.

Após a limpeza, processamento e preparação dos dados foi criado um *script PHP* para transformar os dados para o formato *.arff*, necessário para a entrada na ferramenta de extração de regras de associação *Weka*.

2.2. Extração dos padrões e visualização dos dados

O arquivo .arff preparado e organizado com os dados do vestibular foi submetido na ferramenta *Weka* para a extração de regras de associação, utilizando o algoritmo *Apriori*. As regras de associação obtidas foram avaliadas e visualizadas por meio da ferramenta *Weka*.

3. RESULTADOS E DISCUSSÕES

O pré-processamento foi fundamental e a limpeza e transformação dos dados proporcionou a redução do tamanho de 171 *Megabytes* para 19,5 *Megabytes*, e do número de registros de 741.304 registros para 25.021 registros. Os registros removidos eram casos de duplicidade e/ou atributos faltosos. A extração das regras de associação por meio da ferramenta *Weka* proporcionou a obtenção de alguns padrões de comportamento, sendo os principais descritos na figura 1. As regras não foram semelhantes às extraídas no trabalho de Madeira em 2013, devido a diferença de tempo entre as análises.

A visualização das regras de associação realizada por meio da ferramenta *Weka* possibilitou uma melhor interpretação dos resultados, conforme pode ser observado na figura 1.

#	Porcentagem	Se	Então
1	94	micro=Tem em casa e usa regularmente	internet=Diariamente
2	91	automovel=Sim	antes_entrar=Sim
3	91	local_origem=Zona Urbana tipo_instituicao=Apenas em escola pública	seletivo=Ótimo
4	91	local_origem=Zona Urbana antes_entrar=Sim	seletivo=Ótimo
5	90	estado=MG local_origem=Zona Urbana	seletivo=Ótimo
6	90	vestibular_if=Nenhuma vez	antes_entrar=Sim
7	90	estado=MG tipo_instituicao=Apenas em escola pública antes_entrar=Sim	seletivo=Ótimo
8	90	tipo_instituicao=Apenas em escola pública antes_entrar=Sim	seletivo=Ótimo
9	90	local_origem=Zona Urbana	seletivo=Ótimo
10	90	local_origem=Zona Urbana internet=Diariamente	seletivo=Ótimo

Figura 1: Visualização das regras de associação.
Fonte: Do autor, 2018.

4. CONCLUSÕES

A análise de dados do vestibular do IFSULDEMINAS - Câmpus Muzambinho possibilitou a identificação de perfis dos candidatos, assim como foi feito em 2013 por Madeira. Alguns perfis foram os mesmos, porém outros apresentaram-se de forma distinta, o que evidencia a necessidade de análises por diferentes períodos e anos.

REFERÊNCIAS

- FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27-34, 1996.
- MADEIRA, Fernanda Delizete. **Mineração de dados educacionais: Um estudo de caso aplicado ao processo seletivo de IFSULDEMINAS - Campus Muzambinho**. 2013. 132 f. TCC (Graduação) - Curso de Ciência da Computação, Instituto Federal do Sul de Minas - Campus Muzambinho, Muzambinho, 2013.
- MESSIAS, Icaro Brito de Carvalho. **SADC - SISTEMA DE APOIO AO DIAGNÓSTICO CLÍNICO: UMA FERRAMENTA WEB MÓVEL BASEADA EM SISTEMA DE RECOMENDAÇÃO PARA UNIDADES BÁSICAS DE SAÚDE**. 2018. 47 f. TCC (Graduação) - Curso de Ciência da Computação, Instituto Federal do Sul de Minas - Campus Muzambinho, Muzambinho, 2016.
- MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.
- REZENDE, Solange Oliveira. Mineração de Dados. **A Universalidade da Computação: Um Agente de Inovação e Conhecimento**, São Leopoldo, p.397-433, jul. 2005.