

USO DE DATA MINING NO AUXÍLIO AO COMBATE À EVASÃO ESCOLAR NO INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA DO TOCANTINS – CAMPUS COLINAS DO TOCANTINS

Fernando Turibio de Moura¹

¹Especialista em Banco de Dados - ULBRA. e-mail: <fernando.moura@ifto.edu.br>

Resumo: O presente artigo tem como objetivo levantar dados acerca da evasão escolar dos alunos dos cursos integrado e subsequente ao ensino médio em informática do Instituto Federal de Educação Ciência e tecnologia do Tocantins (IFTO), Campus Colinas do Tocantins, do período de 01 de agosto de 2014 à 01 de fevereiro de 2016, utilizando a técnica de mineração para redução da evasão escolar na instituição. Ao final foi realizada uma estimativa do perfil do aluno evadido utilizando como base os dados coletados.

Palavras-chave: evasão escolar, mineração de dados

1 INTRODUÇÃO

A tecnologia tornou relativamente fácil o acúmulo de dados, ampliando o uso de banco de dados que podem ser usados para dar suporte aos objetivos fundamentais de um negócio. As informações coletadas possuem um valor significativo e os dados armazenados em seus bancos de dados são vasculhados por profissionais especializados, a procura de tendências e padrões (SOUSA, 2003).

A evasão escolar é um tema que historicamente faz parte dos debates e reflexões no âmbito da educação pública brasileira e que ocupa espaço de relevância no cenário das políticas públicas e da educação em particular. As discussões acerca da evasão escolar têm tomado como ponto central de discussão a interseção entre os papéis da família e da escola em relação à vida escolar da criança (QUEIROZ, 2001). A legislação vigente, que estabelece diretrizes para a educação no Brasil, determina a responsabilidade da família e do Estado no dever de orientar a criança em seu percurso sócio educacional. A Lei de Diretrizes e Bases da Educação Nacional (LDBEN), de 1996, é bastante clara a esse respeito. Em seu art. 2º, fica expresso que:

A educação, dever da família e do Estado, inspirada nos princípios de liberdade e nos ideais de solidariedade humana, tem por finalidade o pleno desenvolvimento do educando, seu preparo para o exercício da cidadania e sua qualificação para o trabalho.

A despeito disso, o que se observa é que a escolarização neste país cresceu em termos quantitativos, mas ainda não tem sido plena no que se refere ao alcance de todos os cidadãos. E essa disparidade é ainda maior no que se refere à conclusão de todos os níveis de escolaridade. Ou

seja, a evasão escolar continua limitando o acesso de nossos jovens à cidadania plena e, como consequência, mantém-se na pauta das discussões e reflexões realizadas pelo Estado e pela sociedade civil, especificamente pelas organizações e movimentos relacionados à educação no âmbito da pesquisa científica e das políticas públicas (FREITAG, 1984).

A Mineração de Dados é uma das tecnologias mais promissoras da atualidade. Um dos fatores desse sucesso é o fato de dezenas, e muitas vezes centenas, de milhões de reais serem gastos pelas companhias na coleta dos dados e, no entanto, nenhuma informação útil é identificada (LAROSE, 2014). Em seu trabalho, Han, Pei, Kamber, 2011 refere-se a essa situação como “rico em dados, pobre em informação”. Além da iniciativa privada, o setor público e o terceiro setor (Organizações não governamentais) também podem se beneficiar com a Mineração de Dados.

2 MATERIAIS E MÉTODOS

Os principais objetivos da mineração de dados são descobrir relacionamentos entre dados e fornecer subsídios para que possa ser feita uma previsão de tendências futuras baseadas no passado. A mineração de dados pode ser aplicada como um processo de verificação e como um processo de descoberta. No processo de verificação, o usuário sugere uma hipótese acerca da relação entre os dados e tenta prová-la aplicando técnicas como análises estatística e multidimensional sobre um banco de dados contendo informações passadas.

Em mineração de dados, no processo de descoberta do conhecimento não é feita nenhuma suposição antecipada. Para tanto é preciso identificar as bases de dados a serem utilizadas e neste caso foi realizado um estudo em cima do módulo acadêmico do Sistema integrado de gestão acadêmica(SIGA-EPCT), chamado SIGA-EDU, analisando como sua arquitetura está estruturada, buscando em seu banco dados que pudessem ser utilizados para o desenvolvimento deste trabalho. A figura 1 ilustra a ferramenta.

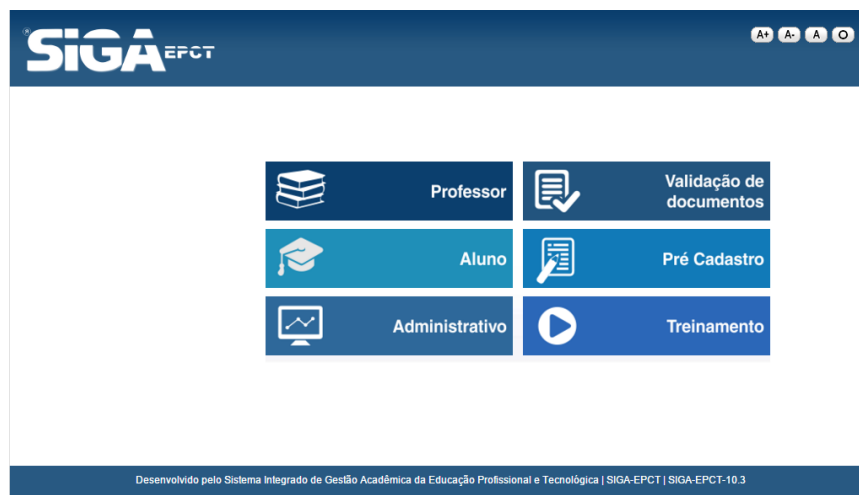


Figura 1 – Imagem da tela inicial do SIGA-EDU.

Foi realizada a aplicação de técnicas de mineração de dados sobre dados dos alunos do campus de Colinas do Tocantins do IFTO. Para utilizar as ferramentas de extração e mineração de dados, houve a necessidade de consolidar um banco de dados a partir de fontes de dados oriundas de diversos setores do Instituto, visto que não existe um único banco de dados com toda a informação corporativa.

Devido à heterogeneidade das fontes de dados, este processo foi realizado de forma manual e foram considerados apenas dados dos cursos técnico em informática subsequente e médio. Está sendo implementada uma estrutura automática para integração de instâncias destas diversas fontes de dados heterogêneas.

Dentro do banco consolidado, os dados analisados representam 1 ano e 6 meses de atividades da instituição (período que compreende do ano de 2014 a 2016), existem 149 alunos dos cursos de informática matriculados no sistema dos quais 44% das instâncias representam dados de alunos que evadiram e os outros 56% representam dados de alunos que não evadiram. A fim de realizar a mineração de dados, foram utilizados algoritmos de classificação implementados na ferramenta Waikato Environment for Knowledge Analysis (WEKA) (HALL et al, 2009).

O WEKA é um produto da Universidade de Waikato (Nova Zelândia) e foi implementado pela primeira vez em sua forma moderna em 1997. Ele usa a GNU (GNU is not unix) sistema baseado no UNIX e a licença General public license (GPL). O software foi escrito na linguagem Java™ e contém uma GUI (Graphical user interface) para interagir com arquivos de dados e produzir resultados visuais (tabelas e gráficos).

Ele também tem uma API (Application programming onterface) geral, assim é possível incorporar o WEKA, como qualquer outra biblioteca, a seus próprios aplicativos para fazer coisas como tarefas de mineração de dados automatizadas no lado do servidor.



Figura 2 – Aplicativo de mineração de dados WEKA.

Para se utilizar a ferramenta adequadamente primeiramente é necessário executar alguns passos da descoberta de conhecimento, dentre eles a fase de seleção ou extração dos dados, pré-processamento e formatação dos dados para desta forma fazer a mineração dos dados em busca de padrões e relações que fossem interessantes.

Como o número de registros é relativamente baixo em função do número de alunos e do pouco tempo de funcionamento do campus, não foi preciso utilizar técnicas de amostragem, podendo ser utilizado todo o extrato de dados. Os dados foram extraídos da base, tanto por meio de instruções SQL (Structured Query Language) como por meio da visualização dos relatórios disponíveis na Coordenação de Registros Escolares (CORES) como histórico escolar e boletins dos alunos.

Para fins de estudo de evasão foi levantado também o perfil do alunado (idade, sexo, módulo em que ocorreu a evasão dentre outros) fazendo um estudo aprofundado na parte institucional. Depois de obtidos, os dados passaram por uma filtragem utilizando programas de planilha eletrônica e tabelas dinâmicas nas mesmas para descobrir irregularidades ou dados ausentes e descartá-los.

4 RESULTADOS E DISCUSSÕES

A partir do uso dos algoritmos de mineração divididos em, classificação, segmentação e árvore de decisão foi possível obter dados interessantes acerca dos discentes do IFTO – Campus Colinas do Tocantins. Conforme denotado no item 2 (Materiais e métodos) do artigo a taxa de evasão corresponde a 44% do total de alunos analisados (cerca de 56). O gráfico abaixo mostra a porcentagem de evasão por modalidade de ensino:

Porcentagem de alunos evadidos por modalidade

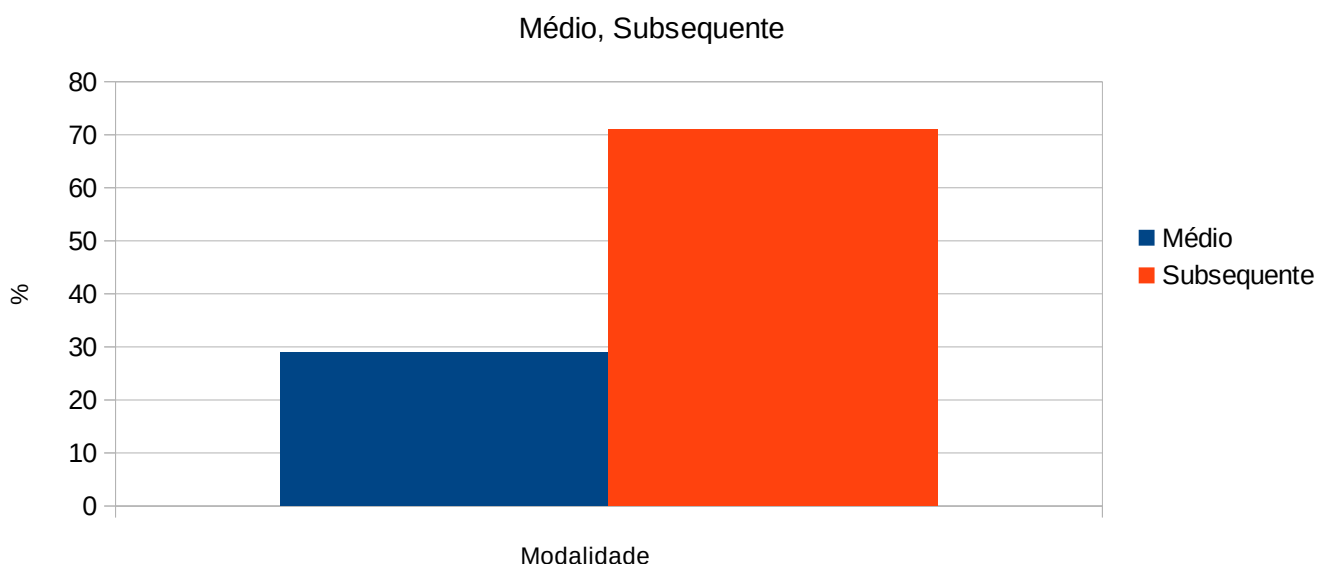


Gráfico 1 – Taxa de evasão por modalidade de ensino.

Conforme o Gráfico 1, 71% dos alunos evadidos pertencem a modalidade subsequente ao ensino médio e 29% pertencem a modalidade integrada. Dentro do escopo de alunos do ensino subsequente, 50% são do sexo masculino e 50% do sexo feminino e 79% dos alunos evadidos nesta modalidade são solteiros. Dado a gravidade do índice de evasão concentrado na modalidade subsequente ao ensino médio, foi feita uma estimativa do perfil de risco de evasão escolar dentro deste grupo.

Dentro do grupo de discentes da modalidade subsequente ao ensino médio conseguiu-se ainda 2 perfis considerados de risco, sexo feminino com idade de 23 anos e estado civil solteira e outro com sexo masculino idade de 30 anos e estado civil solteiro. Foram utilizados os algoritmos, para classificação de cluster, K-means e o algoritmo de classificação com base em árvore de

decisão J-48 para classificar os dados (conforme figura 3 e 4), considerando também devidas modificações nos atributos dos dados para utilizar os algoritmos citados.

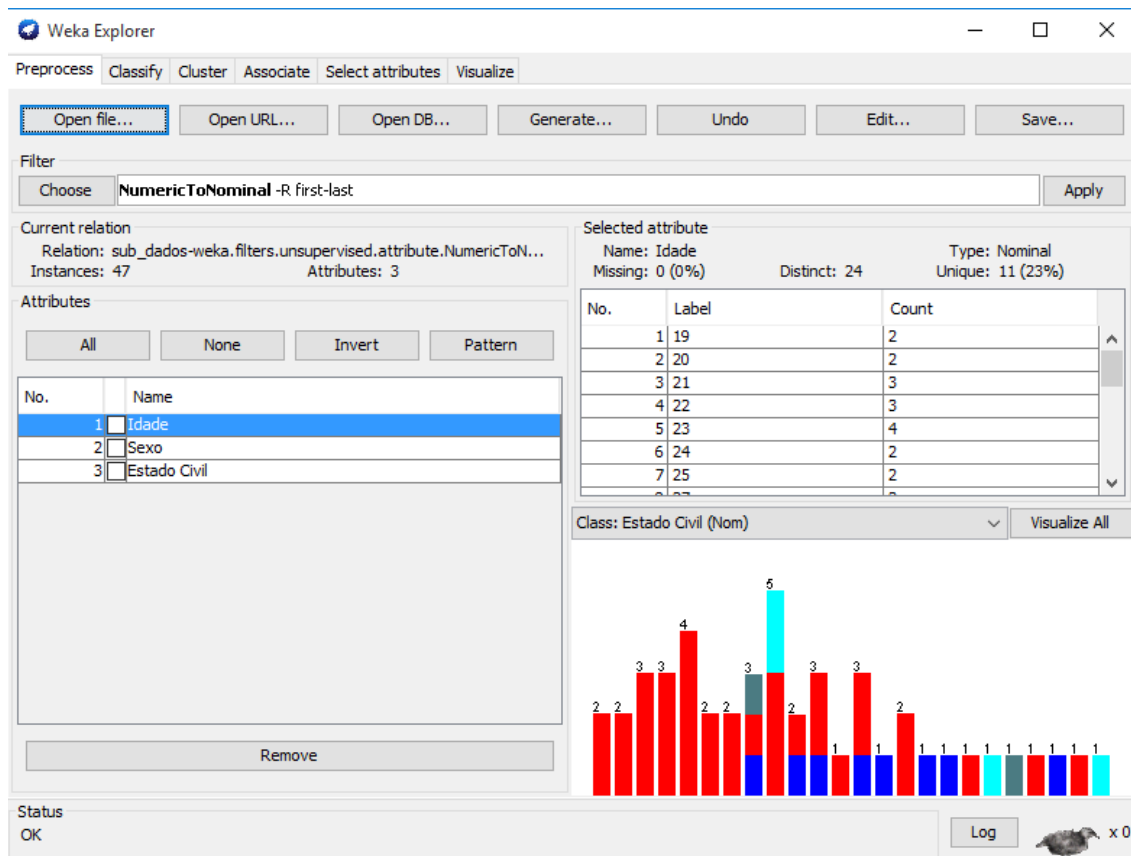


Figura 3 – Uso de algoritmos de mineração na ferramenta WEKA.

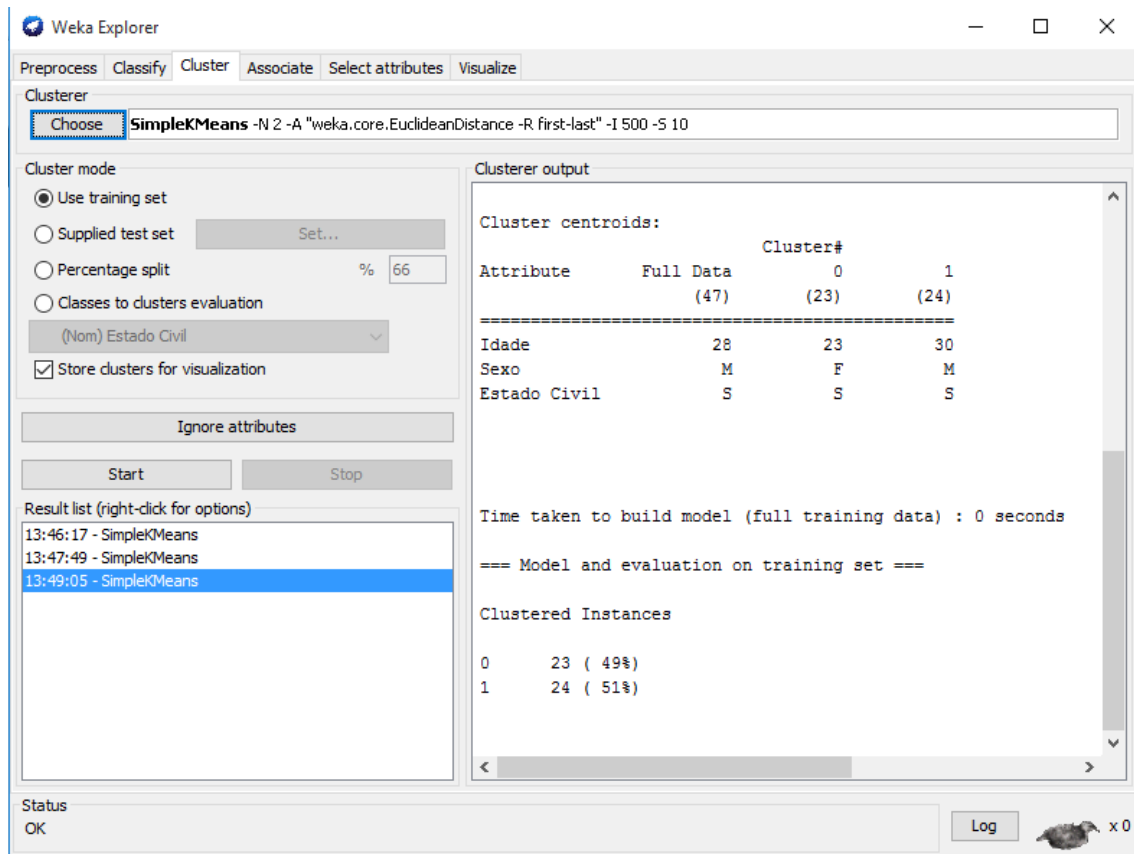


Figura 4 – Uso do algoritmo SimpleKMeans no Weka.

5 CONSIDERAÇÕES FINAIS

O uso da metodologia da exploração de dados no ambiente educacional contribuiu de forma eficiente na descoberta de padrões e de informações cruciais acerca da evasão escolar, com a descoberta de que a evasão escolar no curso de informática do Campus Colinas do Tocantins acontece principalmente na modalidade subsequente ao ensino médio, já se pode tomar atitudes focadas na prevenção de evasão na modalidade supracitada, seja por oferta de bolsas estudantis, seja em um maior acompanhamento destas turmas bem como em um atendimento aos alunos do sexo masculino e feminino, solteiros e com idade acima dos 22 anos, no qual se encaixa um potencial perfil de aluno evadido.

Este estudo focou-se principalmente em fatores internos, ou seja, dados que pertencem a instituição de ensino e que estavam disponíveis. Em busca de resultados mais específicos e um estudo mais amplo, sugere-se para trabalhos futuros, foco em fatores externos à instituição como por exemplo a vida dos alunos em risco fora da escola, fatores familiares, econômicos e adversos

que muitas vezes o ambiente escolar não tem controle ou nem mesmo estas informações por não estarem em seu domínio.

REFERÊNCIAS

FREITAG, Barbara. Escola, estado e sociedade. In: **Coleção educação universitária**. Moraes, 1984.

HALL, Mark et al. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, v. 11, n. 1, p. 10-18, 2009.

HAN, Jiawei; PEI, Jian; KAMBER, Micheline. **Data mining: concepts and techniques**. Elsevier, 2011.

LAROSE, Daniel T. **Discovering knowledge in data: an introduction to data mining**. John Wiley & Sons, 2014.

QUEIROZ, Lucileide Domingos. Um estudo sobre a evasão escolar: para se pensar na inclusão escolar. **Associação Nacional de Pós-Graduação e Pesquisa em Educação (Anpad)**. Disponível em www.anped.org.br/reunioes/25/lucileidedomingosqueirozt13.rtf. Acesso em, v. 3, 2001.

SOUSA, Paulo de Tarso Costa de et al. **Mineração de dados para indução de um modelo de gestão do conhecimento**. 2003.