

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
FACULDADE DE CIÊNCIAS EXATAS E TECNOLOGIA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO
PROJETO DE TRABALHO DE CONCLUSÃO DE CURSO

**MINERAÇÃO DE DADOS COMO PARTE DO PROCESSO DE
DESCOBERTA DO CONHECIMENTO**

ROGÉRIO DUARTE DE SOUZA

DOURADOS – MS
2013

ROGÉRIO DUARTE DE SOUZA

**MINERAÇÃO DE DADOS COMO PARTE DO PROCESSO DE
DESCOBERTA DO CONHECIMENTO**

Orientador: Prof. Me. Everton Castelão Tetila

Trabalho de Conclusão de Curso apresentado à Universidade Federal da Grande Dourados (UFGD) como exigência parcial para graduação no curso de Bacharelado em Sistemas de Informação.

DOURADOS – MS
2013

ROGÉRIO DUARTE DE SOUZA

**MINERAÇÃO DE DADOS COMO PARTE DO PROCESSO DE
DESCOBERTA DO CONHECIMENTO**

Orientador: Prof. Me. Everton Castelão Tetila

Trabalho de Conclusão de Curso apresentado à Universidade Federal da Grande Dourados (UFGD) como exigência parcial para graduação no curso de Bacharelado em Sistemas de Informação.

Aprovado em: ___/___/_____

BANCA EXAMINADORA

Prof. Me. Everton Castelão Tetila
Universidade Federal da Grande Dourados - UFGD

Prof. Me. Murilo Táparo
Universidade Federal da Grande Dourados - UFGD

Prof. Me. Willian Paraguassu Amorim
Universidade Federal da Grande Dourados – UFGD

DEDICATÓRIA

Dedico esta vitória a meus pais, meus irmãos, minha cunhada e minha sobrinha Beatriz. Além de meus amigos, os melhores amigos.

AGRADECIMENTOS

Primeiramente a Deus, por ter me dado força e sabedoria para concluir este trabalho, além do dom precioso da vida.

A meus pais e familiares, os meus avôs Manoel Duarte de Souza (*in memoriam*), que me mostrou os verdadeiros valores da vida; e José Ribeiro da Silva (*in memoriam*), que me presenteou com seu amor e carinho desde que era apenas um bebê.

Especial agradecimento a meu orientador, Prof. Me. Everton Castelão Tetila, que me acompanhou nesta jornada e transmitiu seus conhecimentos e experiências de forma clara e concisa. Estendendo também este agradecimento ao professor Me. Murilo Táparo e o professor Me. Willian Paraguassu Amorim.

EPÍGRAFE

“No que diz respeito ao empenho, ao compromisso, ao esforço, à dedicação, não existe meio termo. Ou você faz uma coisa bem feita, ou não faz.” Airton Senna da Silva.

RESUMO

A Descoberta de Conhecimento nos Bancos de Dados, ou *Knowledge Discovery in Databases* (KDD), é um processo que compreende seis fases: seleção de dados, limpeza de dados, enriquecimento, transformação ou codificação de dados, mineração de dados e o relatório e exibição da informação descoberta.

Mineração de dados é um processo não trivial de extrair ou minerar conhecimento previamente desconhecidos, porém, potencialmente úteis, de grandes volumes de dados, com o intuito de identificar anomalias, estabelecer padrões, relações ou associações capazes de transformar dados brutos em informações estrategicamente valiosas. É processada para identificar diversos fatores e tendências nas atividades de negócios, controle de qualidade, detecção de fraudes, pesquisas científicas, controle de processos, entre outros.

Waikato Environment for Knowledge Analysis (WEKA) é um *software* livre do tipo *open source* para mineração de dados. Foi desenvolvido na Linguagem de programação Java pelo curso de Ciência da Computação da Universidade de Waikato, na Nova Zelândia, no ano de 1993, dentro das especificações da GPL (*General Public License*). Consolidou-se, ao longo dos últimos anos, como a ferramenta de mineração de dados mais utilizada em ambiente acadêmico.

Este trabalho apresenta a mineração de dados realizada nos históricos acadêmicos do curso de Bacharelado de Sistema de Informação (BSI) da Universidade Federal da Grande Dourados (UFGD) no período de 2006 a 2009, com intuito de descobrir as causas potenciais ou os fatores-chave que ocasionam altos índices de evasão e reprovação e baixos índices de aprovação e diplomação.

ABSTRACT

The Knowledge Discovery in Databases, or Knowledge Discovery in Databases (KDD) is a process that consists of six stages: data selection, data cleansing, enrichment, transformation or data coding, data mining and reporting and display of information discovery.

Data mining is a nontrivial process of extracting or mining knowledge previously unknown, but potentially useful for large volumes of data in order to identify anomalies, establish patterns, relationships or associations that transform raw data into valuable information strategically. It is processed to identify several factors and trends in business activities, quality control, fraud detection, scientific research, process control, among others.

Waikato Environment for Knowledge Analysis (WEKA) is a free software like open source for data mining. Was developed in the Java programming language for the course of Computer Science, University of Waikato, New Zealand, in 1993, within the specifications of the GPL (General Public License). Consolidated over the last few years, as the tool most widely used data mining in academic environment.

This paper presents the data mining performed in historical students of Bachelor of Information System (BSI) of the Federal University of Grande Gold (UFGD) in the period 2006-2009, in order to discover the causes or potential key factors that cause high rates of repetition and dropout and low approval ratings and graduation.