

CENTRO UNIVERSITÁRIO FEEVALE

INGO JOST

MINERAÇÃO DE DADOS PARA ADOÇÃO DE PRÁTICAS DE
MARKETING EM AMBIENTE ACADÊMICO

(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo, setembro de 2008.

INGO JOST

ingo@feevale.br

MINERAÇÃO DE DADOS PARA ADOÇÃO DE PRÁTICAS DE
MARKETING EM AMBIENTE ACADÊMICO

(Título Provisório)

Centro Universitário Feevale
Instituto de Ciências Exatas e Tecnológicas
Curso de Ciência da Computação
Anteprojeto de Trabalho de Conclusão

Professor orientador: Juliano Varella de Carvalho

Novo Hamburgo, setembro de 2008.

RESUMO

Empresas dos mais diversos segmentos falham em não aproveitar todo o conhecimento existente em suas bases de dados, o que poderia ser útil tanto para transformação desses dados em informação, como no auxílio à tomada de decisões. Para suprir essa deficiência, surgiu a área de Descoberta de Conhecimento, mais conhecida pelo nome de uma de suas etapas: Mineração de Dados (*Data Mining*). Este é um processo que pode ser aplicado mediante diversas técnicas, destacando-se Classificação, Associação e Agrupamento, sendo estas implementadas por diferentes algoritmos. Estas técnicas serão estudadas e aplicadas para suprir a necessidade de um centro universitário de descobrir informações no seu conjunto de dados de vestibulandos. Após o tratamento destes dados, serão mapeados perfis de alunos, com características como: opções de curso, município de residência, escolas frequentadas, dentre outras. Este trabalho visa aplicar as técnicas de *Data Mining* para que se identifique uma relação entre os perfis mapeados e a demanda por cada curso. Sendo assim, disponibilizar-se-á ao setor de Marketing, informações que permitirão a adoção de práticas específicas para cada área (por exemplo, Ciências Exatas ou Ciências Humanas) ou curso.

Palavras-chave: *Data Mining*. Técnicas de Classificação. Marketing de precisão. Mapeamento de Perfis.

SUMÁRIO

MOTIVAÇÃO	5
OBJETIVOS	8
METODOLOGIA	9
CRONOGRAMA	10
BIBLIOGRAFIA	11

MOTIVAÇÃO

As corporações têm buscado, no decorrer das últimas décadas, digitalizar os seus dados e adquirir ou desenvolver sistemas de informação que manipulam e alimentam essas bases. No entanto, este gerenciamento das informações é utilizado normalmente apenas para a gestão e otimização de processos, não sendo aproveitado para uma análise mais detalhada capaz de propiciar novas oportunidades às instituições.

Com esse objetivo, surge a área de Descoberta de Conhecimento em Bases de Dados, cujo nome é originado do inglês *Knowledge Discovery in Databases* (KDD). Segundo Goldschmidt (2005), a Descoberta do Conhecimento é composta pelas seguintes etapas operacionais:

- Pré-processamento: etapa em que os dados recebidos passam por uma preparação para a próxima etapa. Esta preparação consiste na filtragem dos dados, excluindo os irrelevantes e os que possuem inconsistências.
- Mineração de Dados: quando é realizada a busca propriamente dita pelo conhecimento. Existem diversos algoritmos que implementam diferentes técnicas para a obtenção do conhecimento, destacando-se Associação, Classificação e Agrupamento.
- Pós-processamento: aproveitamento das informações adquiridas, sendo realizada a interpretação e avaliação da importância do conhecimento descoberto, se houver.

A Mineração de Dados é utilizada nas mais diversas áreas, desde a descoberta de pesos de atributos em um sistema de Raciocínio Baseado em Casos, conforme Silveira

(2003), até a definição de padrões em pacientes de Diabetes, de acordo com Neves (2003). Independente do foco, são utilizadas as técnicas já citadas, inclusive os mesmos algoritmos. Dentre estas técnicas, destacam-se:

- **Agrupamento:** procura separar em grupos, registros com características as mais homogêneas possíveis, ou seja, que possuam propriedades comuns. Além disso, a técnica de agrupamento busca a maior distinção entre os grupos definidos (GOLDSCHMIDT, PASSOS, 2005).
- **Classificação:** consiste em classificar os registros em categorias (classes) pré-definidas, possibilitando, quando da inserção de novos registros, que estes já sejam classificados automaticamente (GOLDSCHMIDT, PASSOS, 2005).
- **Associação:** busca associações entre os atributos de um conjunto de dados, conforme Frank e Witten (2005). O exemplo mais comum para a ilustração é o de transações de compras: obtém-se regras de associação com a ocorrência de produtos em diferentes transações. Por exemplo, um mercado pode auferir que 80% dos compradores de pão, também compram leite. A partir destes conhecimentos descobertos, diversas atitudes podem ser tomadas, como por exemplo, deixar esses produtos próximos (ou distantes), promover a venda conjunta dos produtos, reduzir o preço de um deles e aumentar o preço do outro, dentre outras medidas cabíveis.

Existem diversas ferramentas *freeware* que permitem a utilização das técnicas de KDD, como *RapidMiner*¹, *Pentaho*,² sistema de *Business Intelligence* que possui módulo de *Data Mining* e *Weka*,³ ferramenta *open-source* que pode ser acoplada a outras, como o próprio *Pentaho*, por exemplo.

O pacote *Weka* (*Waikato Environment for Knowledge Analysis*) é composto por bibliotecas que implementam diversos algoritmos das técnicas de KDD. Inerente ao *software*, está a interface gráfica que permite a escolha da forma de apresentação dos resultados (gráficos, árvores). Não obstante, podem ser desenvolvidas aplicações que utilizam apenas a camada dos algoritmos de Mineração de Dados (FRANK, WITTEN, 2000).

¹ <http://www.rapidminer.com>

² <http://www.pentaho.com>

³ <http://www.cs.waikato.ac.nz/ml/weka/>

Através das técnicas e tecnologias apresentadas, tem-se como motivação deste trabalho a criação de uma ferramenta que extraia conhecimento da base de dados de um centro universitário. Esta instituição, que conta com milhares de alunos, procura uma solução para melhor distribuir seus esforços em relação ao Marketing, área que vem sendo bastante explorada por aplicações de *Data Mining*, como verificado em *KDNuggets*⁴, portal referência em KDD.

Dentre os exemplos na área, destacam-se processos de relacionamento de clientes, conforme Almeida (2002), ou de mapeamento de perfis, como a empresa *Sigma*⁵ que possui uma ferramenta com diversos canais de comunicação para que as empresas conheçam as características de seus consumidores e o *Talisma Knowledgebase Software*⁶, que permite direcionar o contato da empresa com o cliente, a partir de transações já realizadas, evitando gastos com comunicação.

Este projeto manipulará informações pessoais de alunos, procurando peculiaridades comuns em, por exemplo: escolha de curso, município de residência, idade, escolas frequentadas. Mapear-se-á perfis de vestibulandos para que se encontre associações entre as características dos alunos e a escolha por cursos, possibilitando ao departamento de Marketing adotar práticas direcionadas para institutos, como o de Ciências Exatas, ou especificamente para cursos.

Fica a total critério da equipe de Marketing a forma de aproveitamento do conhecimento adquirido, juntamente com as práticas, que podem ser desde o direcionamento de propagandas até o agendamento de visitas de apresentação conforme a escola ou região de maior procura pelo curso (ou menor).

⁴ [http:// www.kdnuggets.com/](http://www.kdnuggets.com/)

⁵ <http://www.sigmamarketing.com/>

⁶ http://www.talisma.com/tal_products/knowledgebase.aspx

OBJETIVOS

Objetivo Geral:

Extrair conhecimento de base dados de centro universitário, buscando mapear perfis de alunos e a relação destes com a demanda por cursos, possibilitando a adoção de ações publicitárias direcionadas para institutos, como o de Ciências Exatas, ou especificamente para cursos.

Objetivos específicos:

- Apresentar a área de Descoberta de Conhecimento, suas etapas e principais técnicas;
- Estudar algoritmos que implementam essas técnicas;
- Referenciar aplicações que utilizam *Data Mining*;
- Analisar e compreender algumas bases de dados do centro universitário;
- Tratar os dados a serem recebidos do centro universitário;
- Estudar as técnicas / algoritmos que podem apresentar os resultados mais satisfatórios para o problema em questão;
- Estudar a ferramenta *Weka*, para aplicação das técnicas e apresentação de resultados;
- Apresentar o relatório final com os resultados obtidos na implementação do projeto.

METODOLOGIA

1. Elaboração do anteprojeto, definindo o problema a ser abordado.
2. Pesquisa em referencial bibliográfico (artigos, livros, monografias, teses) para a fundamentação teórica.
3. Preparação da base de dados para que sejam aplicados os algoritmos.
4. Estudo aprofundado de técnicas de *Data Mining* e algoritmos que as implementam.
5. Definição dos algoritmos a serem utilizados.
6. Estudo da ferramenta *Weka*.
7. Elaboração do relatório de Trabalho de Conclusão. I
8. Aplicação das técnicas.
9. Interpretação dos resultados obtidos em conjunto com a área de *Marketing*.
10. Análise da utilidade das informações para o centro universitário.
11. Elaboração do relatório de Trabalho de Conclusão II.

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Ago	Set	Out	Nov
1	■	■		
2		■		
3		■		
4		■	■	
5			■	
6			■	■
7		■	■	■

Trabalho de Conclusão II

Etapa	Meses			
	Mar	Abr	Mai	Jun
8	■	■		
9		■		
10		■		
11	■	■	■	■

BIBLIOGRAFIA

NEVES, Cledjalma Ferreira. Descobertas de Padrões Usando Técnicas de Extração de Conhecimento. **Palmas: Centro Universitário Luterano de Palmas - Práticas de Estágio, 2003**

SILVA, Marcelino Pereira dos Santos. Mineração de Dados – Conceitos, Aplicações e Experimentos com o *Weka*. **Livro da Escola Regional de Informática Rio de Janeiro - Espírito Santo Vitória - ES, Rio das Ostras – RJ, 2004.**

SILVEIRA, Rosemari de Freitas. Mineração de Dados Aplicada à Definição de Índices em Sistemas de Raciocínio Baseado em Casos. **Porto Alegre: CPGCC da UFRGS, 2003**

WITTEN, Ian H.; FRANK, Eibe. *Data Mining: Practical Machine Learning Tools and Techniques*. **San Francisco: Elsevier, 2005**

WITTEN, Ian H.; FRANK, Eibe. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. **San Diego: Morgan Kaufmann Publishers, 2000**