

CENTRO UNIVERSITÁRIO FEEVALE

CLÁUDIO AURÉLIO DA SILVA

DESCOBERTA DE CONHECIMENTO NA BASE DE DADOS DE UMA
ACADEMIA DE MUSCULAÇÃO
(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo, setembro de 2008.

CLÁUDIO AURÉLIO DA SILVA

claudio@feevale.br

DESCOBERTA DE CONHECIMENTO NA BASE DE DADOS DE UMA
ACADEMIA DE MUSCULAÇÃO

(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

As técnicas de mineração de dados, de maneira concisa, são utilizadas para auxiliar a extração de conhecimento. Algumas bases de dados não são aproveitadas da melhor maneira possível, pois informações desconhecidas, que podem se tornar importantes para as organizações, deixam de ser exploradas pelos *decision makers*. O software Weka, através de algoritmos de *data mining*, possibilita um aproveitamento dos dados com maior eficácia. A boa compreensão desse software, das técnicas e dos algoritmos, permite que o conhecimento sobre o ambiente em questão seja realizado de maneira mais completa e eficiente. Este trabalho visa extrair conhecimento da base de dados de uma academia de musculação, através do uso de técnicas de mineração de dados, a fim de obter dados estatísticos sobre os clientes e as atividades em geral dentro desta organização.

Palavras-chave: Mineração de Dados. Descoberta de Conhecimento. Análise de Dados. Classificação. Weka.

SUMÁRIO

MOTIVAÇÃO.....	5
OBJETIVOS.....	7
METODOLOGIA.....	8
CRONOGRAMA.....	9
BIBLIOGRAFIA.....	10

MOTIVAÇÃO

Em 1989 foi formalizado um termo para denominar o abrangente conceito de buscar conhecimento a partir de bases de dados: o KDD – *Knowledge Database Discovery*. Historicamente a Descoberta de Conhecimento em Bases de Dados foi consolidada a partir de várias disciplinas, sendo que entre elas pode-se citar a Estatística, a Inteligência Computacional, o Reconhecimento de Padrões e Banco de Dados (GOLDSCHMIDT, 2005).

KDD é um processo, de várias etapas, para descoberta de informações que sejam úteis e estejam implícitas nas grandes bases de dados. O exponencial aumento no volume de dados que não podem ser restaurados de uma maneira adequada pelos limites e capacidades de consultas dos SGBD's atuais é o que faz com que a utilização do KDD tenha uma grande importância na atualidade (BORGES, 2006).

O uso de técnicas para exploração de grandes quantidades de dados para descoberta de padrões e relações, que demandariam grande trabalho por parte do ser humano, é definido como *Data Mining* (CARVALHO, 2005). A Mineração de Dados é a principal etapa do processo de KDD, onde efetivamente é feita a busca por conhecimentos que possam se tornar úteis no conjunto da aplicação da Descoberta de Conhecimento em Bases de Dados (GOLDSCHMIDT, 2005).

Para a solução de diversos casos que relatam os problemas de mau aproveitamento nas bases de dados, a aplicação de técnicas de mineração auxilia na abstração de conhecimento. Carvalho (2005) cita um exemplo de caso de segmentação de mercados, onde a venda foi

otimizada através da análise e mineração na base de dados que contém os registros de venda, a fim de indicar quais as tendências de aquisição de determinados clientes, considerando vários aspectos. Também exemplifica a situação de uma empresa que utilizou as técnicas de redes neurais artificiais para previsão de mercados financeiros e se tornou uma das mais importantes no ramo, nos EUA, durante sete anos consecutivos, tendo crescimento de sua carteira, neste período, de 25% a 100% ao ano.

Existem diversas técnicas para Descoberta de Conhecimento em Bases de Dados, sendo que neste trabalho será utilizada a de Classificação. A técnica tem como objetivo classificar casos em classes distintas, levando em conta os atributos comuns a um determinado conjunto de objetos, pertencentes a uma base de dados. Esse modelo gerado possibilita a descoberta das classes de novos objetos a serem adicionados na base (SOUSA, 1998).

O processo de classificação é definido por dois passos, onde no primeiro, as características dos dados formam um modelo de classificação, e no segundo, esse modelo criado é empregado com o objetivo de classificar novos objetos. Segundo Passini e Toledo (2002), a construção do modelo pode ser dividida em três fases: o treinamento, onde são estabelecidos parâmetros para se treinar o mesmo, a fase de teste, onde a precisão do modelo é testada com a aplicação de dados diferentes, e a aplicação, que é responsável pela execução da técnica.

Dessa forma, este trabalho visa extrair conhecimento da base de dados de uma academia de musculação, através do uso de técnicas de classificação, a fim de obter dados estatísticos sobre os clientes, além de encontrar alguma relação das atividades dos alunos com a sua frequência na academia, evolução do peso e medidas dos alunos de acordo com faixa etária, sexo entre outras informações. Auxiliar a tomada de decisão dentro da academia, promovendo ações que resultem em um conhecimento mais profundo de seus clientes, e por consequência permitir uma análise de tendências, é uma motivação adicional para a realização deste trabalho.

OBJETIVOS

Objetivo geral

Aplicar uma das técnicas de KDD na base de dados de uma academia de musculação a fim de permitir uma análise de tendências e ter um conhecimento mais profundo das atividades dos clientes e serviços mais utilizados, para auxiliar a tomada de decisões dentro da organização.

Objetivos específicos

- Descrever o processo de KDD e Mineração de Dados, apresentando fundamentação teórica sobre o assunto;
- Estudar o software Weka e como os resultados são apresentados;
- Executar todas as etapas do processo de KDD;
- Aplicar o algoritmo na base de dados;
- Analisar os resultados obtidos e torná-los aplicáveis na academia;
- Apresentar o relatório final com as conclusões e proposta de melhora na organização.

METODOLOGIA

As atividades a serem realizadas para a conclusão deste trabalho estão organizadas da seguinte maneira:

1. **Elaboração do anteprojeto** e a definição do escopo do trabalho com o auxílio do professor orientador;
2. **Pesquisa bibliográfica** sobre o KDD;
3. **Pesquisa bibliográfica** sobre Mineração de Dados e técnica de Classificação;
4. **Estudo das características da base de dados e da ferramenta WEKA;**
5. **Redação do Trabalho de Conclusão I;**
6. **Revisão e entrega do Trabalho de Conclusão I;**
7. **Definição e modelagem da estrutura escolhida** para o tratamento dos dados;
8. **Aplicação do algoritmo nos dados selecionados** utilizando o software WEKA;
9. **Estudo e análise** dos resultados obtidos;
10. **Conclusão.** Desenvolvimento do relatório apresentando os resultados;
11. **Redação do Trabalho de Conclusão II;**
12. **Revisão e entrega do Trabalho de Conclusão II;**

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Ago/08	Set/08	Out/08	Nov/08
1	■			
2	■	■		
3		■		
4		■	■	
5		■	■	■
6				■

Trabalho de Conclusão II

Etapa	Meses				
	Fev/09	Mar/09	Abr/09	Mai/09	Jun/09
7	■				
8		■	■		
9			■	■	
10				■	
11			■	■	■
12					■

BIBLIOGRAFIA

BORGES, Helyane Bronoski. **Redução de Dimensionalidade de Atributos em Bases de Dados de Expressão Gênica**. Curitiba, PR: 2006. 123 p. Dissertação (Mestrado) – Programa de Pós Graduação em Informática. Pontifícia Universidade Católica do Paraná, 2006.

CARVALHO, Luis Alfredo Vidal de. **DataMining : A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Rio de Janeiro: Ciência Moderna, 2005.

FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery: An overview. In: FAYYAD et al. **Advances in Knowledge Discovery and Data Mining**. G. Cambridge-Mass:AAAI/MIT Press, 1996.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: Um Guia Prático: Conceitos, Técnicas, Ferramentas, Orientações e Aplicações**. Rio de Janeiro, RJ: Elsevier, 2005.

JERONIMO, Paulo Marcelo. **Estudo sobre: Data Mining : Data Warehouse : Cases - Data Warehouse**. Novo Hamburgo, RS: 2001. 73 p. Monografia (Bacharelado em Ciência da Computação) – Instituto de Ciências Exatas e Tecnológicas, Centro Universitário Feevale, 2001.

KRANZ, Paulo Henrique. **Business Intelligence: Estudo Aplicado em Cooperativa Médica**. Novo Hamburgo, RS: 2004. 103 p. Monografia (Bacharelado em Ciência da

Computação) – Instituto de Ciências Exatas e Tecnológicas, Centro Universitário Feevale, 2004.

PASSINI, Sílvia Regina Reginato; TOLEDO, Carlos Miguel Tobar. Mineração de Dados para Detecção de Fraudes em Ligações de Água. XI SEMINCO – SEMINÁRIO DE COMPUTAÇÃO, 2002, Blumenau, SC. **Anais do XI Seminco**. Blumenau, SC: s.n., 2002. p. 229- 242.

SOUSA, Mauro Sérgio Ribeiro de. **Mineração de Dados: Uma Implementação Fortemente Acoplada a um Sistema Gerenciador de Banco de Dados Paralelo**. Rio de Janeiro, RJ: 1998. 75 p. Dissertação (Mestrado) – Programa de Pós Graduação de Engenharia. Universidade Federal do Rio de Janeiro, 1998.