
Data Mining

Prof. Me. Ricardo Ávila
ricardo.avila@outlook.com.br

Objetivo :

Apresentar e aplicar as principais *técnicas de mineração de dados* e o *processo de descoberta de conhecimento* com vistas à identificação de padrões importantes e não óbvios em grandes bancos de dados.

Objetivos Específicos:

- Assimilar os principais conceitos acerca de Mineração de Dados (MD);
- Compreender os passos do processo de descoberta de conhecimento em bases de dados (KDD);
- Realizar a análise exploratória dos dados;
- Compreender as principais tarefas da mineração de dados e relacionar as respectivas técnicas;
- Conhecer e saber aplicar as principais técnicas de mineração de dados;
- Utilizar ferramenta para uso de algoritmos de mineração de dados.

Programa da disciplina

1. Introdução e Conceitos
2. Processo de Descoberta de Conhecimento
3. Análise Exploratória de Dados
4. Classificação
5. Agrupamento
6. Associação
7. Tópicos Avançados
8. Projeto

Avaliação

Será realizada através de lista de exercícios, participação em sala de aula e um projeto final.

A Média Final será calculada pela fórmula:

$$\text{Média Final} = (L1*0.25 + L3*0.25 + P2*0.1 + SE*0.4)$$

Onde:

L1 = Lista de Exercícios (individual)

P2 = Participação (individual)

SE = Apresentação de seminário (em grupo)

- Alunos com Média Final ≥ 6.0 e frequência suficiente estarão aprovados.
- Alunos com frequência insuficiente estarão reprovados.

Introdução

- Grande quantidade de informação armazenada em muitas áreas: comércio, indústria, governo, ciência, etc.
- Exemplos:
 - cada compra em um supermercado fica registrada em uma tabela de transações (com o código de barra)
 - todas as chamadas telefônicas (origem, destino, horário, duração, ...)
 - o SUS mantém registro magnético de todos os atendimentos realizados (hospital, data, médico, diagnóstico, procedimentos realizados, ...)
 - Imagens de satélite, bioinformática, ...

BDs atuais

- BDs atuais são feitos para armazenar e recuperar rápida e eficientemente dados operacionais
- BDs atuais são úteis para recuperar dados específicos, mas **não são capazes** de extrair conhecimento genérico
- Exemplo: um sistema bancário recupera rapidamente as últimas movimentações ou o saldo de uma conta, mas teria dificuldade em descrever o perfil do cliente em relação a outros clientes ou determinar se ele seria um bom pagador em caso de fazer um empréstimo

BD x DM

- Banco de Dados

- Encontre todos os clientes com sobrenome “Silva”.
- Identifique os clientes que compraram mais de R\$1.000,00 no último mês.
- Encontre todos os clientes que compraram leite.

- Data Mining

- Encontre todos os clientes com baixo risco, em caso de realizarem um empréstimo.
- Identifique clientes com hábitos de compras similares.
- Encontre todos os itens que são normalmente comprados junto com leite.

Descoberta de conhecimento em bases de dados

Objetivo: extrair conhecimento novo, útil e interessante, implícito em grandes volumes de dados, e representá-lo de forma acessível para o usuário.

Porque extrair conhecimento?

Exemplo de uso de conhecimento

Michael Dell: venda de assinaturas de jornais

Relações da DCBD com outras áreas

- O processo de DCBD utiliza conhecimento de várias áreas:
 - BD
 - IA: aprendizagem de máquina, redes neurais, representação de conhecimento, ...
 - Estatística
 - ...

DCBD: Exemplo

- Em uma empresa de fornecimento de água encanada uma das maiores despesas é com energia elétrica
- Objetivo da empresa: reduzir o consumo de energia elétrica

DCBD: Exemplo

- Dados:
 - consumo diário de água
 - aspectos do tempo (temperatura, umidade do ar, ...)
 - dia da semana e do mês, feriado, férias, ...
- Objetivo da DCBD: prever o consumo de água, de forma a minimizar o bombeamento, e por conseguinte, o consumo de energia elétrica

Bibliografia

- TAN,P-N;STEIMBACH, M; KUMAR,V. **Introduction to Data Mining**. Boston: Addison Wesley, 2006. 769p.
- HAN,J.; KAMBER, M. **Data mining: concepts and techniques**. Morgan Kaufmann, 2006 – 2. edição.
- ADRIAANS, Pieter, ZANTINGE, Dolf. **Data Mining**. Harlow : Addison-Wesley, 1997. 158p.
- FAYYAD, Usama M. et al. **Advances in Knowledge Discovery and Data Mining**. American Association for Artificial Intelligence, 1996. 611p.
- BERRY, Michael J. A.; LINOFF, Gordon. **Data Mining techniques for marketing, sales and customer support**. New York: John Wiley, 1997. 454p.
- PYLE, Dorian. **Data preparation for data mining**. San Diego: Academic Press, 1999. 540p.
- **Artigos de congressos e revistas científicas**