Aula 5

# Sistema Gerenciador de Banco de Dados

Profª Vívian Ariane Barausse de Moura

## Conversa Inicial

- *Data Mining*
  - Conceitos básicos sobre *Data Mining*
- Relação do processo de mineração de dados
- Fases de um processo de descoberta de conhecimento
- Tipos de conhecimento descobertos durante a mineração de dados
- Fases do processo de descoberta

## *Data Mining*

### Conceitos básicos

- De acordo com Ramakrishnan (2008, p. 737), *data mining*, ou mineração de dados, consiste em encontrar tendências ou padrões interessantes em grandes conjuntos de dados para orientar decisões sobre atividades futuras
- Os padrões identificados por essa ferramenta podem fornecer ao analista de dados ideias úteis e inesperadas

- Essas ideias podem ser mais bem investigadas e podem ser aliadas a outras ferramentas de apoio à decisão
- Segundo Elmasri e Navathe (2011, p. 700), os objetivos da mineração de dados estão diretamente relacionados à descoberta do conhecimento
- A mineração de dados costuma ser executada com objetivos finais ou aplicações

- De uma perspectiva geral, esses objetivos se encontram nas seguintes classes
  - Previsão
  - Identificação
  - Classificação
  - Otimização

- Para Elmasri e Navathe (2011, p. 698), a mineração de dados está diretamente relacionada ao *data warehousing.*
- O objetivo de um *data warehouse* é dar suporte à tomada de decisão com dados
- Para ajudar em certos tipos de decisões, a mineração de dados pode ser usada com *data warehouse*

## Aplicações de mineração de dados

- *Marketing*
  - As aplicações incluem análise de comportamento do consumidor com base nos padrões de compra
- Finanças
  - As aplicações incluem análise de crédito de clientes

- Manufatura
  - As aplicações envolvem otimização de recursos como máquinas
- Saúde
  - Algumas aplicações são para descoberta de padrões

## Relações do processo de mineração de dados

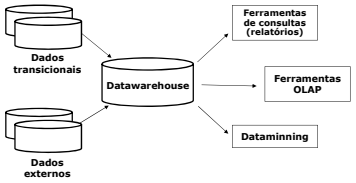
## Relação do processo de mineração de dados com descoberta de conhecimento, estatística e inteligência computacional

- Ramakrishnan (2008, p. 738) defende que a mineração de dados "está relacionada à subárea da estatística chamada análise de dados exploratória, que tem objetivos semelhantes e conta com medidas estatísticas"

- **Relação com demais áreas de estudo, está intimamente relacionada às subáreas da inteligência artificial, chamadas descoberta de conhecimento e aprendizado de máquina**
- **As consultas podem ser realizadas em SQL, usando-se álgebra relacional, com algumas extensões**
- **OLAP fornece idiomas de consulta de nível mais alto, baseados no modelo de dados multidimensional**

- **Mineração de dados fornece as operações de análise mais abstratas**

**Figura 1** – *Data warehouse* e outras tecnologias




FONTE: Caiçara, 2015, pág. 197.

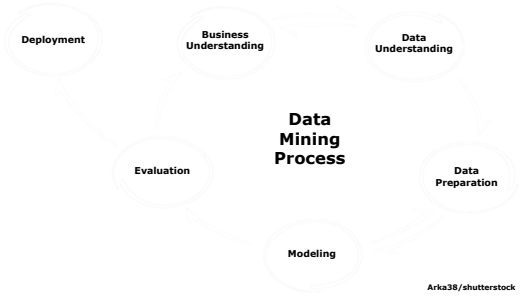
**Problemas**

- **Os dados contêm ruído ou estão incompletos**
- **Se não forem entendidos e corrigidos, é provável que muitos padrões interessantes sejam perdidos e a confiabilidade dos padrões detectados será baixa**

- **O analista precisa decidir quais tipos de algoritmos de mineração são necessários**
- **Aplicá-los em um conjunto bem escolhido de amostras de dados e variáveis, sintetizar os resultados, aplicar outras ferramentas de apoio à decisão e mineração e iterar o processo**

**Fases de um processo de descoberta de conhecimento**

- **A mineração de dados como parte do processo de descoberta do conhecimento é abordada por Elmasri e Navathe (2011) e Ramakrishnan (2008) como "a descoberta de conhecimento nos bancos de dados, que recebe a abreviação KDD (*knowledge discovery in database*); esse processo é composto de algumas etapas".**




Arka38/shutterstock

1. **Obtenção e normalização de dados**
   - **Realizada a seleção dos dados, sobre itens específicos ou categorias de itens**
2. **Limpeza de dados**
   - **O ruído e as exceções são removidos e são realizadas correções ou eliminações de registros incorretos**
3. **Seleção e transformação**
   - **A seleção e a transformação de dados podem ser feitas para melhorar os dados com informações adicionais e reduzir a quantidade de dados**

4. **Mineração**
   - **Aplicados algoritmos de mineração de dados para extrair padrões interessantes**
5. **Avaliação do conhecimento**
   - **Os padrões são apresentados para os usuários finais. O resultado de qualquer etapa no processo KDD pode nos levar de volta a uma etapa anterior para refazermos o processo com o novo conhecimento obtido**

**Tipos de conhecimentos**

- **A palavra "conhecimento" pode assumir diferentes aplicações. É interpretado de forma livre como algo que envolve algum grau de inteligência**
- **Para que chegue nesse patamar, é necessário passar pela transformação de dados brutos da infor-mação para o conhecimento**
- **Conhecimento dedutivo e indutivo**

Yurii_designq/Shutterstock

- **Conhecimento dedutivo deduz novas informações sobre o dado indicado, com base na aplicação de regras lógicas de dedução previamente especificadas**
- **Conhecimento indutivo descobre novas regras e padrões com base nos dados fornecidos**


Elsamnov/shutterstock

## Conhecimentos descobertos na mineração de dados

- Regras de associação
  - Quando uma compradora adquire uma bolsa, ela provavelmente compra sapatos

- Hierarquia de classificação
  - Uma população pode ser dividida em cinco faixas de possibilidade de crédito com base em um histórico de transações de créditos anteriores
- Padrões dentro da série temporal
  - Dois produtos mostraram o mesmo padrão de vendas no verão, mas um padrão diferente no inverno

## Conhecimentos descobertos na mineração de dados

- Padrões sequenciaiss
  - Se um paciente passou por uma cirurgia de ponte de safena para artérias bloqueadas e um aneurisma e, depois, desenvolveu ureia sanguínea alta um ano após a cirurgia, ele provavelmente sofrerá insuficiência renal nos próximos 18 meses

- Agrupamento
  - Uma população inteira de dados de transação sobre uma doença pode ser dividida em grupos com base na similaridade dos efeitos colaterais produzidos

## Fases do processo de descoberta

## Regras de associação

- Uma das principais tecnologias em mineração de dados envolve a descoberta de regras de associação. Muitos algoritmos foram propostos para descobrir várias formas de regras
- O banco de dados é considerado uma coleção de transações, cada uma envolvendo um *itemset* (conjunto de itens)

- **Ramakrishnan (2008) indica que essa regra deve ser lida como se segue: "Se uma caneta é comprada em uma transação, é provável que tinta também seja comprada nessa transação"**

**Regra de associação**

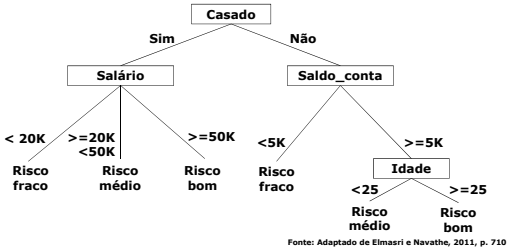
**{caneta} ⇨ {tinta}**

Fonte: Adaptado de Ramakrishnan, 2008, p. 744.

## Classificação

- **O processo de aprender um modelo que descreve diferentes classes de dados também é chamado de aprendizado supervisionado. Utilizando o exemplo da aplicação bancária, os clientes que solicitam cartão de crédito podem ser classificados com risco fraco, risco médio ou risco bom**

**Figura 3 – Árvore de decisão da amostra para aplicações de cartão de crédito**




Fonte: Adaptado de Elmasri e Navathe, 2011, p. 710

## Padrões sequenciais

- **Exemplo citado: se um paciente passou por uma cirurgia de ponte de safena para artérias bloqueadas e um aneurisma e, depois, desenvolveu ureia sanguínea alta um ano após a cirurgia, ele provavelmente sofrerá insuficiência renal nos próximos 18 meses**

- **A descoberta de padrões sequenciais é baseada no conceito de uma sequência de conjunto de itens**
- **A detecção de padrões sequenciais é equivalente à detecção de associações entre eventos com certos relacionamentos temporais**
- **Uma sequência de ações ou eventos é buscada**

## Padrões dentro da série temporal

- **Exemplo citado: dois produtos mostraram o mesmo padrão de vendas no verão, mas um padrão diferente no inverno**
- **Sequências de eventos: cada evento pode ser certo tipo fixo de uma transação**

- **A sequência desses eventos constitui uma série temporal**
- **A série temporal pode ser comparada estabelecendo medidas de similaridade para identificar as vendas que se comportam de modo semelhante**

**Agrupamento**

- **Exemplo citado: uma população inteira de dados de transação sobre uma doença pode ser dividida em grupos com base na similaridade dos efeitos colaterais produzidos**

- **O objetivo do agrupamento é particionar um conjunto de registros em grupos tais que os registros dentro de um grupo sejam similares entre si, e os registros pertencentes a dois grupos diferentes sejam diferentes**
- **Cada grupo é chamado de agrupamento, e cada registro pertence a exatamente um agrupamento**