

# Mineração de Dados

## 01 - Introdução à Mineração de Dados

---

Marcos Roberto Ribeiro



Instituto Federal Minas Gerais - Campus Bambuí

2018

- As organizações atuais estão produzindo cada vez mais dados
- Muitas vezes, as ferramentas tradicionais de análise de dados não podem ser usadas nestes dados
- As vezes, a natureza não trivial dos dados também impede o uso de ferramentas tradicionais
- A mineração de dados combina métodos tradicionais de análise de dados com algoritmos sofisticados para processar dados nestas situações
- Além disso, com a mineração de dados, foi possível analisar novos tipos de dados e avaliar de uma maneira diferente dados analisados por ferramentas tradicionais

## Negócios

- Criação de perfis de clientes
- Vendas direcionadas
- Detecção de fraudes

## Ciência

- Análise de dados climáticos
- Análise de genomas

# O Que é Mineração de Dados?

- A mineração de dados é o processo de descoberta automática de **informações úteis** em grandes bases de dados
- Estas informações úteis não poderiam ser descobertas com ferramentas tradicionais
- Além disto, tais informações podem ser usadas para fazer previsões (ex.: qual a chance de um futuro cliente comprar determinado produto)

# Mineração de Dados e Descoberta de Conhecimento

Mais precisamente, a mineração de dados é uma etapa do processo de descoberta de conhecimento em bancos de dados, do inglês, *knowledge discovery database (KDD)*



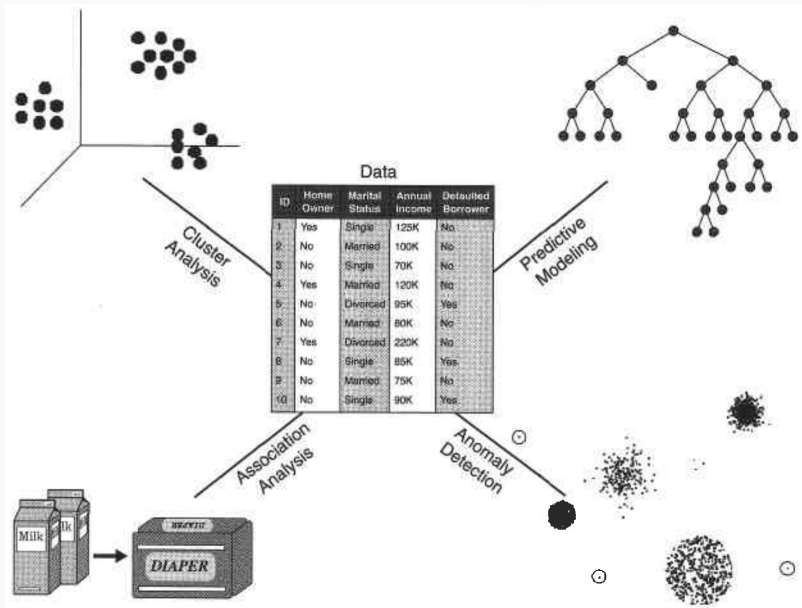
## Preparação dos Dados

- Seleção
- Limpeza
- Transformação

## Análise dos Resultados

- Visualização
- Interpretação
- Predição

# Tarefas de Mineração de Dados



# Tarefas de Mineração de Dados

- Análise de Agrupamentos
- Análise de Associação
- Detecção de Anomalias
- Modelagem de Previsão

# Modelagem de Previsão

- A modelagem de previsão visa construir um modelo para prever valor de um atributo com base nos valores dos demais atributos
- Tipos de Modelagem de Previsão:
  - Classificação:** Previsão de valores discretos. Ex.: determinar se um cliente irá comprar um produto (sim / não)
  - Regressão:** Previsão de valores contínuos. Ex.: prever o preço futuro de uma ação
- O ideal minimizar o erro entre o valor previsto pelo modelo e o valor real



# Exemplo: Prevendo o Tipo de uma Flor

- O banco de dados de flores Íris possui atributos sobre as pétalas e sépalas das flores
- Podemos classificar as flores usando apenas a largura e o comprimento da pétala

## Largura

**pequena:** menor que 0,75

**média:** entre 0,75 e 1,75

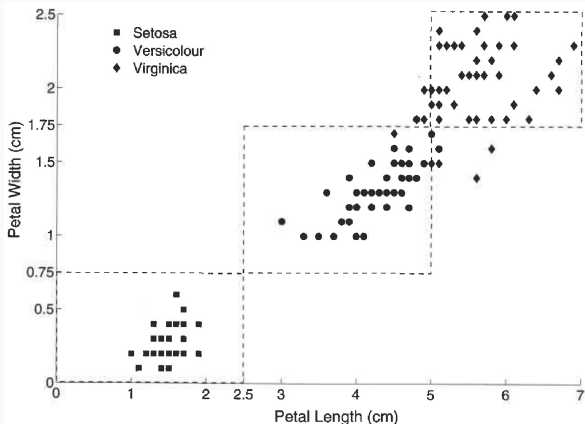
**grande:** maior que 1,75

## Comprimento

**pequeno:** menor que 2,5

**médio:** entre 2,5 e 5,0

**grande:** maior que 5,0



# Análise de Associação

- A análise de associação é usada para descobrir padrões que representem relacionamentos significantes entre os dados
- Geralmente, os padrões descobertos são representados por regras de implicação
- Devido ao tamanho exponencial do espaço de busca, a análise de associação deve extrair os padrões de forma eficiente

## Exemplo: Carrinho de Compras

ID da Transação	Itens
1	Bread, Butter, Diapers, Milk
2	Coffee, Sugar, Cookies, Sakoon
3	Bread, Butter, Coffee, Diapers, Milk, Eggs
4	Bread, Butter, Salmon, Chicken
5	Eggs, Bread, Butter
6	Salmon, Diapers, Milk
7	Bread, Tea, Sugar, Eggs
8	Coffee, Sugar, Chicken, Eggs
9	Bread, Diapers, Milk, Salt
10	Tea, Eggs, Cookies, Diapers, Milk

- Um exemplo de padrão descoberto nas compras é  $\{\text{Diapers}\} \rightarrow \{\text{Milk}\}$
- Isto significa que quem compra fraldas também compra leite

- A análise de agrupamento é usada para agrupar objetos são mais semelhantes
- Um exemplo interessante é o agrupamento de fotos de acordo co os rostos que aparecem nelas

# Detecção de Anomalias

- As técnicas de detecção de anomalias servem para determinar se determinados objetos são muito diferentes de grupos de objetos
- Uma aplicação interessante é a detecção de fraudes em cartões de crédito



BRAGA, L. P. V. (2005).

*Introdução à mineração de dados.*

E-Papers, Rio de Janeiro, 2 edition.



TAN, P.-N., STEINBACH, M., and KUMAR, V. (2009).

*Introdução ao data mining: mineração de dados.*

Ciência Moderna, Rio de Janeiro.