

ASPECTOS GERAIS



- = Exploração de **grandes quantidades** de dados em busca de **padrões consistentes** potencialmente úteis. Mas não explica esses padrões
- Extrai informações **implícitas** não conhecidas previamente
- Descobre novas
 - Correlações
 - Padrões
 - Tendências
- Emprega:
 - Técnicas estatísticas
 - Teorias
 - Tecnologias
 - Processos (*Machine Learning*)
 - Aprendizado de máquina
- Áreas com **ganhos significativos**:
 - Marketing
 - Manufatura
 - Finanças
 - Saúde

OBJETIVOS (P.I.C.O.)

- Previsão (Prever comportamentos futuros)
- Identificação (Identificar a existência de um item, evento ou atividade)
- Classificação (Dividir os dados em classes/ categorias)
- Otimização (Otimizar o uso de recursos limitados)

ANOMALIAS

- = Ponto fora da curva (*outlier*) (exceções, aberrações, ...).
- **Não obedecem** ao comportamento ou modelo dos dados
- Podem ser **tratadas** ou **descartadas**

data mining
= MINERAÇÃO DE DADOS =

DESCOBERTA DO CONHECIMENTO



* TÉCNICAS DE PRÉ-PROCESSAMENTO

- Agregação
- Amostragem
- Redução de dimensionalidade
- Seleção de subconjuntos de recursos
- Criação de recursos
- Binarização e discretização
- Transformação de variáveis

CLASSIFICAÇÃO

- = **Categorizar** elementos segundo múltiplos **atributos** e **critérios**
- É um tipo de **aprendizado supervisionado**: as classes são **pré-definidas** (Antes da análise dos resultados)
- Ferramentas** importantes:
 - Árvore de decisão
 - Redes neurais

REGRESSÃO

- = **Encontrar uma curva/ equação** para ajustar os pontos dos dados
- Prevê um **valor**, ao invés de uma classe.
- **Exemplo:**



data mining
= TÉCNICAS =

REGRAS DE ASSOCIAÇÃO

- Analisam-se vários elementos/ eventos para **descobrir regras** antes desconhecidas que os associam
- Se [algo acontecer], então [algo acontecerá].
- **Exemplo:** se o consumidor comprou leite e açúcar, então, 60% das vezes, também comprará café.
- Tipos:
 - Padrões sequenciais
 - Padrões temporais

AGRUPAMENTO (Clusterização)

- Visa **descobrir** diferentes **agrupamentos** de uma massa de dados + agrupá-los de forma que **ajude a análise**
- **Agrupamento** = coleção de registros **similares** entre si, mas diferentes dos demais.
- Grau de associação entre membros de:
 - Mesmo grupo = forte
 - Grupos diferentes = fraco