



KNOWLEDGE
ENGINEERING
GROUP

António Abelha – Hugo Peixoto

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Extração de Conhecimento

Metodologias

Preparação de dados

Regras de associação

Segmentação

Classificação



KNOWLEDGE
ENGINEERING
GROUP

Extração de Conhecimento

Universidade do Minho
Escola de Engenharia
Departamento de Informática

- A partir de dados estruturados, inferir padrões
- Filtrar informação
- Interpretação de dados

Definição



Motivações

- “Explosão” na **disponibilização de dados**;
- **Fontes de dados** abundantes:
 - Negócios: comércio eletrónico, gestão de stocks, administração de recursos humanos;
 - Web: recolha e disponibilização de informação;
 - Ciências: simulação, aquisição de dados automática, saúde;
 - Sociedade: notícias, câmaras digitais de fotografia e de vídeo, leitores de áudio, funerárias.
- **Solidificação** das Tecnologias de Bases de Dados;
- Disponibilização de **ferramentas automáticas** de procura e arquivo de informação;
- **Armazenamento digital** de informação promove aumento significativo na quantidade de dados disponíveis;
- **Custo do arquivo** de informação diminui drasticamente.



Universidade do Minho
Escola de Engenharia
Departamento de Informática

- “We are drowning in data, but starving for knowledge”

Data Mining: Concepts and Techniques

Jiawei Han, Micheline Kamber

Problema

- “Necessity is the mother of invention”

Idem

- “A necessidade aguça o engenheiro”

PT-version



KNOWLEDGE
ENGINEERING
GROUP

Extração de Conhecimento

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Soluções

- Data Warehousing e técnicas de OLAP (On-Line Analytical Processing);
- Data Mining/Extração de Conhecimento:
 - **Descobrir**, em grandes volumes de dados, sem qualquer formulação prévia de hipóteses, **informações genéricas, relevantes e desconhecidas**, por forma a serem utilizadas em processos de tomada de decisão.



KNOWLEDGE
ENGINEERING
GROUP

Extração de Conhecimento

Universidade do Minho
Escola de Engenharia
Departamento de Informática

- A cada 9 meses duplica a capacidade de armazenamento, pelo mesmo custo;
- Os dados armazenados em bases de dados em todo o mundo duplica a cada 20 meses.

Cenário



KNOWLEDGE
ENGINEERING
GROUP

Extração de Conhecimento

Universidade do Minho
Escola de Engenharia
Departamento de Informática

- Qual a quantidade destes dados que serão alguma vez lidos, acedidos ou procurados por utilizadores humanos?
- É necessário extrair conhecimento para dar sentido e tornar úteis os dados.

Questão



KNOWLEDGE
ENGINEERING
GROUP

Extração de Conhecimento

Universidade do Minho
Escola de Engenharia
Departamento de Informática

- Problemas em que haja necessidade de tomar decisões baseadas em conhecimento;
- Situações caracterizadas por ambientes em constante mutação;
- Casos em que existam consideráveis conjuntos de dados disponíveis.

Para que serve Extração de Conhecimento?



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Exemplos

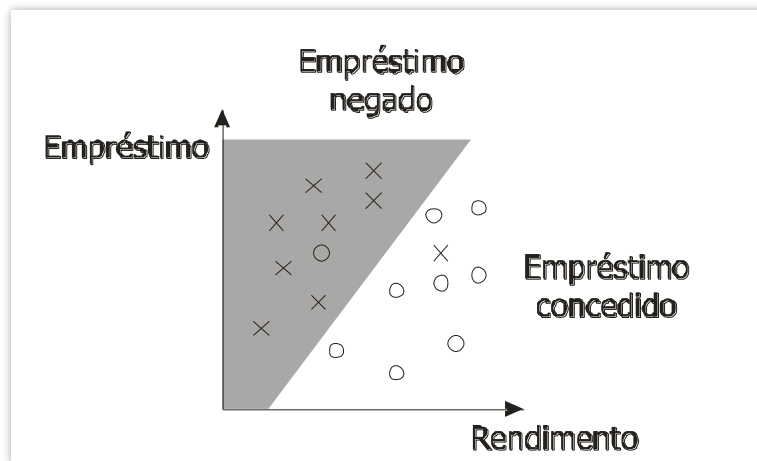
- Identificação de anomalias ou irregularidades em Rx, para apoio ao diagnóstico médico
(a informação contida em chapas de Rx é muito superior à capacidade da visão humana);
- Detecção de fraude com movimentos de cartões de crédito
(detecção de desvios ao padrão normal de utilização);
- Detecção de fraude em serviços de comunicações móveis
(procura de atitudes fraudulentas por identificação de redes sociais);
- Previsão de erros em declarações de impostos
(aplicação de regras de deteção de padrões erróneos).



Objetivos

▪ Classificação:

- Encontrar funções que **associem casos particulares a classes** discretas;
- Este é o objetivo mais comum em projetos de extração de conhecimento.

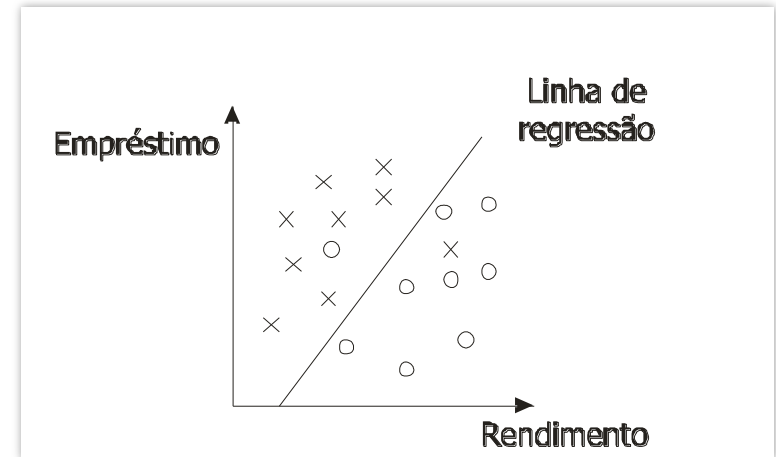


Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo



Objetivos

- **Regressão Linear:**
 - Determinação de uma função que **realize a previsão** de uma variável, representando, de uma forma aproximada, um comportamento variável.

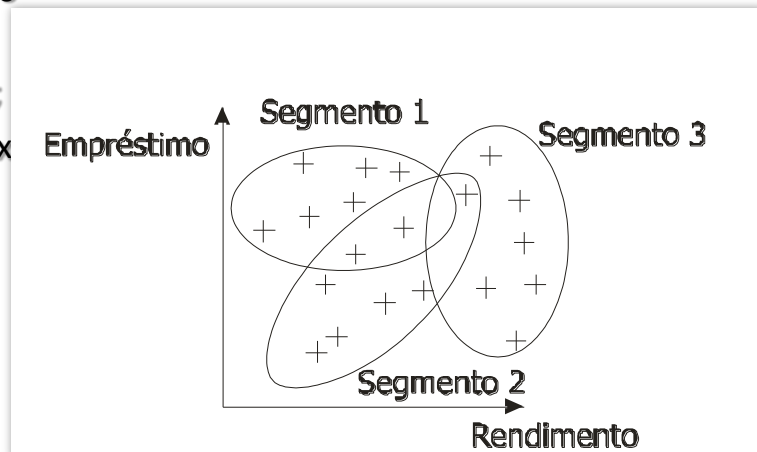


Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo



Objetivos

- Segmentação (Clustering):
 - Identificação de um conjunto finito de segmentos (categorias/clusters) como forma de **descrição dos dados**;
 - A similaridade intra-segmento é alta;
 - A similaridade inter-segmentos é baixa



Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo



KNOWLEDGE
ENGINEERING
GROUP

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Objetivos

- Associação:
 - Procura de modelos que **descrevam dependências** significativas entre variáveis, identificando grupos de dados relacionados.

- Exemplo:
 - Procurar produtos que possam ser colocados em negociação em “pacotes” (o exemplo “canónico” das fraldas e das cervejas).



KNOWLEDGE
ENGINEERING
GROUP

Extração de Conhecimento

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Objetivos

- **Previsão:**
 - **Antevisão de valores** futuros ou **desconhecidos**, baseando a análise em algumas variáveis, geralmente contínuas, e no reconhecimento de padrões.

- **Exemplo:**
 - Prever o preço do petróleo para 2023, com base nos custos anteriores, na evolução política internacional, etc.



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Objetivos

- **Sumariação:**
 - Descrição resumida de subconjuntos de dados;

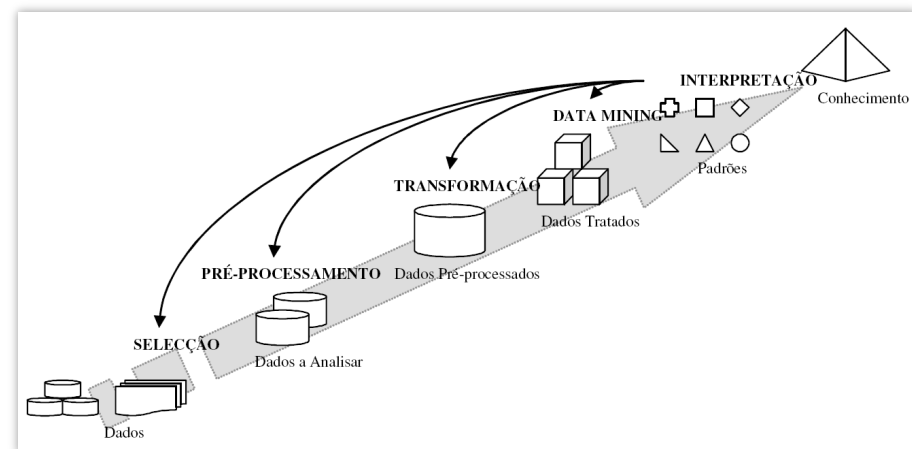
- **Visualização:**
 - Análise gráfica dos dados;

- **Deteção de desvios:**
 - Identificação de alterações significativas em valores normativos.



■ Seleção:

- Os dados provêm de **fontes heterogêneas** e nem sempre apresentam utilidade para a resolução do problema em análise;
- É necessário **escolher** os dados que serão usados no processo de EC.



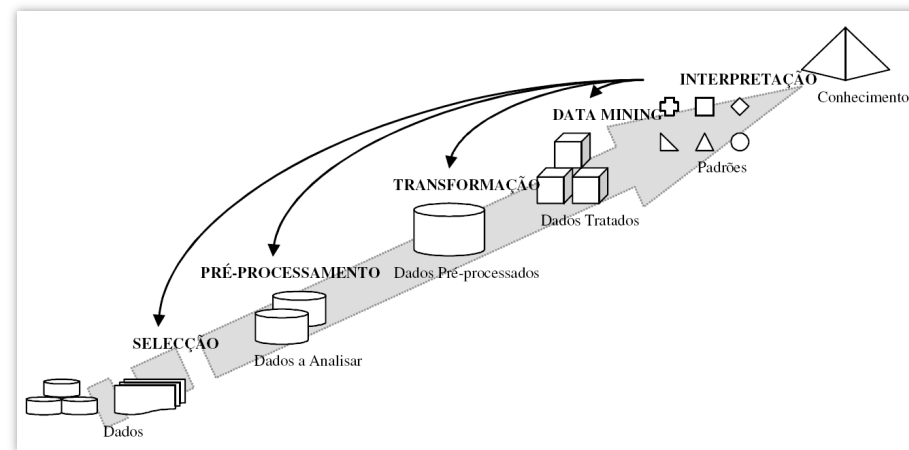
Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo

Processo de Extração de Conhecimento



■ Pré-processamento:

- **Limpeza** dos dados, eliminando ruído, erros e omissões, e **normalizando** a representação da informação;
- Pode representar até 80% do esforço envolvido num processo de extração de conhecimento;
- O sexo pode ser representado por 0/1, M/F, S/N/D.

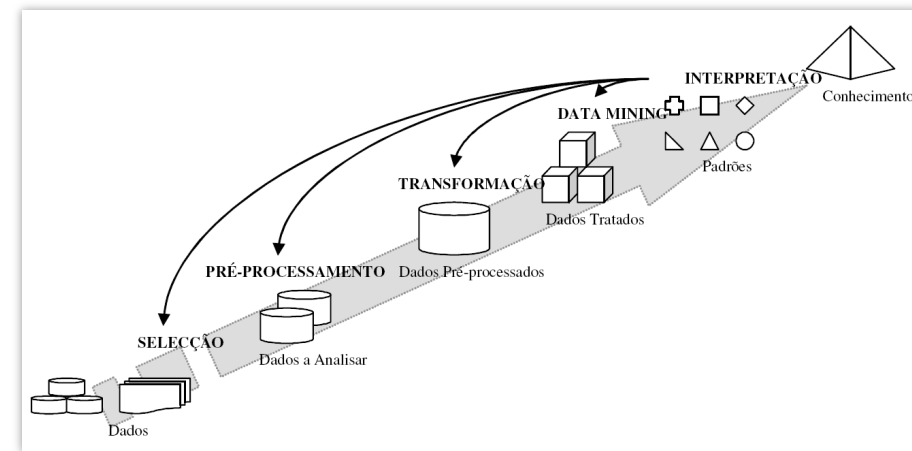


Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo



▪ Transformação:

- Os dados são **classificados** por assunto, cronologicamente, agregando-se ou generalizando-se;
- Constituem um repositório de dados organizado e de grande dimensão (Data Warehouses).

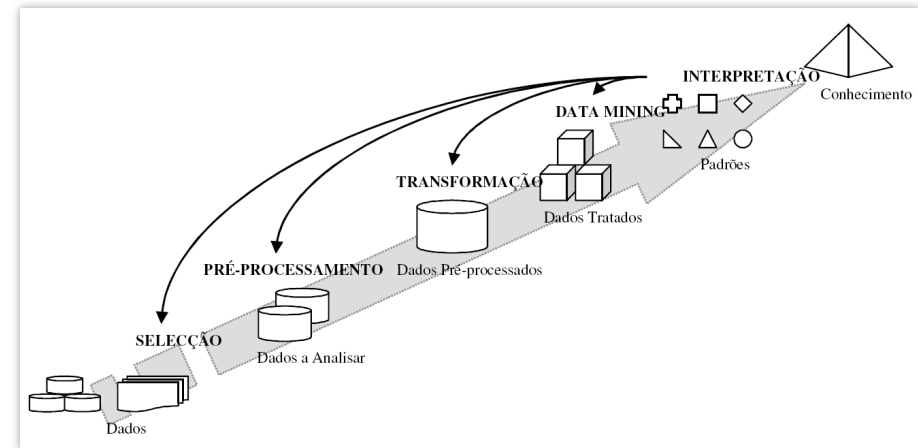


Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo

Processo de Extração de Conhecimento



- **Data-mining:**
 - **Aplicação de algoritmos de extração de conhecimento;**
 - A finalidade é a de seleccionar métodos e técnicas para a resolução do problema em análise.

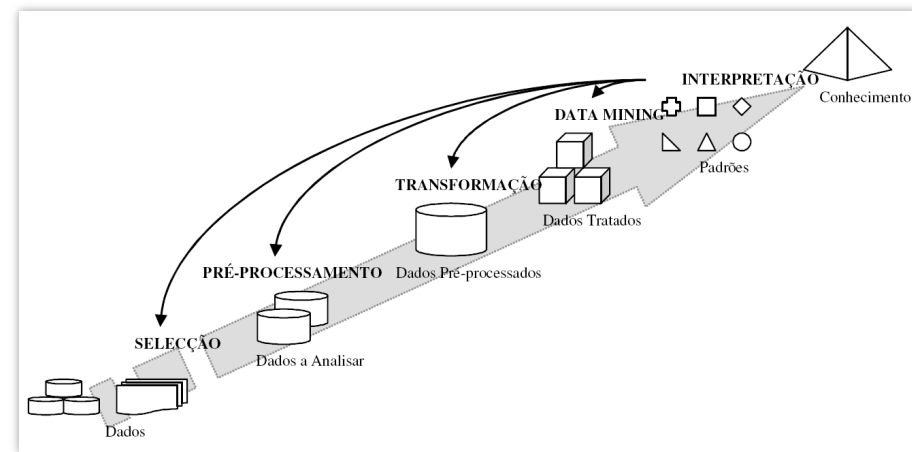


Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo



■ Interpretação:

- Os padrões identificados pela resolução do problema são utilizados para a **tomada de decisão**;
- Considera-se a possibilidade de retornar a qualquer das fases anteriores.



Data Mining – Descoberta de Conhecimento em Bases de Dados
Manuel Filipe Santos, Carla Azevedo



Universidade do Minho
Escola de Engenharia
Departamento de Informática

- **Data Mining: Concepts and Techniques**
Jiawei Han, Micheline Kamber
<http://www.cs.sfu.ca/~han/dmbook>
- **Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations**
Ian Witten, Eibe Frank
- **Data Mining: Descoberta de Conhecimento em BDs**
Manuel Filipe Santos, Carla Azevedo
- **Análise Inteligentes de Dados**
Miguel Rocha, Paulo Cortez, José Maia Neves

Referências bibliográficas