

# **Análise de Dados**

Duarte Trigueiros

3º ano, 3º bimestre

CURSO DE ENGENHARIA INFORMÁTICA

A disciplina ensina a extrair e analisar informação. Tem duas vertentes, o estudo das tarefas e o estudo das técnicas, sendo os alunos orientados no sentido de resolverem problemas, não apenas no sentido de explorarem técnicas. Trata-se de uma disciplina prática, leccionada com o uso de bases de dados empresariais e orientada para a resolução dos problemas que as empresas enfrentam.

Resumo:

Tipos e descrição de dados, estimação de parâmetros, inferência estatística, classificação estatística. Análise de aglomerados. Introdução às redes neuronais. Técnicas de visualização de dados.

Programa:

Primeira parte: a tarefa de extrair e analisar dados.

1. Introdução à análise da informação.
2. Razões e importância da análise da informação nas empresas.
3. Principais aplicações.
4. ETL: extração, transformação e limpeza de dados.
5. A análise de informação:
  - (a) As técnicas disponíveis, o que se consegue com cada.
  - (b) Quadro de tarefas vs técnicas.
6. Primazia das tarefas, gestão por objectivos. Critérios de selecção de técnicas para
  - (a) levar a cabo as tarefas e
  - (b) adaptar-se à organização e suas limitações.
7. Metodologia de desenvolvimento CRISP-DM

Segunda parte: técnicas para a análise de informação.

8. Tipos e descrição de dados.
9. Modelação linear, método dos mínimos quadrados, pressupostos.
10. Regressões lineares simples e múltiplas, estimação de parâmetros. Inferência.
11. Regressões logísticas.
12. Análise de Factores, noção e modelação das fontes de variabilidade.
13. Mapas de Kohonen.
14. Técnicas de visualização de dados.
15. Análise de Clusters: K-means, two-steps.
16. Retro-propagação, o perceptrão de camadas múltiplas, métodos de treino e redução,
17. Indução de árvores: C5 do ID3, GRI. Métodos de redução.
18. Análise de associações (Link Analysis).

Bibliografia:

Dahar (1998), "Seven Methods of Extracting Data for Business Intelligence".

Manual do utilizador, "Clementine".

Trigueiros (2000), folhas da cadeira de Data Mining, ISCTE, Lisboa.