



INSTITUTO POLITÉCNICO DE COIMBRA  
INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO DE COIMBRA

Aprovação do Conselho Pedagógico

20/3/2019

Aprovação do Conselho  
Técnico-Científico

10/4/2019

## **Ficha de Unidade Curricular (FUC) de Data Mining**

**Curso(s):** Mestrado em Análise de Dados e Sistemas de Apoio à Decisão

**Ano Curricular:** 1ºano

**Trimestre curricular:** 3º Trimestre

**Ano lectivo:** 201819

**Docente Responsável:** Fernando Paulo dos Santos Rodrigues Belfo

## Unidade Curricular

Designação:	Data Mining
Curso(s):	Mestrado em Análise de Dados e Sistemas de Apoio à Decisão
Ano curricular:	1ºano
Semestre curricular:	3º Trimestre
Número de ECTS:	5
Horas de contacto:	30

## 1. Corpo Docente

\*\*\*Replicar pelo nº de docentes, sendo que o primeiro deverá ser o responsável da UC.

Nome: Fernando Paulo dos Santos Rodrigues Belfo  
Gabinete: 2.10  
Email: pbelfo@iscac.pt

## 2.Funcionamento

### a) Objectivos:

A unidade curricular visa abordar o processo de descoberta de conhecimento em bases de dados e as tarefas e técnicas mais comuns do Data Mining. Pretende-se que os alunos entendam as possíveis tarefas do Data Mining, designadamente a classificação, segmentação, estimação ou associação. Pretende-se ainda abordar um conjunto de técnicas geralmente usadas na implementação do Data Mining, como árvores de decisão, regras de associação, regressão linear, redes neuronais artificiais, algoritmos genéticos ou redes de Bayes.

Outro importante objetivo é o desenvolvimento dum projeto de Data Mining, usando dados reais duma organização ou dados abertos. Como metodologia de implementação será utilizada o CRISP-DM (Cross-Industry Standard Process for Data Mining), designadamente as fases de análise e compreensão do negócio, seleção e compreensão dos dados, preparação dos dados, modelização e avaliação dos modelos. O software de referência utilizado na prática será o WEKA.

### b) Regime de frequencia e metodologia de avaliação:

As aulas serão lecionadas em regime teórico-prático, em salas de computadores. Será utilizada a metodologia expositiva na apresentação dos conceitos da unidade curricular, apoiada na experimentação prática através da realização de exercícios de aplicação dos conceitos transmitidos, alguns em papel e outros nos computadores. A experimentação prática mais importante será conseguida através dum projeto de Data Mining onde os alunos escolherão uma organização ou fenómeno real, que analisarão e usarão para aplicar conceitos, tarefas e técnicas apreendidas na teoria.

O processo de avaliação compreenderá duas componentes, a realização de um exame final escrito (EFE), sem consulta, que decorrerá em qualquer uma das épocas de exames regulamentares e a realização de um projeto prático (PP), cujos resultados serão apresentados em sessão especialmente marcada para o efeito, que decorrerá em qualquer uma das épocas de exames regulamentares. A Nota Final da unidade curricular será calculada como a média aritmética dessas duas componentes, arredondada ao inteiro mais próximo. A sua fórmula é a seguinte:

$$\text{Nota Final} = (\text{EFE} + \text{PP}) / 2$$

A Nota Final calcula-se fazendo um arredondamento do resultado da fórmula anteriormente apresentada para o valor inteiro mais próximo. No caso de haver uma equidistância para o inteiro inferior e para o superior, o arredondamento é feito para o inteiro superior. Por exemplo, se a nota final der 9,5 valores, então arredondará para os 10 valores.

### c) Programa:

1. Introdução à Descoberta de Conhecimento em Bases de Dados
2. Metodologias e Especificações
3. Atividades e Tarefas

Atividades preditivas (ou supervisionadas)

- Classificação
- Previsão
- Análise de Tendências (Séries Temporais)

#### Atividades descritivas (ou não-supervisionadas)

- Agrupamento
- Sumarização (e visualização)
- Associação

#### 4. Técnicas de Data Mining

- Árvores de Decisão
- Regras de Associação
- Regressão Linear
- Redes Neurais Artificiais
- Conjuntos Difusos e Lógica Difusa
- Redes de Bayes

#### **d) Bibliografia:**

- Berry, M, & Linoff, G. (1997). *Data Mining Techniques: For Marketing, Sales, and Customer Support*. New York: John Wiley & Sons.
- Camilo, C., & Silva, J. (2009). *Mineração de dados*. Goiânia: Universidade Federal de Goiás.
- Chakrabarti, S., Cox, E., Frank, E., Güting, R., Han, J., Jiang, X., Neapolitan, R. (2008). *Data Mining: Know It All*. Burlington: Morgan Kaufmann.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-By-Step Data Mining Guide*. U.S.A.: SPSS, CRISP-DM Consortium.
- Delen, D.. (2014). *Real-World Data Mining*: FT Press.
- Larose, D. (2005). *Discovering Knowledge in Data*. Hoboken: John Wiley & Sons, Inc.
- North, M. (2012). *Data mining for the masses: A Global Text Project Book*.
- Santos, M., & Azevedo, C. (2005). *Data Mining: Descoberta de Conhecimento em Bases de Dados*. Lisboa: FCA.
- Witten, I., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann.
- Loureiro, A., Lourenço, J., Costa, E. & Belfo, F. (2014). *Indução de Árvores de Decisão na Descoberta de Conhecimento: Caso de Empresa de Organização de Eventos*. VI Congresso Internacional de Casos Docentes em Marketing Público e Não Lucrativo, Coimbra, Portugal, December 19
- Pimenta, C.; Ribeiro, R.; Sá, V.; and Belfo, F. (2018). *Fatores que Influenciam o Sucesso Escolar das Licenciaturas numa Instituição de Ensino Superior Portuguesa*. In *Proceedings of the 18.<sup>a</sup> Conferência CAPSI held in Santarém, Portugal, 12-13 October 2018*. Paper 65.