



**ISEL**

INSTITUTO SUPERIOR  
DE ENGENHARIA DE LISBOA

# **AMD**

## Aprendizagem e Mineração de Dados

UC obrigatória de MEIM e optativa de MEIC  
(inverno 2021/2022)

<https://isel.pt/disciplinas/aprendizagem-e-mineracao-de-dados-meim>

# Sumário

1. Enquadramento
2. Objetivos e resultados de aprendizagem
3. Conteúdos programáticos
4. Avaliação
5. Planeamento
6. Docentes e Contactos
7. Bibliografia
8. Ferramentas e recursos

# 1. Enquadramento

- AMD usa conceitos de:
  - Bases de dados e sistemas de informação
  - Aprendizagem automática
  - Programação
  - Estatística

Precedida por AA na LEIM

Sucedida por MDLE (UC optativa) no MEIM e no MEIC

## 2. Objetivos

AMD visa dotar os estudantes das seguintes competências:

1. Construir *dataset* a partir de repositórios, e.g., modelo relacional ou texto Web, considerando a estrutura e semântica, com o objetivo de colocar hipóteses e interpretar resultados

2. Preparar dados via des-normalização, composição e discretização

3. Explorar as características, opções, vantagens e limitações dos métodos de classificação:

- a) de suporte estatístico

- b) baseados na indução de árvores de decisão

- c) baseados em aprendizagem competitiva

## 2. Objetivos

AMD visa dotar os estudantes das seguintes competências:

4. Introduzir a análise de séries temporais; adaptação de *dataset* para aplicar (neste contexto) métodos de classificação supervisionada
5. Explorar métodos não-supervisionados baseados em instâncias
6. Explorar os métodos de procura de regras de associação e evidenciar a diferença em relação à classificação e agrupamento
7. Avaliar a aprendizagem via estimação de erro suportado nas noções de conjuntos de treino, validação e teste

# 3. Conteúdos Programáticos

- I. Gerar e exportar *dataset* a partir do modelo relacional e dados Web; domínios numéricos e nominais e valores omissos
- II. Abordagens não-supervisionadas e supervisionadas à discretização
- III. Classificação com formulação de Bayes e estimadores Laplace
- IV. Indução de árvores de decisão; informação intrínseca, ganho de informação, rácio do ganho e “gini index”; métodos ID3 e C4.5; sobre-ajuste e poda (“pre/post-pruning”); “learning-vector-quantization”, operadores de atração e repulsa e passo-de-aprendizagem

# 3. Conteúdos Programáticos

V. Agrupamento e classificação baseada em instâncias; funções distância com atributos numérico, nominal e omissos; procura de vizinhos com KD-Tree e suporte ao kNN (classificação) e K-means (“clustering”)

VI. Regras de associação; “market-basket analysis”, “rule-space” e avaliação (suporte e confiança); método APRIORI e H-Mine

VII. Taxa de erro e conjuntos de treino, validação e teste; validação cruzada e “bootstrap”; erros e custos; matriz confusão, Kappa e ROC (uni/multi-classe)

## 4. Avaliação - regra de cálculo

- Classificação Final =  
 $0,5 * \text{Classificação Teórica} + 0,5 * \text{Classificação Prática}$
- **Componente teórica ( $\geq 9,5$  valores)**
  - Teste global final na época normal ou na época de recurso
- **Componente prática ( $\geq 9,5$  valores)**
  - Realização de fichas durante o semestre
    - exercícios para explorar e consolidar a compreensão de conceitos teóricos
    - exercícios cuja resolução contribui com uma componente a integrar no projeto final
  - Realização de projeto final (iniciado durante o semestre)
  - Discussão final do projeto e das fichas



## 4. Avaliação – notas importantes

- **Todos os trabalhos entregues serão previamente analisados através de uma ferramenta informática de detecção de plágio**
- **A existência de situações de plágio no trabalho prático levará à:**
  - **anulação de todos os trabalhos envolvidos**
  - **reprovação imediata na unidade curricular, de todos os estudantes envolvidos**
- **Só serão aceites os trabalhos cujos autores coincidam com a constituição dos grupos no sistema Moodle**
- **Eventuais desistências deverão ser comunicadas ao docente da turma**
- **Não serão aceites entregas do trabalho prático, após a data limite**

# 5. Planeamento

Semana	Atividade
4/out	Início do período letivo
18/out – 22/out	Kick-off do primeiro projeto (projeto final A)
22/nov– 26/nov	Kick-off do segundo projeto (projeto final A1)
06/dez-10/dez	Kick-off do terceiro projeto (projeto final B)
22/jan	Final do período letivo
jan/fev	Realização da discussão e do teste global

## 6. Docentes e Contactos

Docente da turma MI1D-MM1D - Artur Ferreira

- [aferreira@deetc.isel.ipl.pt](mailto:aferreira@deetc.isel.ipl.pt) [artur.ferreira@isel.pt](mailto:artur.ferreira@isel.pt)

Docente da turma MI1N-MM1N - Paulo Trigo (regente da UC)

- [ptrigo@deetc.isel.ipl.pt](mailto:ptrigo@deetc.isel.ipl.pt) [paulo.trigo@isel.pt](mailto:paulo.trigo@isel.pt)

Páginas no sistema Moodle

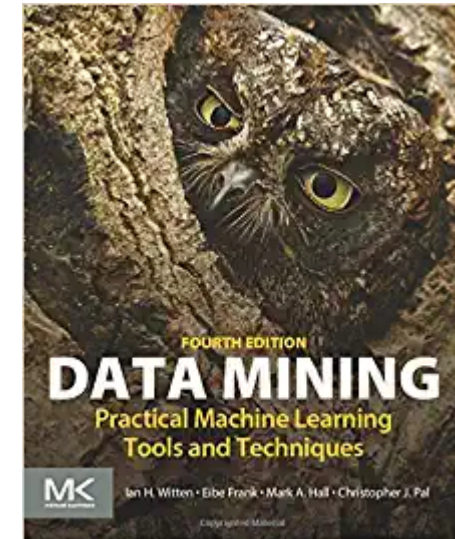
AMD-UCC (conteúdos comuns)

AMD-MM1N-MI1N (turma noite)

AMD-MM1D-MI1D (turma dia)

# 7. Bibliografia

.Witten, H. I., Frank, E., Hall, M. A., and Pal, C. J. (2016). Data Mining - Practical Machine Learning Tools and Techniques. (4th ed.). Morgan-Kaufmann



Orange Data Mining Library  
Documentation. (2018).  
Orange Data Mining.

<https://readthedocs.org/projects/orange-data-mining-library/downloads/pdf/latest/>



# 8. Ferramentas e recursos

- Linguagem Python
- Orange