

## 1. Introdução

A primeira meta (M1) da avaliação prática, tem como objetivo produzir uma abordagem sistemática e organizada à preparação de um *dataset* para a sua posterior submissão a algoritmos de *Supervised Learning* e *Unsupervised Learning*.

Assim, a seleção do *dataset* deverá ter em consideração que M1 será o suporte para o desenvolvimento da meta seguinte: M2 (*supervised learning*; *unsupervised learning*).

## 2. Seleção do Dataset

A seleção do *dataset* deve ser efetuada tendo em consideração que deve:

- conter dados reais<sup>1</sup>;
- configurar um problema de classificação binária;
- possuir entre 8 e 12 *features* (atributos);
- possuir uma dimensão tal que permita a constituição de um *training set* e de um *test set* com um número de instâncias suficiente para assegurar uma validação consistente. Como tal, o número total de instâncias deve estar compreendido no intervalo:  $1000 \leq N \leq 10000$ ;
- ter pelo menos 1 atributo de cada tipo (*numerical*, *categorical*, *ordinal*).

Tendo em consideração os requisitos estabelecidos, os alunos têm total liberdade para pesquisar e selecionar o *dataset*.

---

<sup>1</sup> No contexto desta avaliação, a data de obtenção dos dados não é relevante.

### 3. Meta M1

Esta meta de avaliação deve ser implementada de acordo com os seguintes requisitos:

- Deve ser efetuada uma descrição completa dos dados assim como a correspondente visualização;
- Os dados devem ser preparados para serem submetidos a algoritmos de *Supervised Learning* e *Unsupervised Learning*;
  - Todas as operações de limpeza/transformação/codificação dos dados devem ser devidamente justificadas;
  - Devem ser implementados *pipelines* de transformação dos dados.

### 4. Avaliação

Na implementação desta meta de avaliação devem ser observadas as seguintes regras:

- os grupos são formados por 2 elementos;
- deve ser elaborada uma pequena apresentação (ppt; máximo 10 minutos) a apresentar no dia 22/11/2021. Os dois elementos do grupo terão de ter uma participação equivalente durante a apresentação.
- deve ser elaborado um pequeno documento em que apresente de forma totalmente objetiva e condensada (máx. 10 páginas) a abordagem seguida;
- implementação com base em Jupyter Notebook;
- esta meta de avaliação representa 3 valores (15%) da nota final da disciplina.
- atempadamente será disponibilizado um pequeno formulário com os horários possíveis para a realização da apresentação;
- todos os elementos de avaliação (\*.ipynb; \*.csv; \*.doc; \*.ppt) devem ser entregues na plataforma moodle, num ficheiro comprimido (.zip), **até às 23h59 de 20/11/2021**, com a seguinte nomenclatura:

*primeiroUltimoNomeAluno1\_nºAluno1\_primeiroUltimoNomeAluno2\_nºAluno2*

**Exemplo:** *FernandoPessoa\_1231234\_MiguelTorga\_1234123*