



UNIDADE CURRICULAR: Raciocínio e Representação do Conhecimento

CÓDIGO: 21097

DOCENTE(s): Joaquim Neto

ANO LETIVO: 2025/2026

OBJETIVO

O trabalho a desenvolver no âmbito do E-fólio B tem como objetivo aplicar as técnicas de aprendizagem supervisionada estudadas, recorrendo à linguagem R:

- **Árvores de decisão;**
- **K-vizinhos mais próximos (k-NN);**
- **Redes neuronais.**

Pretende-se que o estudante seja capaz de construir modelos de classificação, avaliar o seu desempenho e interpretar os resultados obtidos.

GERAÇÃO DO DATASET

Cada estudante deve gerar o seu dataset individual utilizando a função **gerar_meu_dataset()**, incluída no ficheiro **gerar_meu_dataset_efB.R**, disponibilizado com o presente enunciado, que deverá descarregar.

Para gerar o seu dataset, deve executar os seguintes comandos no RStudio.

- **source("path_to_file/gerar_meu_dataset_efB.R")**
- **dados <- gerar_meu_dataset(SEU_NUMERO_ESTUDANTE)**

Onde: path_to_file, deverá estar na forma(exemplo): **"c:/xxx/yyy/gerar_meu_dataset_efB.R"**

Obterá algo semelhante ao que se apresenta na imagem a seguir.

```

> dados <- gerar_meu_dataset(1996)

=====
Dataset gerado com sucesso
Número de estudante: 1996
Tema: Seguros
Número de observações: 75
Ficheiro: dataset_seguros_1996.csv
Distribuição da classe:

  AltoRisco BaixoRisco
           37         38
=====

> head(dados)
  Idade AnosCarta Acidentes Tipoveiculo UsoAnual  Classe
1    42         22         3           3    12587 AltoRisco
2    23          5         2           2     6524 AltoRisco
3    35          7         4           2     8150 AltoRisco
4    51         16         3           2     6952 AltoRisco
5    59         35         1           3    11167 BaixoRisco
6    75         52         1           1    10160 BaixoRisco

```

O ficheiro .csv gerado é guardado na pasta de trabalho, que pode obter com o comando `getwd()`

Há que considerar o seguinte:

- O dataset será diferente para cada estudante;
- A temática será automaticamente definida;
- O dataset será sempre o mesmo para o mesmo número de estudante.

TRABALHO A DESENVOLVER

O estudante deve começar por identificar as variáveis preditoras e a variável-alvo, descrevendo o significado de cada uma no contexto do tema atribuído. Deve ainda analisar a distribuição da variável de classe e apresentar sumários estatísticos relevantes dos dados.

Pretende-se, de seguida, que aplique **as três técnicas de aprendizagem supervisionada estudadas — árvores de decisão, k-vizinhos mais próximos (k-NN) e redes neuronais — ao dataset fornecido**, construindo os respetivos modelos, avaliando o respetivo desempenho com métricas adequadas e interpretando os resultados obtidos. Deve sempre fundamentar/justificar as escolhas/decisões efetuadas.

Espera-se, ainda, que compare os diferentes métodos, identificando semelhanças e diferenças, e que discuta qual apresenta melhor desempenho no problema em análise.

O estudante deverá ainda apresentar uma análise crítica dos resultados obtidos, refletindo sobre o desempenho dos modelos e a adequação ao problema em estudo. Deve ainda discutir a qualidade dos dados utilizados, identificar eventuais limitações e possíveis melhorias.

ENTREGÁVEIS

Deve ser entregue arquivo ZIP com, pelo menos, os seguintes ficheiros:

- Relatório em PDF
- Ficheiro CSV com o dataset
- Código em R utilizado

CRITÉRIOS DE AVALIAÇÃO

O E-Fólio vale 4,0 valores, de acordo com os seguintes critérios de avaliação:

- Análise dos dados: 10%
- Árvore de decisão: 20%
- k-NN: 20%
- Rede neuronal: 20%
- Avaliação e comparação: 15%
- Discussão crítica: 10%
- Relatório (Qualidade, organização, apresentação, clareza e fundamentação) – 5%

NOTAS IMPORTANTES

- Trabalho individual;
- Cada *dataset* é único e só são considerados trabalhos cujo *dataset* seja o gerado com o respetivo número de estudante
- Código deve ser claro;
- Descontos: Detecção de fraude (total ou parcial): 100%
- O texto do relatório não pode ser manuscrito e não pode ser em formato de imagem; apenas serão aceites imagens de *screenshots* do R, designadamente do código e dos gráficos.

Bom Trabalho

Joaquim Neto

Mai de 2026