**Avaliação Formativa**

**Equipe (máximo 3 alunos)**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Treinamento e avaliação de classificadores para problema de classificação de câncer de mama a partir de imagem digital, em maligno ou benigno. A base foi criada por pesquisadores da Universidade de Wisconsin nos Estados Unidos.**

**Link para a base no Scikit Learn:**

<https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html>

Qtde de atributos igual a 31, sendo:

classe (0=maligno ou 1=benigno)

30 atributos contínuos extraídos do núcleo das células em imagem de câncer de mama.

Tipo dos atributos: numéricos

Qtde de instâncias: 569 (212 malignos e 357 benignos)

Protocolo experimental a ser utilizado: validação cruzada com 5 folds.

Indutores a serem avaliados: Árvores de decisão, KNN, Naive Bayes.

**Tabela de Resultados**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Indutor** | **Taxa de Acerto (%)** | **F1\_score** | **Precisão** | **Recall** |
| Árvores de Decisão |  |  |  |  |
| KNN |  |  |  |  |
| KNN weighted (com pesos) |  |  |  |  |

**Considerando o melhor resultado encontrado, responda:**

A.1) Qual a taxa de acerto de cada classe?

R: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A.2) Informe a matriz de confusão.

R: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A.3) Informe o valor dos parâmetros utilizados no treinamento.

R: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A.4) Há diferença significativa entre acurácia e F1\_score?

R: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Leitura do arquivo:**

**from** **sklearn.datasets** **import** [load\_breast\_cancer](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html#sklearn.datasets.load_breast_cancer)

X, y = [load\_breast\_cancer](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html#sklearn.datasets.load_breast_cancer)(return\_X\_y=**True**)

1. **Treinamento e avaliação de modelos de regressão para estimativa do nível de progressão de diabetes de uma pessoa.**

Qtde de atributos igual a 10, sendo:

10 atributos numéricos (idade, sexo, massa corpórea, média pressão sangue, dentre outros)

Alvo = nível progressão diabetes (valor entre 25 e 346)

Tipo dos atributos: numéricos

Qtde de instâncias: 442

Nome do arquivo: diabetes

Protocolo experimental a ser utilizado: validação cruzada com 5 folds.

Indutores a serem avaliados: Árvores de regressão, KNN.

Link para a base no Scikit Learn

<https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html>

**Tabela de Resultados**

|  |  |  |
| --- | --- | --- |
| **Indutor** | **Coeficiente de Determinação (R2)** | **Erro Médio Absoluto (MAE)** |
| Árvores de Regressão |  |  |
| KNN |  |  |
| KNN weighted (com pesos) |  |  |

Considerando o melhor resultado encontrado, responda:

B.1) Informe o valor dos parâmetros utilizados no treinamento do modelo.

Leitura do arquivo:

**from** **sklearn** **import** datasets

X, y = [datasets.load\_diabetes](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html#sklearn.datasets.load_diabetes)(return\_X\_y=**True**)

**Forma de entrega: PDF**

**Data de entrega: 5/04/2022**