**Avaliação Somativa 2**

**Aluno**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Treinamento e avaliação de um modelo 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>

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**)

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.

Encontre uma solução avaliando técnicas monolíticas e *ensembles*.

* Alternativas de ensembles de classificadores a serem avaliadas: BaggingClassifier() e RandomForestClassifier()
* Alternativas de classificadores monolíticos a serem avaliados: MLPClassifier() e SVC().

Apresente abaixo (na tabela) o melhor resultado observado e abaixo da tabela insira a matriz de confusão.

Busque pela melhor acurácia (taxa de acerto) e o menor esforço computacional.

**Apenas o Melhor Resultado**

|  |  |  |
| --- | --- | --- |
| **Técnica** | **Parâmetros** | **Taxa de Acerto (%)** |
|  |  |  |

- < insira aqui a matriz de confusão>

- Responda: qual a taxa de acerto de cada classe?

Classe 0 (malignos):

Classe 1 (benignos):

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

Nome do arquivo: diabetes

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

Link para a base no Scikit Learn

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

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

**Tabela de Resultados**

|  |  |  |
| --- | --- | --- |
| **Técnica** | **Parâmetros** | **MAE (Mean Absolute Error)** |
|  |  |  |

* Alternativas de ensembles de regressores a serem avaliados: BaggingRegressor() e RandomForestRegressor().
* Alternativas de regressores monolíticos a serem avaliados: MLPRegressor() e SVR().

**ENTREGA EM ATÉ 48h.**

**HORÁRIO LIMITE: 24h de 16/06/2022**

**Forma de entrega: PDF no AVA.**

**Anexar documento no formato PDF !!!!!**