|  |  |
| --- | --- |
|  | Curso de Data Science  Prof. MSc. Eng. Marcelo Bianchi |

**Trabalho 2 – Curso de Data Science**

Integrantes do Grupo:

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Utilizar UCI Web Site <https://archive.ics.uci.edu/ml/index.php>



Escolhe **um dataset de classificação** no site UCI. A restrição é que cada grupo deverá escolher apenas um (1) único dataset não podendo haver repetições entre os grupos.

**Machine Learning – Predição - Usando Classificação**

1. ( 0,5 ponto ) Importar o Data Set e aplicar a técnica Missing Data
2. ( 0,5 ponto ) Dividir o dataset entre o Training Set e o Test Set
3. ( 0,5 ponto ) Aplicar Feature Scaling (Se for aplicável, senão for então justificar)
4. ( 0,5 ponto ) Aplicar Dummy Variable (Se for aplicável, senão for então justificar)
5. (7,0 pontos) Aplicar e Comparar os resultados dos algoritmos de Classificação

•(1,0 ponto) Logistic Regression (Classification)

• (1,0 ponto) KNN - K Nearest Neighborhood

• (1,0 ponto) SVM - Support Vector Machines

• (1,0 ponto) Kernel SVM

• (1,0 ponto) Naive Bays

• (1,0 ponto) Decision Tree (Classification)

• (1,0 ponto) Random Forest (Classification)

1. (0,5 ponto ) Construir os Gráficos
2. (0,5 ponto) Construir a Confusion Matrix