

Enunciado – AC2a

Até 3 pessoas.

Entregável: PDF

Nome:

RA:

Nome:

RA:

Nome:

RA:

Predição de Cancer

Informações do atributo:

- a) ID Número de identificação
- b) Diagnóstico (M = maligno, B = benigno) **(avariável para predecir)**

Lista de exercicios

- 1- Lee o dataset “Cancer_diagnostic.csv” e segui os seguintes passos.
- 2- Qual é o tamanho do dataset?
- 3- Verifique colunas e tipos de dados
- 4- Mostra o resumo estatístico do dataset (pd.describe)
- 5- O dataset está balanceado?
- 6- Mostra uma gráfica de pizza do dataset
- 7- Mostra o gráfico de boxplot do dataset
- 8- Tem valores ausentes? Se sim quais?
- 9- Realiza a padronização (standardization)
- 10- Se tiver outlier faz a remoção de deles e mostra o boxplot do dataset.
- 11- Divide o dataset em training (80%) e testing (20%)
- 12- Faz a classificação usando o algoritmo KNN com $K = 3$, $K = 5$, $K = 7$ e mostra as métricas de accuracy, recall e precision para cada K