

Nome: Sarah Abrahão Ferreira

Número de Matrícula: 2022.1023.4

Turma: 03

## Trabalho 1: Diferenciando classes

Em primeiro lugar, foi necessário importar as bibliotecas para visualização de dados e exercer a programação em Python pelo Google Collab. Posteriormente, foi escolhido o conjunto de dados “Health care: heart attack possibility” (Cuidados com a saúde: possibilidade de ataque cardíaco), do site

<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.

```
[28] # Carregar o conjunto de dados Heart Attack Possibility
data = pd.read_csv('heart.csv')

[29] #Atribuir os nomes corretos a cada variável
data.columns = [
    'age',
    'sex',
    'cp',
    'trestbps',
    'chol',
    'fbs',
    'restecg',
    'thalach',
    'exang',
    'oldpeak',
    'slope',
    'ca',
    'thal',
    'target',
]
```



Figura 1: Pulmão

Fonte: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

Após importar os dados, foi gerada a seguinte tabela:

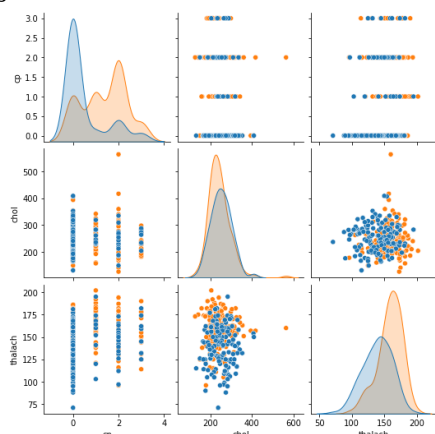
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
5	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
6	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1

Por consequência, foi preciso explorar a variável de interesse nesse projeto que é a previsão do incidente: se é provável (likely – 1) ou não provável (not likely – 0) que ocorra um ataque cardíaco.

```
[30] #frequência por probabilidade
data['target'].value_counts()

1    165
0    138
Name: target, dtype: int64
```

Nesse momento, foram gerados gráficos contendo a relação entre todos os dados:



Analisando esses gráficos, foram escolhidas outras três variáveis: “cp – Chest Pain type”, “chol – cholesterol in mg/dl” e “thalach – maximum heart rate achieved”. Depois, foram utilizados os códigos que geram a figura, nomeiam os eixos, configuram a legenda e realizam a animação:

```
[15] # filtro: 4 variáveis
data = data[
    'target',
    'cp',
    'chol',
    'thalach'
]

#carrega os eixos criando um e mais se necessário
ax = fig.gca(projection='3d')

#cria lista de cor considerando qualidade
colors = data['target'].map({0: 'g', 1: 'r'}).values

ax.scatter(
    data['cp'],
    data['chol'],
    data['thalach'],
    c=colors
)

ax.legend(handles=legend_elements, loc='best')

#ajusta o eixo do gráfico
plt.tight_layout()
plt.show()
```

Então, foi gerado o seguinte gráfico:

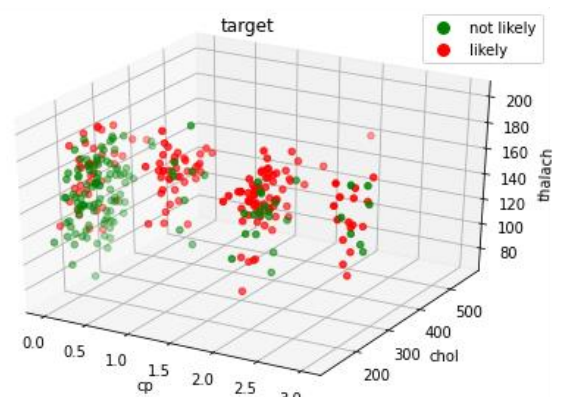


Figura 2: Visualização de 4 dimensões do Heart Attack Possibility – Fonte: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

Por fim, conclui-se que o gráfico contém quatro características e dentro de uma delas, a da probabilidade do ataque cardíaco (target). Percebe-se que entre essas duas probabilidades há uma relação geométrica e que elas podem ser separadas de acordo com suas características que aparecem nos eixos X, Y e Z do gráfico.

O trabalho analisado é importante porque auxilia nos cuidados com a saúde. Assim, mostrando quando há o risco de acontecer um ataque cardíaco. Com isso, ampara tanto os médicos quanto a população. Já que é possível ver se a pessoa tem tendência ou não ao ataque cardíaco, analisando suas dores no peito, colesterol e a frequência cardíaca máxima.