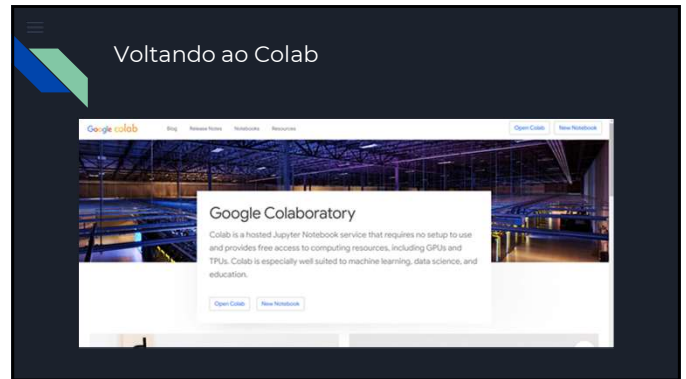




1



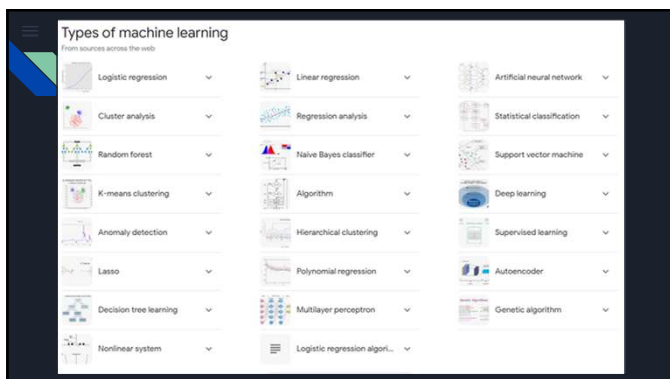
2



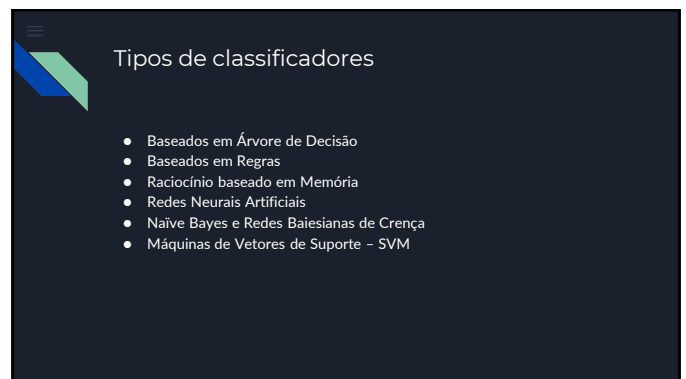
3



4



5



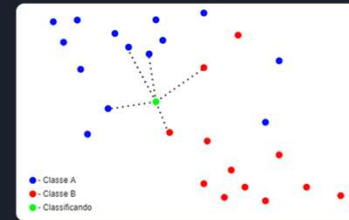
6

## Que vamos utilizar

- Decision Tree - DT
- Random Forest - RF
- Support Vector Machine - SVM
- k-Nearest Neighbors - KNN
- Multilayer Perceptron - MLP

7

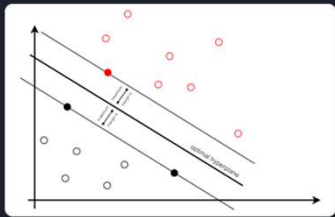
## k-Nearest Neighbors - KNN



"[...] para classificar uma nova instância, ele mede a distância deste exemplo com cada instância presente na base de treino (ALY, 2005)"

8

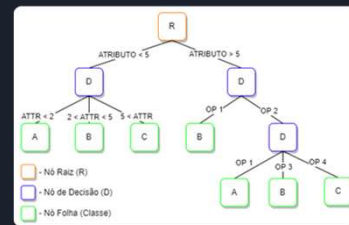
## Support Vector Machine - SVM



seu objetivo "é encontrar o hiperplano de separação com a maior margem; esperando que quanto maior a margem, melhor a generalização do classificador (Hudson, 2006)"

9

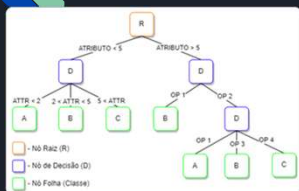
## Decision Tree - DT



"[...] é um classificador que utiliza da lógica da criação de árvores de decisão como forma de obtenção de resultado."

10

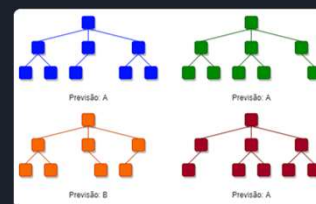
## Random Forest



?

11

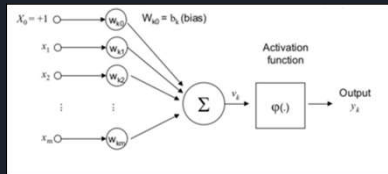
## Random Forest - RF



"Esse comitê de árvores aleatórias vem do entendimento de que, ao combinar vários modelos, os erros de um provavelmente serão compensados por outros, e como resultado, o desempenho geral de previsão do comitê será melhor do que quando usado um único classificador (SAG; ROKACH, 2018)."

12

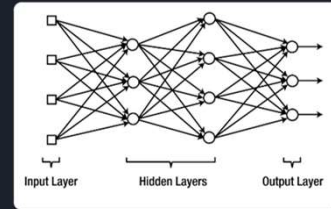
## Multilayer Perceptron - MLP



"Multilayer Perceptrons é um modelo de rede neural inspirado na estrutura e função do cérebro que aprende com os dados e se especializa no reconhecimento de padrões (SOU et al., 2021)"

13

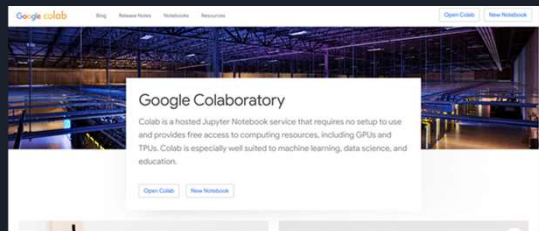
## Multilayer Perceptron - MLP



"Multilayer Perceptrons é um modelo de rede neural inspirado na estrutura e função do cérebro que aprende com os dados e se especializa no reconhecimento de padrões (SOU et al., 2021)"

14

## Voltando ao Colab



15

## Desafio 2:

Procurar outros classificadores do Scikit Learn e aplicar eles na mesma base de dados para ver qual funciona melhor neste cenário.



16

## Machine Learning ~ Métricas

por André L. Jeller Selleti



17

## Acurácia

Classe A: 90 (majoritária)  
Classe B: 10 (minoritária)



Acurácia de 90%

18

## Precision

A	12	0	15	0
B	0	1	2	0
C	0	0	299	0
D	2	0	0	0
	A	B	C	D

$$Pr = \frac{VP}{VP + FP}$$

19

## Recall

A	12	0	15	0
B	0	1	2	0
C	0	0	299	0
D	2	0	0	0
	A	B	C	D

$$Re = \frac{VP}{VP + FN}$$

20

## Precision e Recall = F1-Score

$$F1 = \frac{2 * Pr * Re}{Pr + Re}$$

21

## Macro Average F1-Score

$$Mavg = \frac{\sum_{j=1}^l F1_{,j}}{l}$$

22

## Precision

A	12	0	15	0
B	0	1	2	0
C	0	0	299	0
D	2	0	0	0
	A	B	C	D

$$Pr = \frac{VP}{VP + FP}$$

$$Pr = \frac{12}{12 + (0 + 0 + 2)} \approx 0,86$$

23

## Recall

A	12	0	15	0
B	0	1	2	0
C	0	0	299	0
D	2	0	0	0
	A	B	C	D

$$Re = \frac{VP}{VP + FN}$$

$$Re = \frac{12}{12 + (0 + 15 + 0)} \approx 0,44$$

24

## Precision e Recall = F1-Score

$$F1 = \frac{2 * Pr * Re}{Pr + Re}$$

$$F1 = \frac{2 * 0,86 * 0,44}{0,86 + 0,44} = \frac{0,7568}{1,3} \approx 0,58$$

25

## Macro Average F1-Score

$$Mavg = \frac{\sum_{j=1}^l F1_{ij}}{l}$$

$$Mavg = \frac{(0,58 + 0,50 + 0,97 + 0,0)}{4} \approx 0,51$$

26

## Macro-Avg VS Acurácia

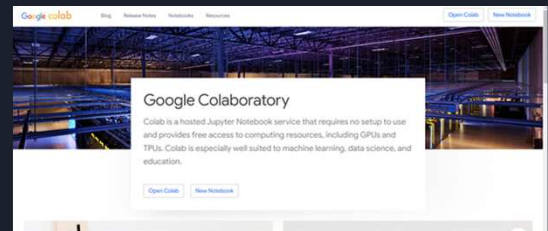
A	12	0	15	0
B	0	1	2	0
C	0	0	299	0
D	2	0	0	0
	A	B	C	D

Macro-Avg = 0,51

Acurácia = 94%

27

## Voltando ao Colab



28

## Desafio 3:

Após testar os Classificadores, agora teste as diferentes métricas, para ver se o melhor classificador continua sendo o mesmo.



29

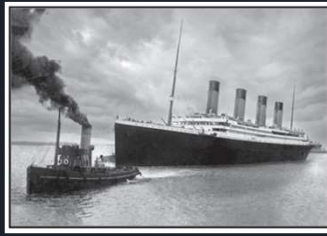
## A se considerar - Problemas com os dados

- Ruídos
  - Distorção da voz de uma pessoa falando.
- Outliers
  - Dados consideravelmente diferentes da maioria, não necessariamente estão errados.
- Dados Faltantes
  - Informação não foi coletada ou não aplicável
- Dados Duplicados
  - Mesma pessoa com vários emails

30

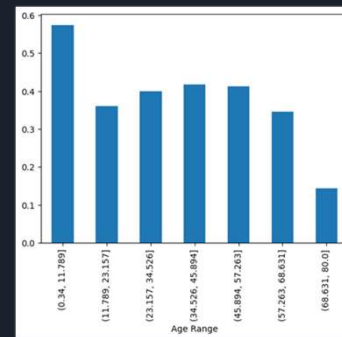
## Desafio 4:

Base Titanic , com dados faltantes  
que vamos arrumar semana que vem

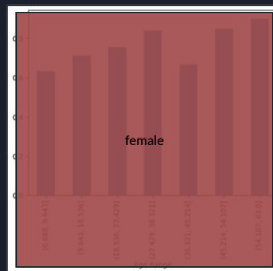
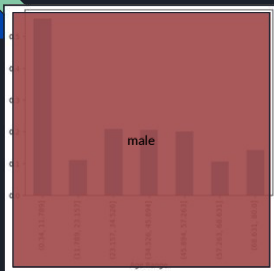


learnenoughpythontobedangerous 377

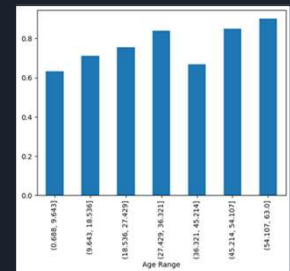
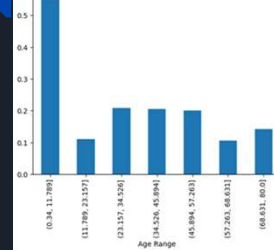
31



32



33



34

Obrigado!

Caso precise tirar alguma dúvida:  
contato@andrejeller.com



35

## Referências e Links

(1) Livro: "Machine Learning"  
ZHOU, Zhi-Hua. Machine learning. Springer Nature, 2021.

Ícones de:  
<https://www.flaticon.com/>

Imagem de capa:  
<https://www.pexels.com/pt-br/foto/3d-abstrato-resumo-abstrair-17483848/>

36