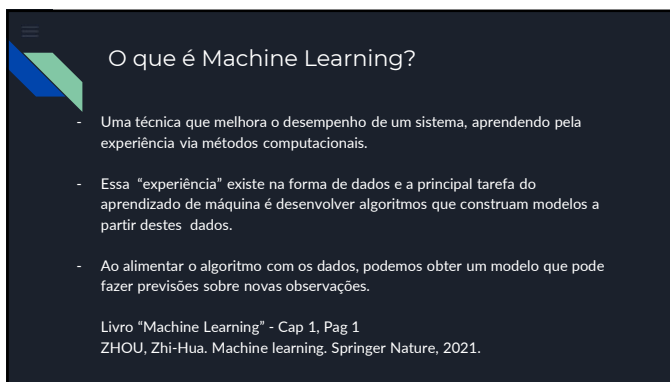




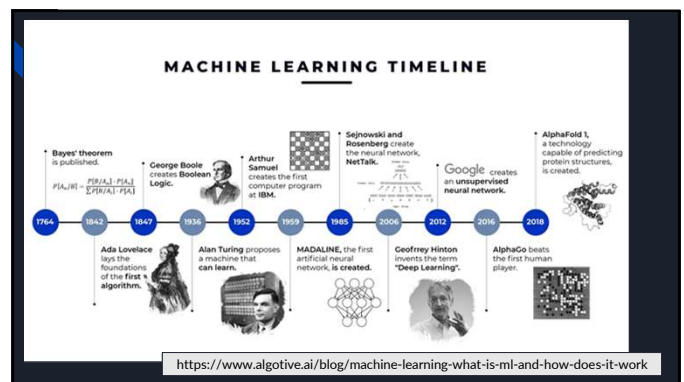
1



2



3



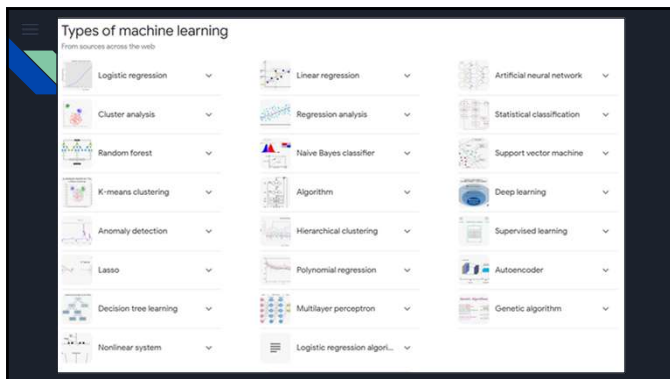
4



5



6



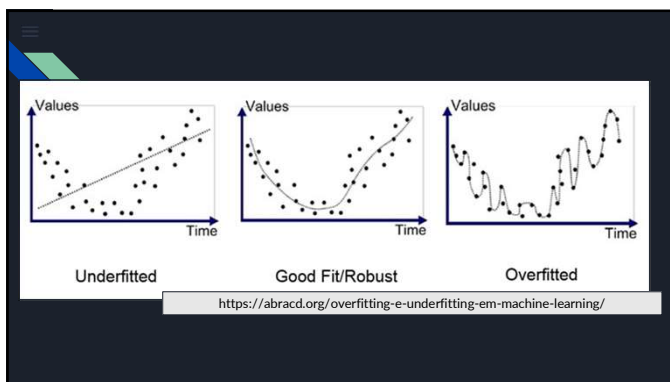
7

Principais conceitos e terminologias:

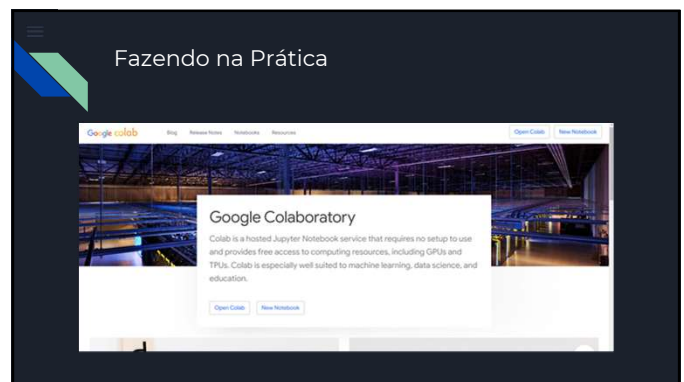
Apesar de vários tipos e de poderem ser agrupados, eles ainda têm suas diferenças. Porém todos seguem algumas coisas em comum como conceitos e estruturas de funcionamento, entre eles:

- Divisão entre Dados de Treinamento, Teste e Validação.
- Features (Características - Entrada) e Labels (Rótulos- Saída).
- Problemas como Overfitting e Underfitting.

8



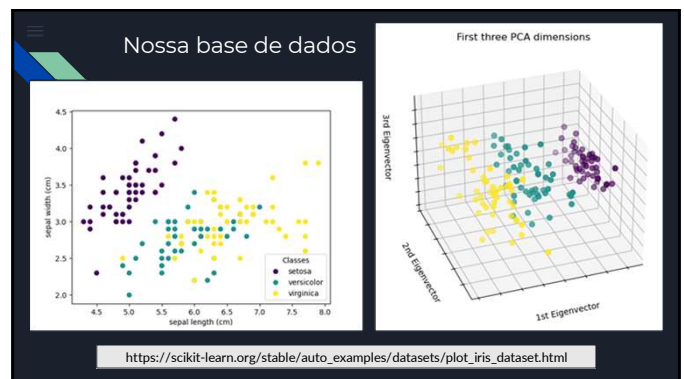
9



10



11



12

Entender o processo básico de treinamento e avaliação

- Coleta e Preparação de Dados.
- Divisão dos Dados em Treinamento e Teste.
- Escolha do Algoritmo de Machine Learning.
- Treinamento do Modelo.
- Teste/Avaliação do Modelo.
- Análise dos Resultados.

13

Desafio 1:

Identificar no código as etapas apresentadas anteriormente

```
1 import numpy as np
2 from matplotlib import pyplot as plt
3
4 from sklearn import tree
5 from sklearn.datasets import load_iris
6 from sklearn.model_selection import train_test_split
7 from sklearn.tree import DecisionTreeClassifier
8 from sklearn.metrics import classification_report
9 from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
10
11 iris = load_iris()
12 X = iris.data
13 y = iris.target
14 classes = iris.target_names.tolist()
15
16 X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
17
18 clf = DecisionTreeClassifier(max_leaf_nodes=3, random_state=0)
19 clf.fit(X_train, y_train)
20 y_pred = clf.predict(X_test)
21
22 print(classification_report(y_test, y_pred))
23
24 cm = confusion_matrix(y_test, y_pred)
25 disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=classes)
26 disp.plot()
27 plt.show()
```

14

Desafio 2:

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



15

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

16

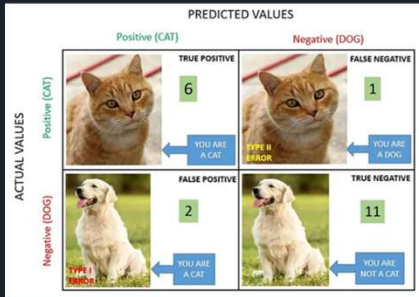
$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

17

		Predicted condition	
		Cancer	Non-cancer
Total	8 + 4 = 12	7	5
	Cancer 8	6	2
	Non-cancer 4	1	3

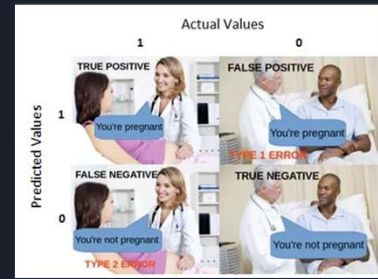
https://en.wikipedia.org/wiki/Confusion_matrix

18



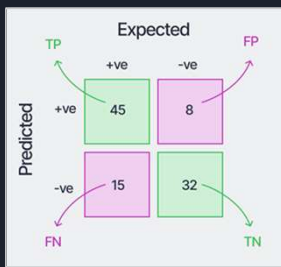
<https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5>

19



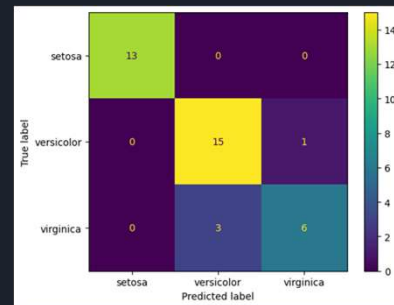
<https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>

20

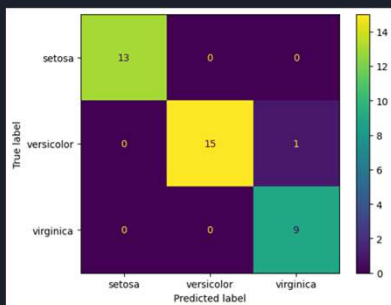


<https://www.v7labs.com/blog/confusion-matrix-guide>

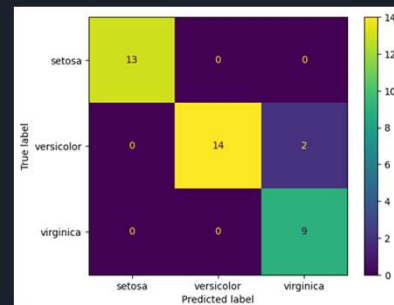
21



22



23



24

		Actual		
		Pregnant	Not	
Predicted	Pregnant	45 TP	55 FP	Type I
	Not	5 FN	395 TN	Type II

<https://towardsdatascience.com/taking-the-confusion-out-of-confusion-matrices-c1ce054b3d3e>

25

Obrigado!

Caso precise tirar alguma dúvida:
contato@andrejeller.com



26

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/>

27