

INTELIGÊNCIA ARTIFICIAL

RELATÓRIO DE CLASSIFICAÇÃO BINÁRIA E MULTICLASSE

Aluna: Lara Luísa Ayrolla Abreu

CLASSIFICAÇÃO BINÁRIA

O notebook criado no Google Colab para fazer a análise da base de dados pode ser acessado através [deste link](#).

A base de dados utilizada possui como tema “**câncer de mama**” e descreve pacientes que possuem um tumor. Ele possui **30 características** para previsão, sendo elas descrições do tumor encontrado, e **569 linhas**, sendo elas diferentes pacientes.

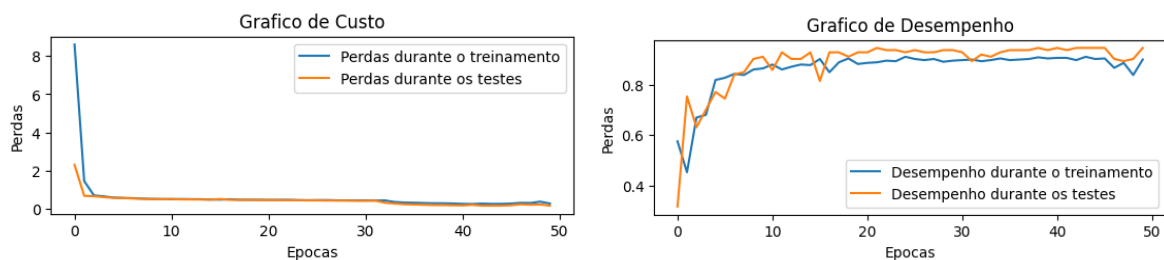
O objetivo da IA é **prever o diagnóstico do tumor** do paciente, para o qual as duas respostas possíveis são:

- **Maligno (M)**
- **Benigno (B)**

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	...
0	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	...
1	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	...
2	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	...
3	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	...
4	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	...
...
564	M	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	...
565	M	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	...
566	M	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	...
567	M	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	...
568	B	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	...

569 rows x 31 columns

Após a execução do código, os resultados e as métricas obtidas foram os seguintes:



Acurácia: 89%

Precisão: 89%

Recall: 92%

F1-score: 91%

AUC ROC: 98.1419%

MATRIZ DE CONFUSÃO		
Valores reais	Valores previstos pelo modelo	
	Benigno	Maligno
Benigno	68	6
Maligno	3	37

RELATÓRIO DE CLASSIFICAÇÃO				
	PRECISÃO	RECALL	F1-SCORE	SUPORTE
B	0.96	0.92	0.94	74
M	0.79	0.93	0.85	40
micro avg	0.89	0.92	0.91	114
macro avg	0.87	0.92	0.89	114
weighted avg	0.90	0.92	0.91	114
samples avg	0.90	0.92	0.91	114

Sendo assim, os resultados obtidos foram satisfatórios, visto que a base era pequena, mas havia grande quantidade de características. Os erros foram baixos tanto para falsos positivos, quanto para falsos negativos, o que é importante em um tópico como o câncer, no qual qualquer erro pode causar a morte do paciente.

CLASSIFICAÇÃO MULTICLASSE

O notebook criado no Google Colab para fazer a análise da base de dados pode ser acessado através [deste link](#).

A base de dados utilizada possui como tema “**tipos de drogas**”. Ele possui **5 características** para previsão, sendo elas descrições da saúde de diferentes pessoas, e **200 linhas**, sendo elas diferentes pessoas.

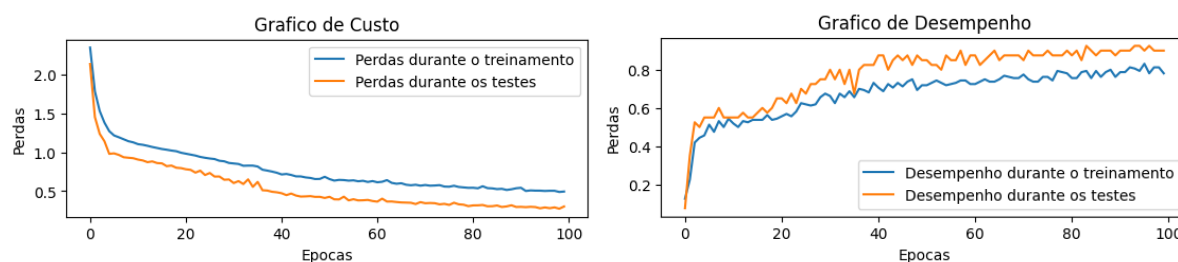
O objetivo da IA é **prever o tipo de droga** utilizado pela pessoa, para o qual há **5 respostas possíveis**, sendo elas:

- drugA
- drugB
- drugC
- drugX
- DrugY

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	DrugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	DrugY
...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

Após a execução do código, os resultados e as métricas obtidas foram os seguintes:



Acurácia: 75%

Precisão: 91%

Recall: 75%

F1-score: 82%

AUC ROC: 97.9569%

MATRIZ DE CONFUSÃO					
Valores reais	Valores previstos pelo modelo				
	drugA	drugB	drugC	drugX	DrugY
drugA	13	0	1	0	1
drugB	0	3	0	0	0
drugC	1	0	3	0	0
drugX	0	0	2	0	1
DrugY	1	0	0	0	14

RELATÓRIO DE CLASSIFICAÇÃO				
	PRECISÃO	RECALL	F1-SCORE	SUPORTE
drugA	1.00	0.87	0.93	15
drugB	1.00	0.33	0.50	3
drugC	0.50	0.50	0.50	4
drugX	0.00	0.00	0.00	3
DrugY	0.93	0.93	0.93	15
micro avg	0.91	0.75	0.82	40
macro avg	0.69	0.53	0.57	40
weighted avg	0.85	0.75	0.79	40
samples avg	0.75	0.75	0.75	40

Sendo assim, os resultados obtidos foram satisfatórios, superando as expectativas. A base era muito pequena, principalmente para uma classificação com 5 classes

possíveis, porém o número de características menor pode ter contribuído para a qualidade do resultado.

Os erros não foram muitos, porém a previsão da substância “drugX” foi de 0%, o que poderia ser um grande problema caso a necessidade de detectar essa droga seja grande.