



CLASSIFICAÇÃO DE TUMORES DE MAMA

Utilizando k-NN

Conheça o Problema

O câncer de mama é uma doença em que células anormais se multiplicam descontroladamente na mama, formando tumores malignos ou benignos. E é o câncer mais comum e letal entre mulheres do mundo todo.

Pensando nesses dados desenvolvemos um modelo de machine learning para classificar os tumores de mama como malignos ou benignos.



DataSet

Breast Cancer Wisconsin (Diagnostic)



- O DataSet Breast Cancer Wisconsin (Diagnostic) conta com **569** casos sendo **212** desses casos malignos e **357** benignos
- Ele é frequentemente utilizado em pesquisas e competições de machine learning devido à sua qualidade e relevância clínica

DataSet

Breast Cancer Wisconsin (Diagnostic)




- O objetivo do DataSet é auxiliar no diagnóstico de tumores de mama, classificando como malignos ou benignos com base em características extraídas de imagens de tecido mamário. O dataset é altamente relevante para a área de saúde, pois pode auxiliar médicos no diagnóstico precoce e preciso de câncer de mama
- O DataSet Utiliza 30 características numéricas, calculadas a partir de imagens digitalizadas de núcleos celulares. Mas neste caso foi utilizada apenas as 10 primeiras características como base

Classificador de Tumores de Mama

Raio (média):	17.46
Textura (média):	21.60
Perímetro (média):	115.37
Área (média):	978.38
Suavidade (média):	0.103
Compacidade (média):	0.277
Concavidade (média):	0.272
Pontos Côncavos (média):	0.124
Simetria (média):	0.202
Dimensão Fractal (média):	0.066

Classificar

Resultado

 O tumor é classificado como: Maligno


OK

Classificador de Tumores de Mama

Raio (média):	12.15
Textura (média):	17.91
Perímetro (média):	78.07
Área (média):	462.80
Suavidade (média):	0.092
Compacidade (média):	0.080
Concavidade (média):	0.047
Pontos Côncavos (média):	0.025
Simetria (média):	0.174
Dimensão Fractal (média):	0.062

Classificar

Resultado

 O tumor é classificado como: Benigno

OK

Acurácia do modelo k-NN: 94.15%

Acurácia média com validação cruzada: 93.85%

Desvio padrão das acurácias: 0.97%

Conclusão

- Em resumo o modelo alcançou uma Eficácia de 94,15%, demonstrando ser muito eficaz na classificação dos tumores.
- Existem diversos projetos que utilizam machine learning para dar assistencia a medicos em diagnosticos do mundo todo, um deles seria o Google DeepMind realizado em parceria com o Serviço Nacional de Saúde do Reino Unido que ja reduziu em 9,4% os falsos negativos.



