

Classificação de Tumores de Mama usando Aprendizado de Máquina

Daniele Carnaúba, Felipe Barros, João Samuel e Leyberson
Assunção

29 de Agosto de 2024

Introdução

- ▶ Câncer de mama é um dos tipos mais comuns de câncer entre mulheres
- ▶ Detecção precoce e precisa é crucial para um tratamento eficaz
- ▶ Objetivo: Classificar tumores mamários como malignos ou benignos usando aprendizado de máquina

Descrição do Conjunto de Dados

- ▶ Conjunto de dados disponível no Kaggle, contendo 569 instâncias
- ▶ Rótulos: Maligno (M) ou Benigno (B)
- ▶ 30 características numéricas que descrevem os núcleos celulares

Rótulos e Variáveis

- ▶ **Rótulo Alvo:** Diagnóstico (M ou B)
- ▶ **Variáveis:**
 - ▶ **Médias:** radius_mean, texture_mean, perimeter_mean, area_mean, smoothness_mean, compactness_mean, concavity_mean, concave points_mean, symmetry_mean, fractal_dimension_mean
 - ▶ **Piores Valores:** radius_worst, texture_worst, perimeter_worst, area_worst, smoothness_worst, compactness_worst, concavity_worst, concave points_worst, symmetry_worst, fractal_dimension_worst
 - ▶ **Erros Padrão:** radius_se, texture_se, perimeter_se, area_se, smoothness_se, compactness_se, concavity_se, concave points_se, symmetry_se, fractal_dimension_se

Questão de Pesquisa

Questão Principal

"Podemos classificar com precisão os tumores mamários como malignos ou benignos com base nas características extraídas de imagens digitalizadas?"

Metodologia

- ▶ Preparação e exploração de dados
- ▶ Seleção e treinamento de modelos
- ▶ Avaliação do modelo
- ▶ Tratamento de dados ausentes

Modelos Considerados

- ▶ Regressão Logística, Árvores de Decisão, Random Forest
- ▶ Máquinas de Vetores de Suporte (SVM), k-Nearest Neighbors (k-NN)
- ▶ Naive Bayes, Redes Neurais Artificiais (ANN)
- ▶ Gradient Boosting Machines (GBM), XGBoost, LightGBM, CatBoost
- ▶ Métodos de Conjunto (Stacking, Bagging)

Tratamento de Dados Ausentes

- ▶ Remover colunas irrelevantes (por exemplo, *Unnamed: 32*)
- ▶ Imputação de valores ausentes: média, mediana, moda
- ▶ Avaliação do impacto: Comparar desempenho antes e depois da imputação

Conclusão

- ▶ Desenvolver um modelo robusto e preciso para a classificação de tumores mamários.
- ▶ Avaliar o sucesso do modelo com base na capacidade de generalização e em métricas de desempenho.
- ▶ Apoiar a tomada de decisões clínicas no diagnóstico do câncer de mama.