

Grupo DataStudent

Douglas Matos Gomes, RA: 170013

Matheus Rotta Alves, RA: 184403

Resumo:

Dado um conjunto de dados coletados sobre diferentes características de núcleos de células de tumores, o problema a ser resolvido é prever se o tumor é benigno ou maligno.

As características incluem raio, textura, perímetro, área, suavidade, compactidade, concavidade, pontos côncavos, simetria e dimensão fractal.

O raio é a média das distâncias do centro aos pontos no perímetro, a textura é o desvio padrão dos valores em tons de cinza, a suavidade é a variação local dos comprimentos do raio, a compactidade é o perímetro ao quadrado sobre a área menos um e a concavidade é a severidade das porções côncavas ao longo do contorno.

Nossa proposta é fazer diferentes seleções de subconjuntos dos dados disponíveis e, para cada seleção, avaliar a capacidade desses atributos de prever se o tumor é maligno ou benigno.

Ademais, separamos as diferentes características em dois grupos principais, numéricas e visuais.

Lista de Requisitos:

- Propiciar subdivisão dos dados em subconjuntos (Seleção);
- Associar um valor de acurácia com cada subconjunto;
- Relacionar as características de cada tumor, determinando se é visual ou puramente numérica;
- Fornecer um diagnóstico associado a cada tumor;

Fonte dos dados:

A fonte da qual foram retirados os dados é o repositório de Machine Learning da escola Bren de Informação e Ciência da Computação da Universidade da Califórnia, em Irvine.

([https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)))

Diagrama UML:

