

**Atividade A3**

**Nome dos Alunos:**

Ana Carolina Vasque de Oliveira

Eduarda Aparecida Guedes

Laion De Souza Medeiros

Thales Cardoso Carvalho Novaes

Julio Caporicci

**Base de Dados: Câncer de mama (Diagnóstico)**

Dentro de uma Base de Dados relevante de analise de imagem digitalizada de punção aspirativa com agulha fina de uma massa mamária, poderemos conter informações sobre diagnóstico (maligno ou benígno), raio do nódulo, textura, perímetro, área, entre outros. Foi considerado, para este estudo, apenas amostras onde houveram um tumor presente.

Um exemplo com algumas variáveis e tipos de dados:

* Diagnóstico "M" ou "B", maligno ou benigno (texto).
* Raio: (número).
* Textura: Variável desvio padrão de escala cinza (número).
* Perímetro: Tamanho médio do tumor central (número).
* Porções Côncavas: Variável decimal (número).

Agora, vamos abordar a aplicação dos métodos:

Árvores de Decisão:

As árvores de decisão podem ser usadas para prever se uma paciente tem câncer de mama. A divisão da base de dados em conjuntos de treinamento e teste, compreendendo quais atributos do câncer especifíca os diagnósticos, para treinar uma árvore de decisão com base nas variáveis e, em seguida, usaria a árvore treinada para fazer previsões.

Redes Neurais:

Redes neurais podem ser usadas para tarefas mais complexas, como atribuições que não estão previstas nos valores. Mesmo que haja um nódulo benigno, é possível, através da rede, categorizar possíveis novos tumores, partindo de outras bases e pacientes, como padrões que geraram benignos e também, posteriormente, geraram malignos.

KNN (k-Nearest Neighbors)

KNN podem auxiliar identificando vizinhos próximos, através de uma variável de escolha, para descobrir se elas são sensíveis ao teste e, por seguinte, definir o atributo de comparação e compensação nos casos de tumor maligno.

**Diagnóstico:**

Variável Target: Valor do diagnóstico do tipo do maligno (M) ou benigno (B)

Justificação: Considerando a existência de um tumor, a principal característica que torna-o ameaçador para quem o tem, é sua classificação de malígno, pois o benigno, apesar de também requerer tratamento, não invade outros tecidos, nem espalha-se, como o maligno, justificando como principal, a variável para determinar o processo.

**Hipótese:**

Para classificar e dimensionar o tumor, vamos realizar as seguintes questões para compreender a proporção e gravidade:  
Qual a área do tumor?

Qual o perímetro do tumor?

Qual a textura do tumor?

Quais as porções côncavas do tumor?

Quais as médias das porções côncavas do tumor?

A área do tumor, junto com o perímetro e as porções côncavas, nos permitirão detalhar em qual processo ocorre o tumor, se está avançado ou não. A textura, também é um indicativo, porém, podemos isolá-la como um atributo de indicativo de dano.

**Método KNN**

Através da utilização do método KNN, é notável a distribuição entre os malignos e benignos, onde os malignos estão concetrados nas faixas iniciais do eixo X e Y, criando margem mais para o eixo X. Podemos traçar alguns padrões a partir da imagem gerada, visto que a distribuição dos pontos no gráfico está bem concentrada e bem dividia, onde alguns pontos azuis se encontram com os vermelhos, mas em sua maioria, há um afastamento entre eles, nos permitindo visualizar os atributos que classificam-os como malignos.

