

# Implementação de classificador para determinar tipo de tratamento em pacientes com câncer de mama através do algoritmo de Bayes ingênuo

Alansidney da S. Júnior, Matheus F. B. Souza, e Pedro H. S. de Paula, Juan Pablo N. T. Ferreira  
Estudantes de graduação em Engenharia da Computação  
na Universidade Federal de Pernambuco

01 de Abril de 2021

## I. OBJETIVOS

Este projeto tem como objetivo o estudo, treinamento, construção, e validação de um modelo classificador utilizando o algoritmo de Bayes ingênuo que seja capaz de receber informações relacionadas a pacientes com câncer de mama e concluir se o mesmo pode ser tratado por meio de radioterapia. Para atingir tal objetivo será utilizado como linguagem de programação Python e como ferramentas para implementação a biblioteca scikit-learn e a plataforma Google Colab.

## II. JUSTIFICATIVA

O Instituto Nacional de Câncer (INCA) estima que para cada ano do triênio 2020/2022, sejam diagnosticados no Brasil 66.280 novos casos de câncer de mama, com um risco estimado de 61,61 casos a cada 100 mil mulheres. O tratamento varia de acordo com o estadiamento da doença, suas características biológicas, bem como das condições da paciente (idade, status menopausal, comorbidades e preferências). Quando a doença é diagnosticada no início, o tratamento tem maior potencial curativo.

O tratamento radioterápico utiliza radiações ionizantes para destruir ou inibir o crescimento das células anormais que formam um tumor. Nem todas as mulheres com câncer de mama têm indicação de radioterapia. Além disso, a radioterapia pode apresentar efeitos colaterais, como seroma, dor mamária, infecção e fratura de costelas em casos raros. Por conta disso, submeter uma paciente com baixa probabilidade de resposta à esse tipo de tratamento pode ser perigoso, além de atrasar a aplicação de outros tratamentos que possam ter mais eficácia para a paciente em questão. Desta forma, torna-se valido a utilização de um classificador para eficientemente direcionar as pacientes ao tratamento.

## III. METODOLOGIA

A metodologia utilizada para a implementação do programa consistirá de uma execução sequencial de várias etapas, dentre elas, a análise e tratamento da base de dados (Breast Cancer Data Set), a separação de amostras usadas tanto na fase de treinamento quanto na fase de teste do algoritmo, o uso de métodos da biblioteca scikit-learn para manipular os dados previamente tratados e fornecer os retornos necessários para a implementação do modelo, e por fim, ser capaz de prever com certa acurácia, a partir das features de entrada, se um câncer pode ou não ser tratado por radioterapia.

Dados da base utilizada:

Variável	Descrição	Valores
age	idade do indivíduo	inteiro
menopause	Indica se o paciente é pré ou pós menopausa no momento do diagnóstico	lt40, ge40, premeno
tumor-size	O maior diâmetro (em mm) do tumor excisado	inteiro
inv-nodes	o número de linfonodos axilares que contêm mama metastático	inteiro
node-caps	Mostra se o câncer se desenvolve dentro do linfonodo	Sim ou não
deg-malig	O grau de malignidade do tumor	inteiro. 1 2 ou 3
breast	O câncer de mama que pode ocorrer em qualquer mama	Left, Right
breast-quad	Quadrante da mama que pode ser dividida em quatro quadrantes	Quadrante
irradiat	Tratável por radioterapia	Sim, Não
class	Demonstra se os dados recorrem para um câncer de mama benigno ou maligno	Sim, Não

## IV. CRONOGRAMA DE ATIVIDADES

Início	Fim	Atividade
27/03/2021	28/03/2021	Pesquisa e escolha da Base de dados
29/03/2021	29/03/2021	Reunião e elaboração de brainstorm para decisão do tema e título do projeto
31/03/2021	01/04/2021	Planejamento e construção da Proposta de projeto
02/04/2021	05/04/2021	Análise da base de dados e estudo das ferramentas scikit-learn e Google Colab
05/04/2021	08/04/2021	Separação dos dados em grupo para treinamento e grupo para teste
09/04/2021	10/04/2021	Construção do modelo proposto usando o classificador de Bayes ingênuo
11/04/2021	15/04/2021	Fase de testes
15/04/2021	19/04/2021	Análise de Elaboração do relatório de projeto
20/04/2021	22/04/2021	Estudo crítico dos resultados obtidos e gravação da apresentação

## V. REFERENCIAS

INSTITUTO NACIONAL DE CÂNCER. Tipos de câncer: Câncer de Mama  
<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>

INSTITUTO NACIONAL DE CÂNCER. Conceito e Magnitude do câncer de mama  
<https://www.inca.gov.br/controlado-cancer-de-mama/conceito-e-magnitude>

INSTITUTO NACIONAL DE CÂNCER. Tratamento para o câncer de mama  
<https://www.inca.gov.br/controlado-cancer-de-mama/acoes-de-controlado/tratamento>

Faculdade Guairacá. Felipe C. Soczek, Regiane Orlovski. Mineração de Dados: Conceitos e aplicação de algoritmos em uma Base de Dados na área da saúde

UCI MACHINE LEARNING REPOSITORY. Breast Cancer Data Set  
<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>

Machine Learning in Python. Scikit-learn  
<https://scikit-learn.org/stable/>

Thiago G Santos. Google Colab: o que é e como usar?  
<https://www.alura.com.br/artigos/google-colab-o-que-e-e-como-usar>