

# COMPARAÇÃO DE MODELOS DE ML PARA DIAGNÓSTICO DE CÂNCER DE MAMA

## BREAST CANCER DIAGNOSIS ML MODEL COMPARISON

Queiroz V. K. M<sup>1</sup> – Instituto Federal do estado do Ceará (IFCE)

### RESUMO:

Este trabalho realiza uma análise comparativa entre três modelos de aprendizado de máquina, a Regressão Logística, Máquina de Vetores de Suporte (SVM) e Árvore de Decisão, aplicados ao diagnóstico de câncer de mama. Usando o Breast Cancer Wisconsin Dataset, os modelos foram avaliados por meio de métricas como acurácia, precisão, recall e F1-score. Os resultados indicam que o modelo de Regressão Logística teve um desempenho levemente melhor, com maior precisão e melhor equilíbrio entre os resultados positivos e negativos. Este estudo sugere que o modelo de Regressão Logística é uma opção eficaz para auxiliar na detecção do câncer de mama.

**Palavras-chave:** Câncer de mama. SVM. Regressão logística. Árvore de Decisão.

### ABSTRACT:

This work performs a comparative analysis between three machine learning models, Logistic Regression, Support Vector Machine (SVM) and Decision Tree, applied to the diagnosis of breast cancer. Using the Breast Cancer Wisconsin Dataset, models were evaluated using metrics such as accuracy, precision, recall and F1-score. The results indicate that the Logistic Regression model performed slightly better, with greater precision and a better balance between positive and negative results. This study suggests that the Logistic Regression model is an effective option to assist in the detection of breast cancer.

**Keywords:** Breast Cancer. SVM. Logistic Regression. Decision Tree.

## 1. INTRODUÇÃO

O câncer de mama é uma das principais causas de morte entre as mulheres no mundo, e o diagnóstico precoce é crucial para aumentar as chances de sucesso no tratamento. Nos últimos anos, modelos de aprendizado de máquina (ML) têm sido amplamente utilizados para ajudar na detecção automática do câncer de mama a partir de dados clínicos e imagens. Este trabalho tem como objetivo comparar três dos modelos mais utilizados – a Regressão

---

<sup>1</sup> Graduanda em Ciência da Computação - Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE Campus Maracanaú. Técnico em Automação Industrial - Escola Estadual de Educação Profissional Raimundo Célio Rodrigues (2019).

Logística, Máquina de Vetores de Suporte (SVM) e Árvore de Decisão – para determinar qual apresenta melhor desempenho na detecção do câncer de mama.

## **2 FUNDAMENTAÇÃO TEÓRICA**

O diagnóstico de câncer de mama pode ser modelado como um problema de classificação binária, onde o objetivo é classificar as amostras como 'benignas' ou 'malignas' com base em atributos extraídos de exames médicos, como mamografias. O conjunto de dados usado neste estudo é o Breast Cancer Wisconsin Dataset Diagnostic, que contém 30 atributos relacionados às características das células mamárias e 569 instâncias.

## **3 MATERIAL E MÉTODO**

A metodologia envolve a aplicação dos modelos de Regressão Logística, SVM e Árvore de Decisão no conjunto de dados do câncer de mama. O processo de validação cruzada com 10 folds será utilizado para garantir que os modelos sejam avaliados de maneira consistente e robusta. As métricas utilizadas para comparação são acurácia, precisão, recall e F1-score. Para obter os melhores parâmetros e modelos será utilizado também uma implementação personalizada do GRID SEARCH CV.

## **4 RESULTADOS E DISCUSSÃO**

Os experimentos foram realizados com os modelos implementados na biblioteca scikit-learn, com uma divisão de 80% dos dados para treinamento e 20% para teste. Os resultados mostraram que o Regressão Logística obteve uma acurácia de 99%, enquanto o SVM 96% e a Árvore alcançou 94%. A Tabela 1 mostra as métricas detalhadas para ambos os modelos. A Figura 1, Figura 2 e a Figura 3 ilustram as curvas ROC, demonstrando que a Regressão Logística apresenta levemente uma maior área sob a curva (AUC).

Tabela 1: métricas obtidas.

Modelo/métricas	Acurácia	Precisão	Revocação	F1
Regressão logística	99%	100%	97%	98%
SVM	96%	95%	95%	95%
Árvore de Decisão	94%	95%	90%	94%

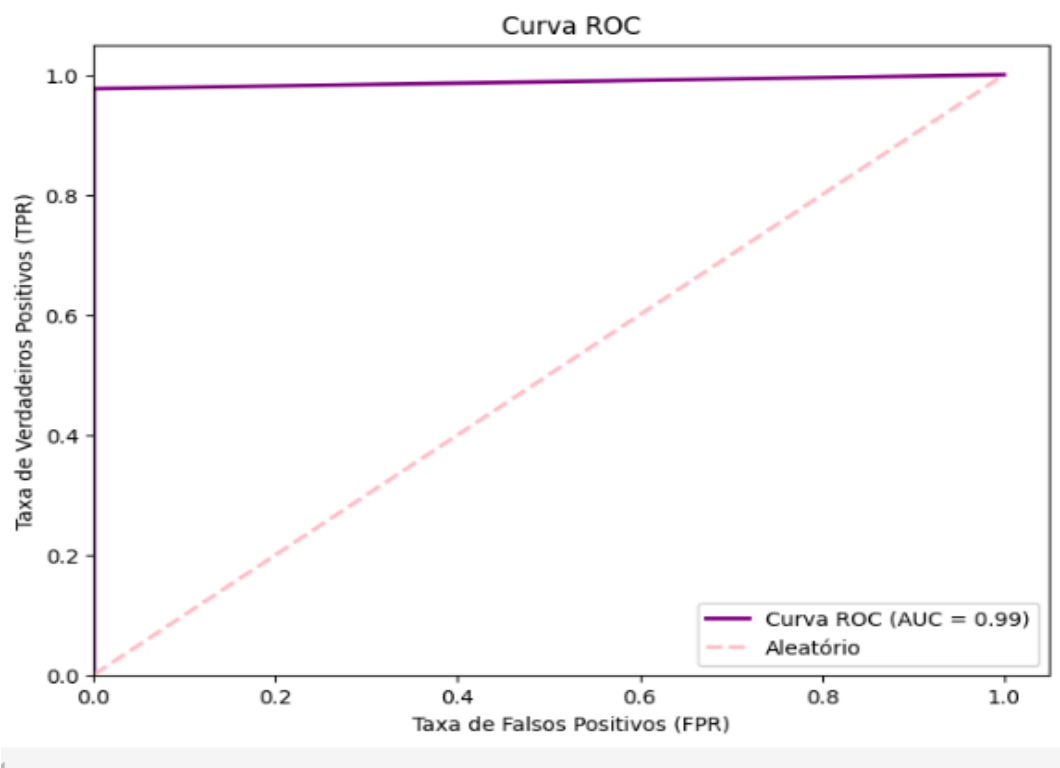


Figura 1: Curva ROC Regressão Logística.

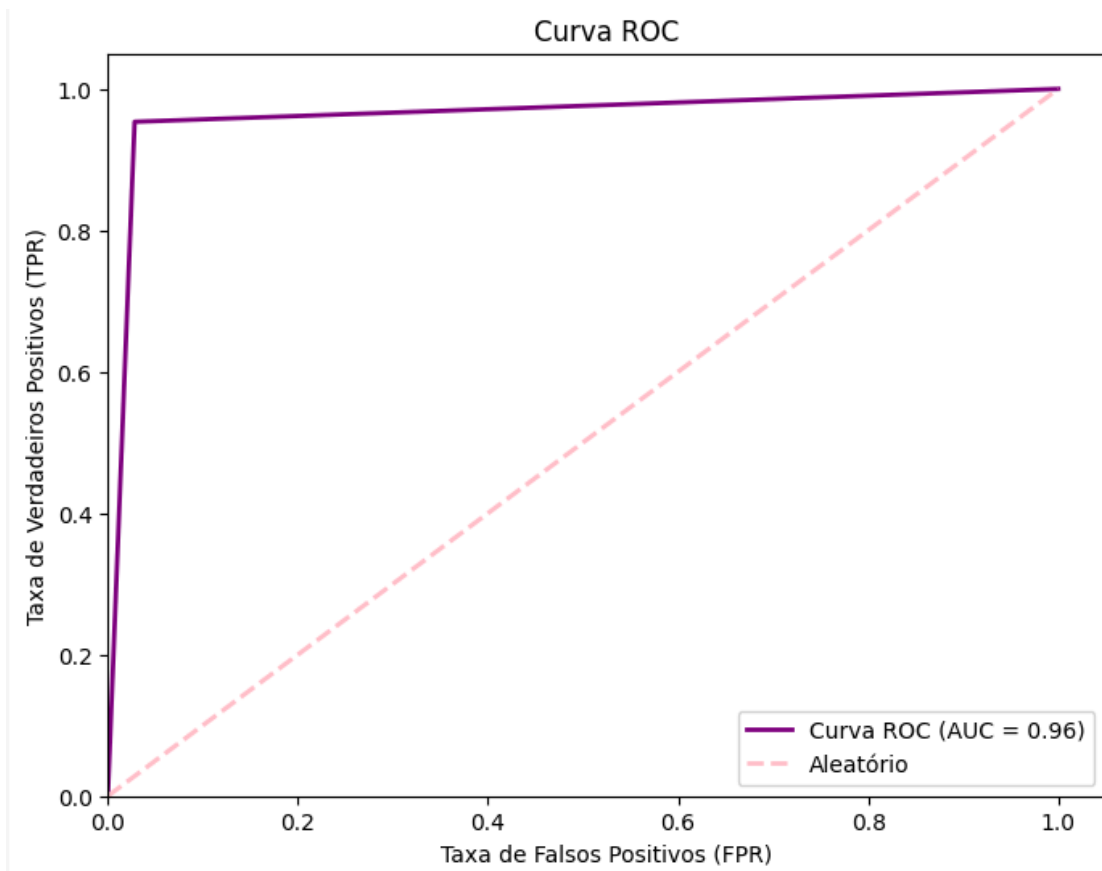


Figura 2: Curva ROC SVM.

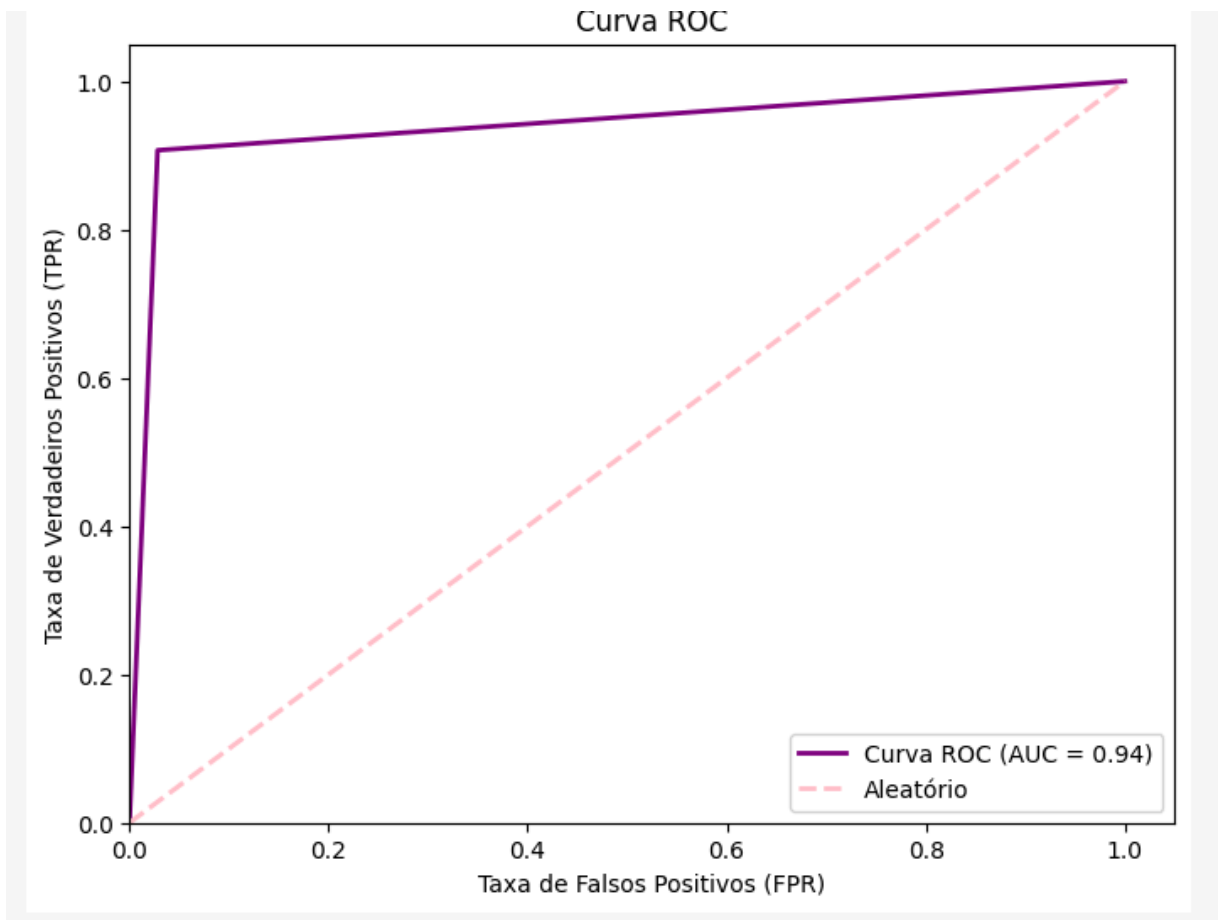


Figura 3: Curva ROC Árvore de decisão.

## 5 CONSIDERAÇÕES FINAIS

Com base nos resultados, pode-se concluir que o modelo de Regressão Logística é ligeiramente superior no diagnóstico de câncer de mama em comparação ao SVM e Árvore de decisão, para casos clínicos é necessário avaliar a revocação e RL se saiu bem.