

# OncoScope: Projeto Tech Challenge 1

**Equipe:** Davi Lopes - RM366331

**Projeto no Github:** <https://github.com/dgllpnv/OncoScope>

**Vídeo do projeto:** <https://youtu.be/I0TwASuSPgg>

## 1. Objetivo do projeto

O OncoScope tem como propósito ser uma solução de diagnóstico assistido para câncer de mama que una técnicas de Machine Learning e um canal de apoio emocional via chatbot.

### 1.1 Objetivos específicos:

- **Precisão Diagnóstica:** Treinar modelos preditivos capazes de classificar lesões benignas e malignas a partir de dados estruturados do dataset Wisconsin Breast Cancer.
- **Integração com suporte emocional:** Implementar um chatbot que, após o diagnóstico, forneça mensagens de orientação e incentivo, lembrando ao usuário a importância de buscar avaliação médica e cuidando do apoio psicológico.
- **Documentação e Disseminação:** Fornecer um relatório técnico claro e um vídeo demonstrativo que facilitem a compreensão do projeto e reprodução.

## 2. Dados e modelos

O OncoScope tem como propósito ser uma solução de diagnóstico assistido para câncer de mama que una técnicas de Machine Learning e um canal de apoio emocional via chatbot.

### 2.1 Dataset

Foi utilizado o **Wisconsin Breast Cancer Dataset**

Total de amostras: 569

Classes: Benigno (357), Maligno (212)

Formato: 30 variáveis numéricas derivadas de imagens digitalizadas de punções aspirativas da mama (FNA).

Seleção de variáveis: Para otimizar desempenho e reduzir o tempo de preenchimento manual, foram escolhidas 10 features mais relevantes segundo análise de correlação e importância de atributos:

worst\_radius – Maior raio  
worst\_texture – Textura  
worst\_perimeter – Perímetro  
worst\_area – Área  
worst\_smoothness – Suavidade  
mean\_radius – Raio médio  
mean\_texture – Textura média  
mean\_perimeter – Perímetro médio  
mean\_area – Área média  
mean\_concave\_points – Pontos côncavos médios

## **2.2 Modelos**

Pré-processamento: Normalização dos dados e divisão em treino/validação/teste.

Modelos testados:

- Logistic Regression
- Random Forest

Modelo escolhido: Logistic Regression, que apresentou o melhor equilíbrio entre precisão, recall e interpretabilidade, sendo salvo como best\_model.pkl para uso no backend.

## **3. Métricas e Interpretação**

Acurácia: Proporção de previsões corretas.

Precisão: Proporção de casos classificados como malignos que realmente são malignos (baixa taxa de falsos positivos).

Recall: Capacidade do modelo de identificar corretamente todos os casos malignos (baixa taxa de falsos negativos).

F1-Score: Média harmônica entre precisão e recall, útil quando há desbalanceamento de classes.

## **4. Fluxo do Sistema e Integração OpenAI**

O sistema é composto por duas camadas principais:

### **4.1 Backend – FastAPI**

- Recebe dados do exame (As 10 variáveis mais relevantes no diagnóstico).
- Carrega o modelo treinado (best\_model.pkl).
- Realiza a previsão (Benigno ou Maligno) e retorna a probabilidade associada.
- Integração com API da OpenAI
- Após previsão, envia o diagnóstico para um prompt customizado.
- Recebe resposta do chatbot com mensagem de acolhimento e orientação.
- Suporta modo multi-turno com conversas persistentes, permitindo interação contínua com o usuário.

### **4.2 Frontend – Angular + Angular Material**

- Formulário responsivo para entrada das 10 variáveis.
- Botões para preenchimento rápido de exemplos benignos e malignos.
- Exibição do resultado e barra de confiança.
- Caixa de chat acoplada para interação com o assistente virtual.

### **Fluxo resumido:**

- Usuário preenche dados > envia para backend.
- Backend prediz resultado > chama OpenAI > retorna resposta inicial.
- Usuário pode continuar conversa com o chatbot.

## **5. Discussão Ética e Limitações**

### **5.1 Aspectos Éticos**

Não substituição do diagnóstico médico: O sistema deixa claro que é apenas apoio à decisão e não substitui consulta com profissional de saúde.

Transparência: Modelo treinado com dataset público, código aberto e documentação disponível.

Privacidade: Nenhum dado pessoal é armazenado, a entrada é processada em tempo real.

## **5.2 Limitações**

Base de dados limitada: Dataset pequeno e homogêneo, podendo não refletir toda a diversidade de casos clínicos reais.

Variabilidade técnica: Modelos de ML podem ter desempenho inferior em populações diferentes da usada no treino.

Dependência de conexão: Chatbot via OpenAI exige internet; fallback offline fornece mensagens mais genéricas e não usa o chatbot de apoio emocional.

Falsos negativos: Apesar de priorizar recall, há risco residual de não identificar casos malignos.

Inserção manual: A inserção dos dados manualmente pode-se tornar cansativa e confusa para o usuário.

## **6. Conclusão**

O OncoScope é uma solução que pode ajudar, tanto com um pré-diagnóstico quanto com um apoio emocional no momento da notícia e indicar a procura de profissionais de saúde.

O projeto alcançou boa performance no teste, possui arquitetura modular (backend + frontend), integração com OpenAI e está documentado para reprodutibilidade.