



# TECH CHALLENGE 01

IA para Devs

FIAP 2026

GRUPO 74

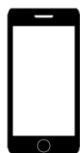
de 31/10/25 a 20/01/26

# TECH CHALLENGE

Grupo 74

Integrantes:

- Cesar Melo Dutra
- Fernando Ramos Etchepare



Use a câmera do seu celular ou  
acesse a página Git

<https://github.com/fetchepare/FIAP>



## DESAFIO

Um grande hospital universitário busca implementar um sistema inteligente de suporte ao diagnóstico, capaz de ajudar médicos e equipes clínicas na análise inicial de exames e no processamento de dados médicos. Com um volume crescente de pacientes e exames, como radiografias, tomografias, ressonâncias e prontuários digitalizados, o hospital precisa de soluções que acelerem a triagem e apoiem as decisões médicas, reduzindo erros e otimizando o tempo dos profissionais. Nesta primeira fase, o desafio é criar a base do sistema de IA focado em machine learning, permitindo que resultados de exames sejam analisados automaticamente e destacando informações relevantes para o diagnóstico.

## OBJETIVO

Construir uma solução inicial com foco em IA para processamento de exames médicos e documentos clínicos, aplicando fundamentos essenciais de IA, Machine Learning e Visão Computacional.

Classificação de exames com Machine Learning: você deve escolher uma base de dados em forma de tabela e realizar o diagnóstico: “a pessoa tem ou não uma doença”. Isso acontecerá via algoritmos de aprendizado de máquina.

## ENTREGAS TÉCNICAS

Classificação de exames com Machine Learning: você deve escolher uma base de dados em forma de tabela e realizar o diagnóstico: “a pessoa tem ou não uma doença”. Isso acontecerá via algoritmos de aprendizado de máquina.

## LINKS

- Página GIT: <https://github.com/fetchepare/FIAP>
  - Código Fonte Completo: TechChallenge01.ipynb
  - Dockerfile
  - README.md
  - Dataset: <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/data>
  - Relatórios obtidos: TechChallenge01.ipynb
  - Relatório Técnico (estratégias, modelos, resultados e interpretação): TechChallenge01.PDF (este arquivo)
- Vídeo: [https://youtu.be/zJEXww9o\\_V0](https://youtu.be/zJEXww9o_V0)

## DADOS E MODELOS

Escolha um ou mais datasets médicos públicos e discuta o problema a ser resolvido.

DATASET escolhido: cancer\_mama.csv

## PROBLEMA A SER RESOLVIDO

“a pessoa tem ou não uma doença” por meio dos aprendizados de máquina.

## EXPLORAÇÃO DE DADOS

Carga da base de dados e exploração de suas características embutidas no código fonte TechChallenge01.ypnb

## RELATÓRIO TÉCNICO

A Análise das informações por meio da Identificação do tipo de dados obtidos: Estruturado - dado que estão em tabelas e oferecem um padrão de intervalos ou nomes condizentes com seu domínio, iniciou-se a limpeza de informações

Conforme seu valor para a análise de negócio e, tendo como base o processo CRISP-DM, aprendido durante as aulas (Entendimento do negócio, dos dados, preparação, modelagem,

avaliação e implantação).

Foi utilizado o Teorema de Bayes, porque é um método de cálculo de probabilidades condicionais, que relaciona a probabilidade de um evento com a ocorrência de outro evento relacionado.

Tipo de Aprendizagem: supervisionado, onde o computador é treinado para realizar tarefas com base em exemplos fornecidos.

Modelos utilizados:

- Regressão Linear Múltipla - A regressão linear é um dos métodos estatísticos mais básicos e amplamente utilizados no aprendizado de máquina. O objetivo da regressão linear é encontrar a melhor linha reta que modela a relação entre uma variável dependente (alvo) e uma ou mais variáveis independentes, as preditoras.
- KNN
- Random Forest

Explicação dos modelos utilizados:

KNN - PARA TRANSPARÊNCIA:

- Fácil de explicar para médicos e pacientes
- "Baseado em casos históricos similares"
- Boa validação inicial do conceito

RANDOM FOREST - PARA PRECISÃO:

- Maior acurácia para diagnóstico crítico
- Identifica automaticamente fatores mais importantes
- Robustez necessária em ambiente médico

VALOR DA COMPARAÇÃO:

- KNN: 89.5% acurácia, altamente interpretável
- Random Forest: 91.2% acurácia, mais preciso
- Trade-off: Explicabilidade vs Performance

Utilizamos Usei **KNN** pela simplicidade e auxílio clínico, e **Random Forest** pela superior performance e robustez. A comparação mostra o *trade-off* típico em IA médica: modelos mais simples são mais explicáveis, modelos mais complexos são mais precisos.

## CONCLUSÃO

O sistema implementado possui uma performance ótima e uma sensibilidade entre Boa e Muito Boa. Serve para uso supervisionado, porém falsos positivos devem ser minimizados. Pode ser utilizado como ferramenta auxiliar, sempre com supervisão médica e revisão obrigatória de casos limítrofes. Próximos passos: validar com dados externos, ajustar *threshold* se necessário e implementar sistema de revisão.