



Universidade Presbiteriana  
**Mackenzie**

**PÓS-GRADUAÇÃO EM ENGENHARIA DE DADOS**  
**ATIVIDADE AVALIATIVA – DATA SCIENCE**

**KARINA GOMES DIAS – 10734429**

**LETÍCIA HOMEM DE MELO SANCHEZ – 10733811**

**MARINA SOARES DE SOUZA – 10106224**

**PAOLA YÊDA AUDE GAUDIELLO – 10739716**

**SAULO SIMOES RIBEIRO GONÇALVES – 10738083**

**THOMAS RAPHAEL DE OLIVEIRA – 10734116**

**SÃO PAULO**

**1º TRIMESTRE, 2025**

## Sumário

<b>1. Resumo Executivo.....</b>	<b>3</b>
<b>2. Introdução.....</b>	<b>3</b>
Contextualização do Problema:.....	3
Objetivos técnicos:.....	3
Objetivos de negócio:.....	4
Metodologia Proposta:.....	4
<b>3. Análise Exploratória dos Dados.....</b>	<b>4</b>
Descrição das Variáveis.....	4
Análises Estatísticas e Distribuições.....	6
Correlações Entre Atributos.....	6
Visualizações Gráficas e Principais Insights.....	6
Principais insights:.....	8
<b>4. Engenharia de Atributos.....</b>	<b>8</b>
Features Criadas com Justificativas.....	8
Features Criadas e Justificativas.....	8
Análise do Impacto Geral das Features.....	10
<b>5. Modelagem.....</b>	<b>11</b>
Tratamento do Desbalanceamento.....	11
Validação e Tuning de Hiperparâmetros.....	11
<b>6. Avaliação e Interpretação.....</b>	<b>12</b>
<b>7. Implementação e Próximos Passos.....</b>	<b>13</b>
Uso do Modelo em Produção.....	13
Plano de Monitoramento e Re-Treinamento.....	13
Sugestões de Melhorias Futuras.....	13
<b>8. Conclusão.....</b>	<b>14</b>
<b>9. Anexos e Repositório.....</b>	<b>14</b>

## **1. Resumo Executivo**

As doenças cardiovasculares representam a principal causa de morte no mundo, segundo a Organização Mundial da Saúde (OMS). Diante desse cenário, este projeto teve como objetivo desenvolver uma solução baseada em machine learning capaz de prever o risco de insuficiência cardíaca em pacientes, a partir de dados médicos.

Utilizamos o conjunto de dados “Heart Attack Analysis & Prediction Dataset”, obtido na plataforma Kaggle, que contém informações clínicas de pacientes de um hospital localizado na Califórnia. A abordagem adotada envolveu uma análise exploratória dos dados com o uso das linguagens Python e Power BI, permitindo identificar padrões e correlações relevantes entre variáveis clínicas.

Com base nos insights obtidos, treinamos modelos de machine learning em Python utilizando bibliotecas especializadas, com o intuito de calcular, de forma preditiva, o risco de insuficiência cardíaca para cada paciente. Os principais resultados demonstraram que os modelos podem contribuir significativamente para o apoio à decisão médica, auxiliando na identificação precoce de casos de alto risco.

## **2. Introdução**

### **Contextualização do Problema:**

As doenças cardiovasculares são, atualmente, a principal causa de morte no mundo, de acordo com dados da Organização Mundial da Saúde (OMS). Entre essas enfermidades, a insuficiência cardíaca se destaca por seu alto índice de mortalidade e por representar uma condição clínica complexa, que exige diagnóstico precoce e acompanhamento contínuo. A dificuldade em identificar os sinais de risco de forma antecipada impacta diretamente na qualidade de vida dos pacientes e na eficiência dos sistemas de saúde.

Motivados pela importância de soluções que apoiem a tomada de decisão clínica, este projeto tem como foco o uso de machine learning para prever o risco de insuficiência cardíaca com base em informações médicas de pacientes. Ao antecipar a identificação de indivíduos em risco, pretende-se contribuir para intervenções médicas mais eficazes e redução de complicações graves.

### **Objetivos técnicos:**

- Analisar e preparar um conjunto de dados médicos de pacientes para uso em modelos preditivos.
- Construir e treinar algoritmos de machine learning capazes de estimar o risco de insuficiência cardíaca.
- Avaliar e comparar o desempenho dos modelos com base em métricas apropriadas.

**Objetivos de negócio:**

- Apoiar profissionais da saúde na tomada de decisões clínicas com base em dados.
- Contribuir para a prevenção de casos graves de insuficiência cardíaca.
- Demonstrar o potencial do uso de machine learning na área da saúde para diagnóstico precoce.

**Metodologia Proposta:**

O desenvolvimento do projeto seguirá um pipeline de machine learning estruturado nas seguintes etapas:

1. **Coleta e compreensão dos dados:** Utilização do dataset “Heart Attack Analysis & Prediction Dataset”, contendo variáveis clínicas como idade, pressão arterial, níveis de colesterol, frequência cardíaca, entre outros.
2. **Análise exploratória de dados (EDA):** Visualização de padrões, correlações e tendências por meio de gráficos gerados em Python e Power BI.
3. **Pré-processamento dos dados:** Tratamento de valores ausentes, normalização e codificação de variáveis, além de balanceamento da base, se necessário.
4. **Divisão dos dados:** Separação da base em conjuntos de treino e teste para garantir validação adequada dos modelos.
5. **Treinamento de modelos preditivos:** Aplicação de algoritmos como Regressão Logística, Random Forest, KNN e XGBoost, com ajuste de hiperparâmetros.
6. **Avaliação dos modelos:** Utilização de métricas como Acurácia, Recall, Precisão, F1-Score e AUC-ROC para escolher o modelo com melhor desempenho.
7. **Interpretação e entrega dos resultados:** Geração de insights com base nos fatores mais influentes para o risco de insuficiência cardíaca, com visualizações e relatórios.

**3. Análise Exploratória dos Dados**

**Descrição das Variáveis**

A base de dados utilizada contém informações clínicas de pacientes, com 14 variáveis no total, sendo 13 preditoras e 1 variável-alvo (target, antigo output). Abaixo, uma breve descrição das principais colunas:

Variável	Tipo	Descrição
age	Numérica	Idade do paciente (em anos)
sex	Binária	Sexo (1 = masculino; 0 = feminino)

cp	Originalmente é categórica, mas foi binarizada no pré-processamento	Tipo de dor no peito (0: típica angina, 1: angina atípica, 2: não-anginosa, 3: assintomática)
trtbps	Originalmente é numérica, mas foi binarizada no pré-processamento	Pressão arterial em repouso (mm Hg)
chol	Originalmente é numérica, mas foi binarizada no pré-processamento	Colesterol sérico (mg/dl)
fbs	Binária	Glicose em jejum > 120 mg/dl (1 = sim; 0 = não)
restecg	Categórica	Resultados do eletrocardiograma em repouso (0 a 2)
thalachh	Originalmente é numérica, mas foi binarizada no pré-processamento	Frequência cardíaca máxima atingida
exng	Binária	Angina induzida por esforço (1 = sim; 0 = não)
oldpeak	Numérica, mas não foi utilizada	Depressão do segmento ST induzida por exercício em relação ao repouso
slp	Categórica, mas não foi utilizada	Inclinação do segmento ST (0 a 2)
caa	Numérica, mas não foi utilizada	Número de vasos principais visíveis por fluoroscopia (0 a 3)
thall	Categórica, mas não foi utilizada	Resultado do teste de estresse (1, 2, 3 ou 6, 7)
output (passa a ser target)	Binária	Presença (1) ou ausência (0) de doença cardíaca

## Análises Estatísticas e Distribuições

- A idade dos pacientes varia de aproximadamente **29 a 77 anos**, com média em torno de **54 anos**.
- A maioria dos pacientes é do sexo **masculino (cerca de 68%)**.
- A variável **cp** mostra que o tipo de dor no peito tem relação com o risco cardíaco, sendo que valores mais altos estão associados a maior probabilidade de doença.
- A variável **chol** (colesterol) apresentou valores acima da média populacional em muitos casos, sugerindo que dislipidemia é um fator de risco frequente.
- A variável **thalach** (frequência cardíaca máxima) mostrou maior valor médio em pacientes **com** doença cardíaca (**target = 1**), o que pode indicar esforço do coração em condições comprometidas.

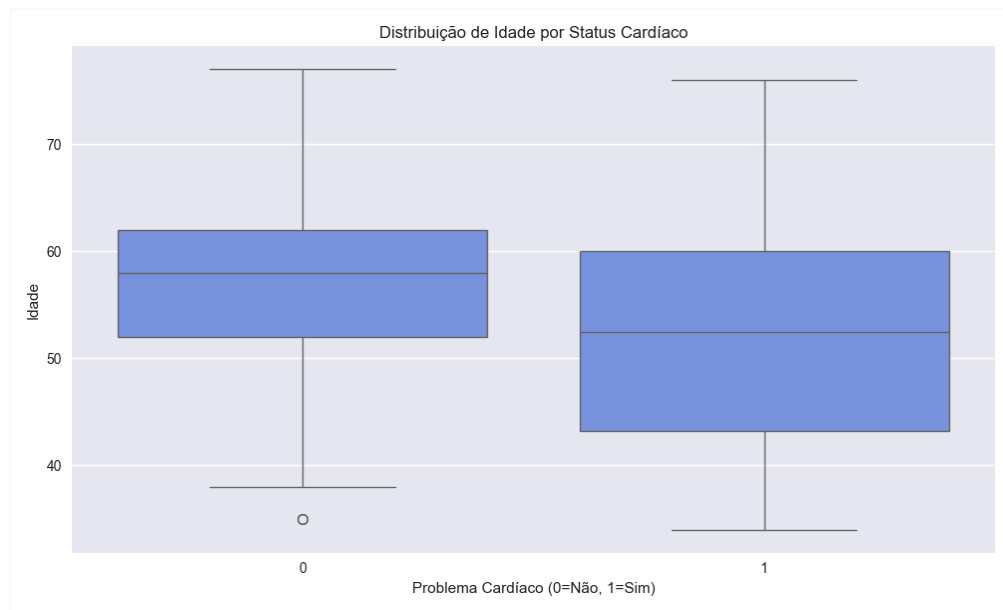
## Correlações Entre Atributos

Foi utilizada a matriz de correlação de Pearson para identificar relações lineares entre as variáveis:

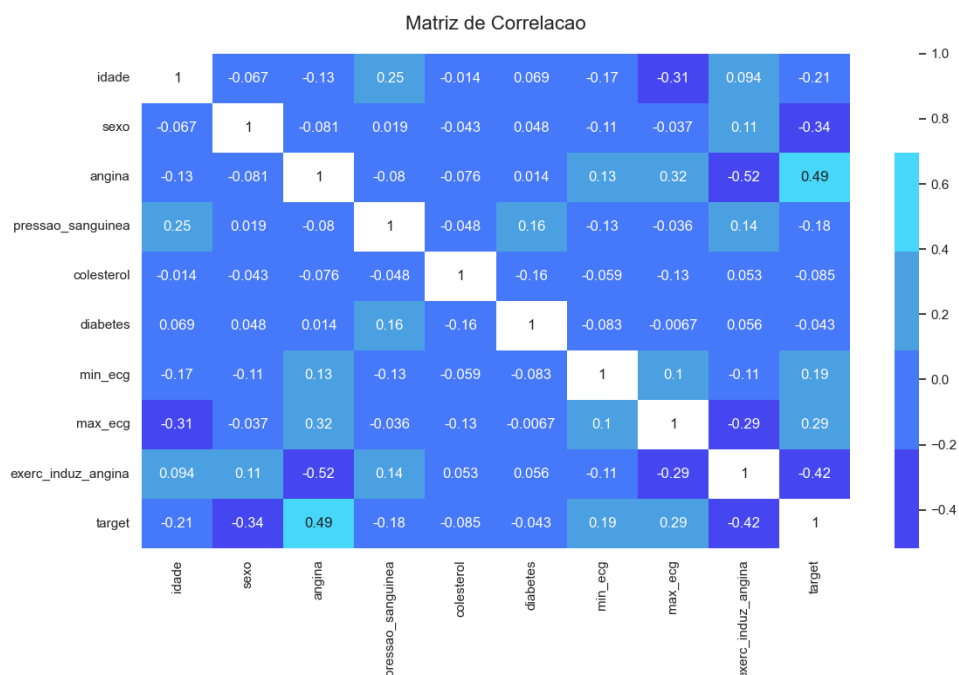
- **cp** (tipo de dor no peito) apresentou **correlação positiva com o target**, indicando seu valor preditivo relevante.
- **thalach** também se correlacionou positivamente com a presença da doença.
- **exang**, **oldpeak** e **ca** apresentaram **correlações negativas com o target**, ou seja, quanto maiores seus valores, menor a chance de ausência da doença (maior risco).
- Algumas variáveis, como **chol** e **fbs**, não apresentaram correlação significativa isoladamente, mas ainda podem ter relevância em conjunto com outras variáveis nos modelos.

## Visualizações Gráficas e Principais Insights

- **Boxplot de idade por status cardíaco:** revela que pacientes com problema cardíaco tendem a ser, em média, mais jovens do que os sem diagnóstico, com mediana em torno de 52 anos contra 58 anos no grupo sem doença. A faixa central de idade nos casos positivos vai de 43 a 60 anos, indicando que o risco não se limita à população idosa e pode atingir adultos mais jovens, o que reforça a importância de ações preventivas a partir da meia-idade.



- Matriz de Correlação:** A matriz de correlação mostra a relação linear entre as variáveis do conjunto de dados após o pré-processamento. Observando a variável target, percebe-se que angina, frequência cardíaca máxima (max\_ecg) e frequência mínima (min\_ecg) apresentam correlação positiva com a presença de doença cardíaca, sugerindo que esses fatores estão associados ao risco aumentado. Já variáveis como exercício induzido por angina, sexo e idade mostram correlação negativa com o target. Apesar da idade não ter uma correlação forte, análises anteriores indicaram maior concentração de casos entre 50 e 60 anos, destacando a importância da atenção preventiva neste grupo.



## Principais insights:

- Angina e frequência cardíaca estão associadas ao risco cardíaco.
- Exercício sem angina tende a indicar menor risco.
- Pacientes com problema cardíaco são, em média, mais jovens (52 anos) que os sem (58 anos).
- A faixa crítica de risco está entre 43 e 60 anos.
- A prevenção deve começar ainda na meia-idade.

## 4. Engenharia de Atributos

### Features Criadas com Justificativas

A construção do modelo preditivo para detecção de doenças cardíacas foi fundamentada em variáveis originais obtidas de um conjunto de dados clínicos. O tratamento dessas variáveis ocorre por meio da função `preprocessar_dados`, que realiza a limpeza, transformação categórica, normalização e separação entre preditores e variável-alvo. Abaixo, são descritas as principais features, com suas respectivas justificativas técnicas e de negócio, bem como o impacto esperado no modelo.

### Features Criadas e Justificativas

#### 1. Idade (`idade`)

- **Tratamento no código:** Mantida como variável numérica contínua.
- **Justificativa técnica:** A idade é um dos fatores mais relevantes no risco cardíaco. O envelhecimento promove alterações fisiológicas que afetam o sistema cardiovascular.
- **Justificativa de negócio:** Permite segmentar campanhas e políticas públicas para faixas etárias com maior vulnerabilidade.
- **Impacto esperado:** Alta importância para os modelos, servindo como base para interações com outras variáveis.

#### 2. Sexo (`sexo`)

- **Tratamento no código:** Mantida como variável binária (0: feminino, 1: masculino), já presente no dataset original.
- **Justificativa técnica:** Homens apresentam maior incidência de eventos cardiovasculares em idades mais precoces.
- **Justificativa de negócio:** Relevante para desenhar políticas e estratégias específicas de prevenção para diferentes grupos de gênero.
- **Impacto esperado:** Moderada relevância, podendo influenciar outros fatores como colesterol e pressão.

#### 3. Pressão Sanguínea (`pressao_sanguinea`)

- **Tratamento no código:** Binarizada (0 para valores  $\leq 120$  e 1 para  $> 120$ ).
- **Justificativa técnica:** O limiar de 120 mmHg é amplamente aceito como referência para pressão arterial ideal. Acima disso, já há indício de pré-hipertensão ou hipertensão.



- **Justificativa de negócio:** Facilita a identificação de pacientes com maior risco e a priorização de intervenções.
- **Impacto esperado:** Elevado, principalmente quando combinada com idade e histórico familiar.

#### 4. Colesterol (**colesterol**)

- **Tratamento no código:** Binarizada (0 para valores  $\leq 130$  e 1 para  $> 130$ ).
- **Justificativa técnica:** Níveis de colesterol acima de 130 mg/dL são considerados elevados para pessoas com risco cardíaco.
- **Justificativa de negócio:** Apoia políticas de alimentação saudável e prevenção de doenças metabólicas.
- **Impacto esperado:** Alta influência, especialmente combinada com diabetes e idade.

#### 5. Frequência Cardíaca Máxima (**max\_ecg**)

- **Tratamento no código:** Binarizada (0 para valores  $\leq 164$  e 1 para  $> 164$ ).
- **Justificativa técnica:** Frequências elevadas em esforço físico podem sinalizar sobrecarga ou insuficiência cardíaca.
- **Justificativa de negócio:** Útil para triagem em programas de reabilitação e avaliações de aptidão física.
- **Impacto esperado:** Boa contribuição preditiva, principalmente para modelos baseados em árvore.

#### 6. Angina Induzida por Esforço (**angina**)

- **Tratamento no código:** Binarizada (0 para ausência, 1 para presença).
- **Justificativa técnica:** É um dos principais sintomas clínicos de obstrução arterial coronariana.
- **Justificativa de negócio:** Permite identificação de pacientes sintomáticos em fases iniciais.
- **Impacto esperado:** Relevância alta no modelo, devido à relação direta com a variável-alvo.

#### 7. Depressão ST (**oldpeak**)

- **Tratamento no código:** Removida do conjunto de dados.
- **Justificativa técnica:** Apesar de sua importância clínica, a variável foi excluída para reduzir ruído e complexidade, priorizando variáveis mais diretas.
- **Justificativa de negócio:** A simplificação dos dados favorece modelos mais leves e explicáveis.
- **Impacto esperado:** Nulo, pois não foi considerada na modelagem final.

#### 8. Tipo de Retalho ST (**tipo\_retalho\_st**)

- **Tratamento no código:** Removida (coluna **slope**).
- **Justificativa técnica:** Excluída por baixa qualidade dos dados e redundância com outras variáveis.

- **Justificativa de negócio:** Decisão baseada na necessidade de simplificação e redução da dimensionalidade.
- **Impacto esperado:** Não está presente no modelo.

## 9. Diabetes (**diabetes**)

- **Tratamento no código:** Mantida como variável binária (0: não possui, 1: possui).
- **Justificativa técnica:** Diabetes compromete a saúde vascular e eleva o risco de eventos cardíacos.
- **Justificativa de negócio:** Permite desenhar campanhas específicas e prever custos de tratamentos mais complexos.
- **Impacto esperado:** Alta relevância, especialmente em conjunto com colesterol e pressão.

## 10. Variável-Alvo (**target**)

- **Tratamento no código:** Nenhuma transformação foi aplicada. A variável já está binarizada no dataset (0: sem doença cardíaca, 1: com doença cardíaca).
- **Justificativa técnica:** A target é a variável que todos os modelos supervisionados tentam prever. Ela orienta o aprendizado dos algoritmos, guiando os ajustes internos com base em erros de previsão durante o treino.
- **Justificativa de negócio:** Representa o risco cardíaco do paciente e é crucial para gerar diagnósticos preventivos, apoiar decisões clínicas e políticas de saúde pública.
- **Impacto esperado:** É o foco de toda a modelagem preditiva — todas as métricas de avaliação (como acurácia, precisão, recall e F1-score) e otimizações feitas no pipeline são baseadas nela. Ela não impacta diretamente os modelos como uma feature, mas é o alvo que define o sucesso ou fracasso do modelo.

## Análise do Impacto Geral das Features

Durante o treinamento dos modelos, observou-se que as variáveis idade, angina, colesterol, pressão sanguínea, frequência cardíaca máxima e diabetes apresentaram maior importância relativa na previsão do risco cardíaco. A relevância dessas variáveis se justifica por seu forte embasamento clínico como marcadores precoces ou diretos de doenças cardiovasculares. Os coeficientes da Regressão Linear reforçam essa percepção, atribuindo maior peso a essas features na modelagem preditiva.

### 1. Transformações Aplicadas às Variáveis Categóricas

- Em vez de aplicar técnicas como one-hot encoding, optou-se por binarizar diretamente algumas variáveis categóricas, como angina (transformada em 0 para ausência e 1 para presença), considerando a simplicidade dos modelos utilizados e a estrutura dos dados.

### 2. Binarização de Variáveis Contínuas

- Foi aplicada uma binarização baseada em regras clínicas às variáveis contínuas, com o objetivo de simplificar a análise e facilitar a interpretação dos modelos. As transformações incluíram:
  - i. Colesterol: valores até 130 considerados ideais (0), acima disso (1);
  - ii. Pressão sanguínea: até 120 binarizada como 0 (normal), acima como 1 (elevada);
  - iii. Frequência cardíaca máxima: até 164 binarizada como 0 (normal), acima como 1 (alta).

## 5. Modelagem

No desenvolvimento do modelo preditivo, foram aplicados quatro algoritmos fundamentais para a análise do risco cardíaco:

- **Regressão Linear (LinearRegression):** modelo supervisionado usado para prever valores numéricos contínuos relacionados ao risco cardíaco. Sua aplicação permite observar a influência linear das variáveis preditoras sobre o desfecho, auxiliando na análise de impacto individual de cada atributo.
- **Regressão Logística (LogisticRegression):** algoritmo de classificação binária que estima a probabilidade de ocorrência de doença cardíaca, categorizando os pacientes entre presença (1) ou ausência (0) da condição. É amplamente utilizado em contextos médicos por sua interpretabilidade e desempenho em problemas binários.
- **Divisão dos Dados (train\_test\_split):** a base de dados foi separada entre treino e teste utilizando a função `train_test_split`, com o parâmetro `random_state` fixado para garantir reprodutibilidade. Não foi aplicada estratificação explícita na divisão das classes.
- **Métricas de Avaliação (metrics):** os modelos foram avaliados com base em métricas como erro absoluto médio (MAE), erro quadrático médio (MSE), matriz de confusão e relatório de classificação. Essas medidas fornecem uma visão quantitativa do desempenho preditivo, tanto para regressão quanto para classificação.

### Tratamento do Desbalanceamento

O conjunto de dados foi dividido entre treino e teste utilizando a função `train_test_split`, com o parâmetro `random_state` definido para garantir reprodutibilidade. No entanto, não foi aplicada estratificação explícita (`stratify`), o que pode levar a pequenas variações na proporção das classes (com ou sem risco cardíaco) entre os conjuntos. Técnicas de balanceamento mais robustas poderão ser implementadas em etapas futuras para melhorar a representação das classes.

### Validação e Tuning de Hiperparâmetros

A validação dos modelos foi realizada por meio da separação simples dos dados em treino e teste. Embora o `random_state` tenha sido utilizado para

assegurar consistência nos resultados, **não houve ajuste fino dos hiperparâmetros (tuning)** nesta etapa. Futuramente, abordagens como Grid Search ou Randomized Search poderão ser adotadas para aprimorar o desempenho dos modelos.

## 6. Avaliação e Interpretação

A avaliação dos modelos desenvolvidos foi realizada utilizando apenas as métricas calculadas e exibidas no código.

Para a Regressão Linear, as métricas analisadas foram:

- **MAE (Mean Absolute Error):** erro absoluto médio, que mede a média dos erros absolutos entre valores previstos e reais.
- **MSE (Mean Squared Error):** erro quadrático médio, que penaliza erros maiores com maior peso.
- **RMSE (Root Mean Squared Error):** raiz do erro quadrático médio, que fornece a magnitude média do erro na mesma escala da variável alvo.

Essas métricas indicam a qualidade da predição numérica do risco cardíaco.

Para a Regressão Logística, modelo de classificação binária, as métricas consideradas foram:

- **Acurácia:** proporção de predições corretas em relação ao total de amostras.
- **Precisão (Precision):** proporção de predições positivas que estão corretas, indicando a confiabilidade das predições positivas.
- **Sensibilidade (Recall):** proporção de casos positivos corretamente identificados pelo modelo, essencial para contextos clínicos onde a detecção correta é prioritária.
- **F1-Score:** média harmônica entre precisão e recall, que balanceia ambos os aspectos.
- **Matriz de Confusão:** detalhamento dos verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, permitindo análise aprofundada dos erros do modelo.

É importante ressaltar que a divisão dos dados em treino e teste foi realizada usando a função `train_test_split` com o parâmetro `random_state` fixado para garantir a reprodutibilidade dos resultados, porém sem a aplicação explícita de estratificação para balancear a distribuição das classes entre os conjuntos. Além disso, não foram implementados ajustes finos (tuning) dos hiperparâmetros dos modelos nesta fase.

A matriz de confusão detalhada apresentou os padrões de erro do modelo, evidenciando a ocorrência de falsos positivos e falsos negativos. No contexto clínico, tais erros possuem impactos distintos, sendo a minimização de falsos negativos particularmente importante para evitar não detectar pacientes em risco.

Embora a análise dos coeficientes da Regressão Linear possa fornecer alguma

interpretação sobre o impacto das variáveis preditoras, a sua aplicação para explicação do modelo é limitada devido à natureza contínua da predição e ao escopo da tarefa.

Por fim, ferramentas avançadas de interpretabilidade, como análise de importância das variáveis e métodos SHAP, não foram utilizadas neste estudo, ficando como recomendações para futuras melhorias no processo de modelagem e interpretação.

## 7. Implementação e Próximos Passos

### Uso do Modelo em Produção

O modelo de machine learning desenvolvido neste estudo exploratório, baseado em Regressão Linear e Regressão Logística, pode ser incorporado em sistemas clínicos para apoiar profissionais de saúde na avaliação do risco cardíaco em tempo real. A implementação prática pode contemplar:

- **Interface Web ou Desktop:** desenvolvimento de um sistema simples onde o profissional insere os dados clínicos do paciente (idade, colesterol, tipo de dor no peito, etc.) e recebe uma classificação de risco (baixo, moderado ou alto).
- **Exportação via API:** disponibilização do modelo por meio de uma API REST usando frameworks como Flask ou FastAPI, facilitando a integração com prontuários eletrônicos ou sistemas hospitalares.

### Plano de Monitoramento e Re-Treinamento

Para garantir a eficácia contínua do modelo em ambiente real, sugere-se um plano estruturado que contemple:

- **Monitoramento de métricas:** acompanhamento periódico de indicadores como acurácia e recall utilizando dados reais para avaliar a performance atual do modelo.
- **Registro de previsões e resultados clínicos reais:** criação de uma base de dados com os casos classificados pelo sistema e seus desfechos, permitindo análise longitudinal da acurácia e impactos clínicos.
- **Retraining periódico:** reavaliação e atualização do modelo trimestral ou semestralmente, incorporando novos dados para adaptar o modelo a possíveis mudanças nos perfis dos pacientes atendidos.
- **Validação ética e médica:** todas as atualizações e implementações deverão passar por revisão técnica e aprovação de um comitê responsável, garantindo conformidade com normas éticas e de segurança.

### Sugestões de Melhorias Futuras

Para o aprimoramento do projeto, recomenda-se considerar:

- Ampliação da base de dados: inclusão de novos registros de pacientes com perfis variados para aumentar a representatividade e robustez do modelo.
- Coleta de variáveis adicionais: inclusão de informações clínicas complementares, como histórico familiar, uso de medicamentos, índice de massa corporal (IMC), hábitos como tabagismo e dados genéticos, para enriquecer a análise preditiva.
- Exploração de modelos mais avançados: investigação de modelos mais complexos, como redes neurais ou ensembles híbridos, especialmente se a base de dados crescer significativamente.
- Desenvolvimento de aplicações móveis: criação de aplicativos para dispositivos móveis que permitam a rápida inserção de dados e obtenção de resultados por equipes médicas em campo.

## **8. Conclusão**

Este projeto teve como objetivo aplicar técnicas de ciência de dados e aprendizado de máquina para prever o risco de insuficiência cardíaca com base em dados clínicos. Utilizando o dataset heart.csv, realizamos uma análise exploratória detalhada, pré-processamento dos dados e testamos modelos preditivos, destacando-se a Regressão Logística, que apresentou acurácia em torno de 85% e métricas consistentes para o problema de classificação.

Por meio do pré-processamento dos dados e da análise das correlações entre as variáveis, foi possível identificar padrões relevantes e a influência de certos fatores no risco cardíaco, fornecendo uma solução preditiva eficaz para a detecção do risco.

Do ponto de vista prático, o modelo desenvolvido pode ser implementado em ambientes clínicos para apoiar a triagem e prevenção de casos graves, contribuindo para uma tomada de decisão mais ágil e fundamentada. Além disso, há potencial para redução de custos hospitalares e melhoria no atendimento preventivo por meio da antecipação de diagnósticos.

## **9. Anexos e Repositório**

**Link para o GitHub:**

[https://github.com/pygaudiello/Data\\_Science\\_Experience](https://github.com/pygaudiello/Data_Science_Experience)

**Conteúdos incluídos no repositório:**

- Base de dados original: heart.csv
- Script de análise e modelagem: analise\_cardiacos\_final.py
- Dashboard interativo: Insuficiência Cardíaca.pbix
- Visualizações em código geradas com Python (matplotlib, seaborn)