

Dataset de previsão de insuficiência cardíaca

Felipe Carvalho de Paula Silva

8 de maio de 2024

0.1 Introdução

Este trabalho tem como objetivo explorar o *Heart Failure Prediction Dataset*, um conjunto de dados relacionados à previsão de insuficiência cardíaca. A gestão eficiente desses dados é fundamental para entender os fatores de risco e desenvolver modelos de previsão confiáveis. Com o crescente volume de informações geradas diariamente, torna-se crucial estabelecer sistemas robustos para armazenamento, recuperação e manipulação desses dados. Nesse contexto, os bancos de dados desempenham um papel fundamental, fornecendo uma estrutura organizada para armazenar e acessar dados de forma eficiente.

Empresas de todos os portes dependem de sistemas de previsão de insuficiência cardíaca para tomar decisões estratégicas, oferecer tratamentos adequados aos pacientes e melhorar a qualidade de vida. Portanto, compreender os princípios e práticas relacionados à análise de dados de saúde é essencial para profissionais da área médica e de tecnologia da informação.

0.2 Dicionário de Dados e Fragmento dos Dados

Tabela 1 – Dicionário de Dados

Campo	Tipo de Dado	Descrição
Age	Numérico	Idade do paciente
Sex	Categórico	Sexo do paciente
ChestPainType	Categórico	Tipo de dor no peito
RestingBP	Numérico	Pressão arterial em repouso
Cholesterol	Numérico	Nível de colesterol
FastingBS	Binário	Glicemia em jejum
RestingECG	Categórico	Resultado do ECG em repouso
MaxHR	Numérico	Frequência cardíaca máxima
ExerciseAngina	Binário	Angina induzida por exercício
Oldpeak	Numérico	Depressão do segmento ST
ST_Slope	Categórico	Inclinação do segmento ST
HeartDisease	Binário	Doença cardíaca

Tabela 2 – Fragmento dos Dados

Age HeartDisease	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope
40 0	M	ATA	140	289	0	Normal	172	N	0	Up
49 1	F	NAP	160	180	0	Normal	156	N	1	Flat
37 0	M	ATA	130	283	0	ST	98	N	0	Up
48 1	F	ASY	138	214	0	Normal	108	Y	1.5	Flat

0.3 Definição de Problema

Como os fatores de saúde, como idade, sexo, pressão arterial, nível de colesterol e outros, influenciam na probabilidade de uma pessoa desenvolver insuficiência cardíaca?

0.4 Análise dos dados

A partir do dataset, é possível concluir que a média do colesterol dos voluntários é de 198,80 mg/dL. A idade com maior ocorrência de problemas cardíacos é 58 anos, algo que nos permite identificar uma idade de risco. Também é possível identificar que aproximadamente metade das pessoas (49,4%) que possuem problemas cardíacos experienciam dores no peito ao fazerem exercícios. A média de idade dos entrevistados é de 53 anos, com um total de 919 pessoas.

0.5 Visualização dos dados

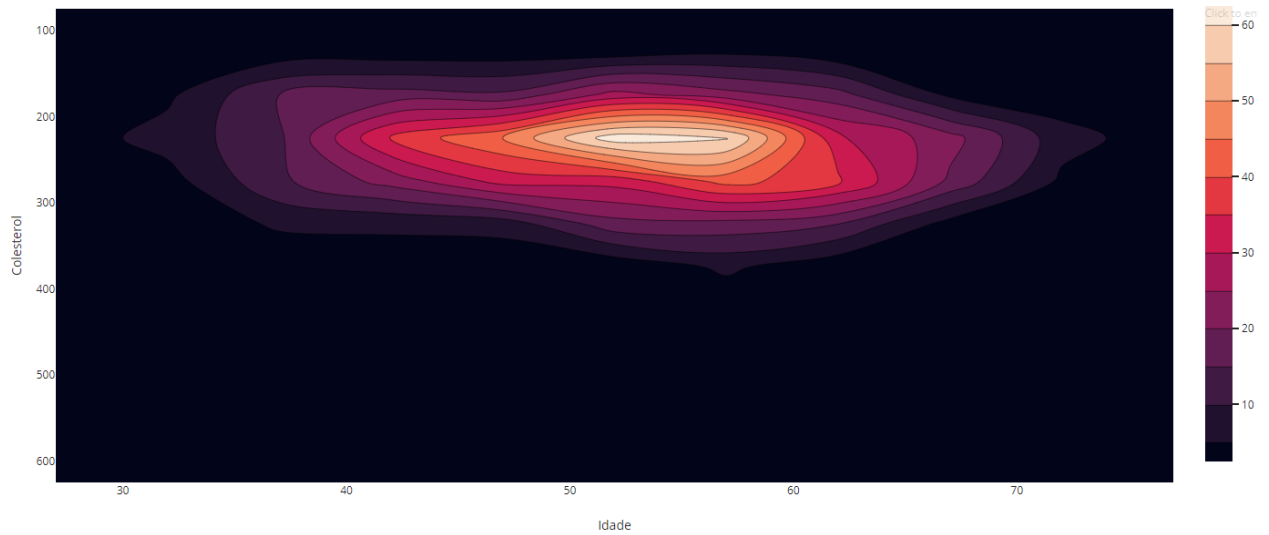


Figura 1 – Mapa de calor de Correlação entre colesterol e idade

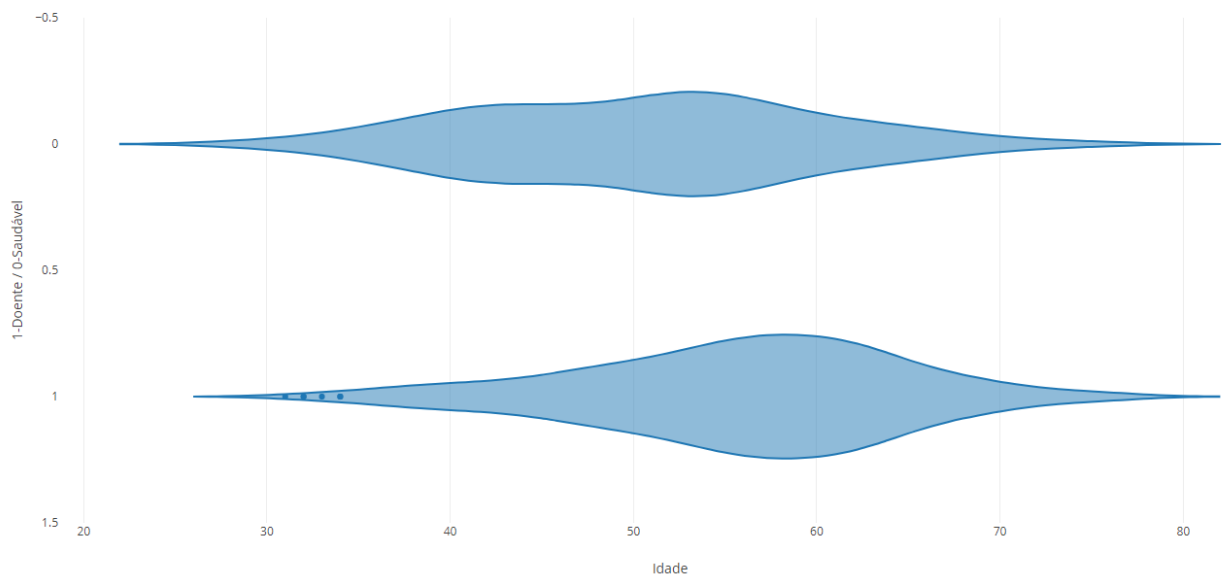


Figura 2 – Idades com maior ocorrência de doença cardíaca

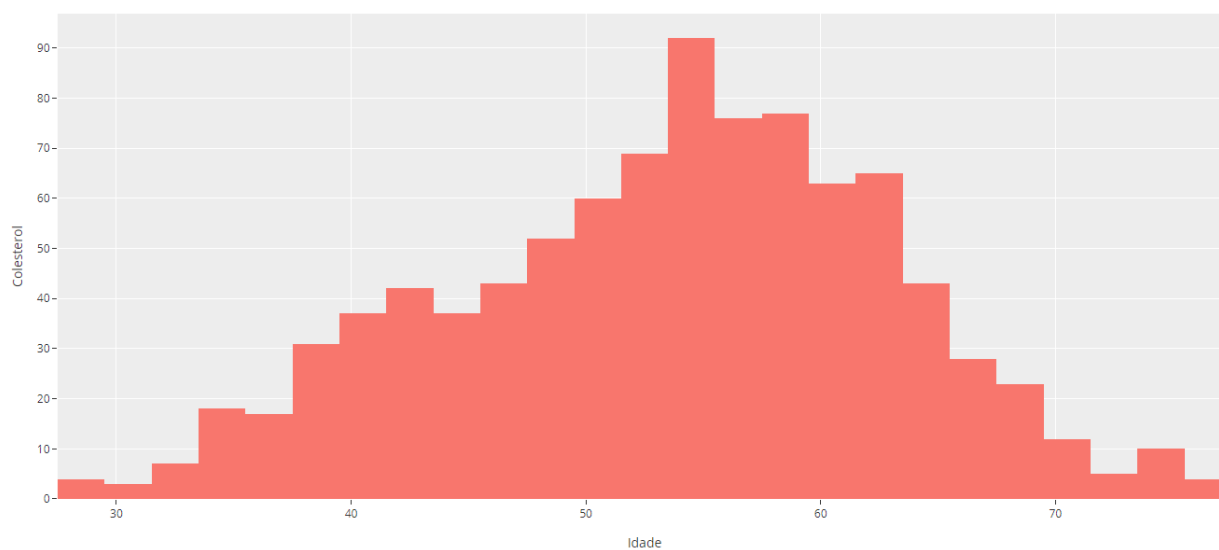


Figura 3 – Correlação entre colesterol e idade

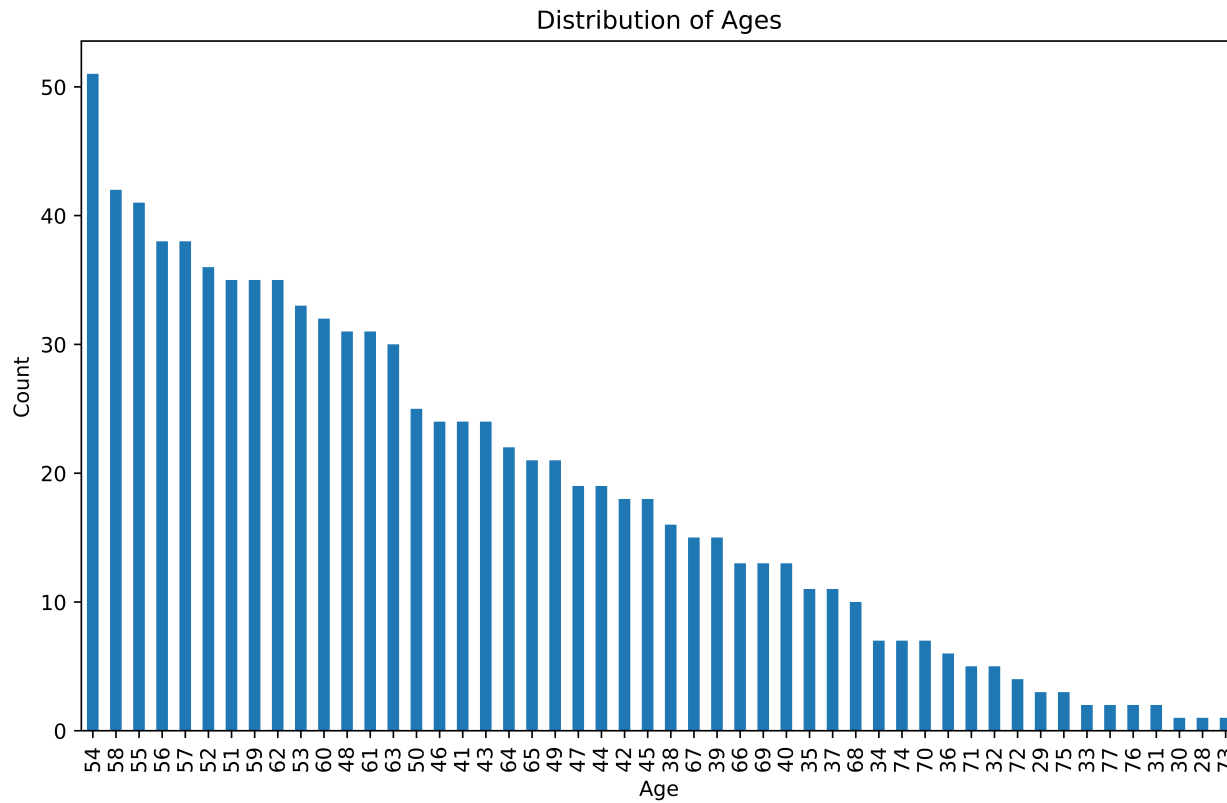


Figura 4 – Idades dos voluntários

0.6 Conclusão

A análise dos dados apresentados nos permite concluir que a idade, sexo, pressão arterial, nível de colesterol e outros fatores de saúde desempenham um papel significativo na probabilidade de uma pessoa desenvolver insuficiência cardíaca. Essas informações são cruciais para a prevenção, diagnóstico e tratamento eficazes da doença cardíaca. Porém, é importante observar que os dados analisados não incluem pessoas mais jovens, tendo uma idade média de 53 anos. Isso pode dificultar a identificação das faixas etárias mais vulneráveis a problemas cardíacos, uma vez que uma parte significativa da população com maior risco pode não estar representada nos dados.

0.7 Fonte

Os dados utilizados neste trabalho foram obtidos do Kaggle. Para acessar o dataset completo, visite: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data>.