



Heart Failure Prediction Dataset

Grupo Iniciante



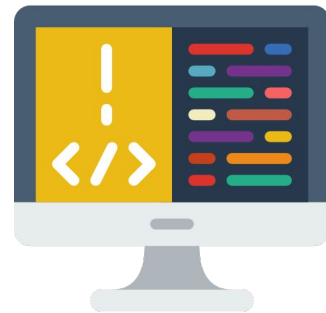
Introdução ao problema e ao Python

Introdução



Como foi falado na reunião de boas vindas, uma das frentes em que os integrantes do Hype atuam são os grupos de estudo, onde temos que resolver um problema utilizando ferramentas de análise de dados.

No grupo de nível iniciante, há um pouco mais de auxílio por parte do mentor, visto que os integrantes possivelmente estão tendo o primeiro contato com as ferramentas e conceitos, porém você é totalmente livre (inclusive encorajado) a continuar os estudos de forma individual, mas claro sempre trazendo as discussões nas nossas reuniões 😊

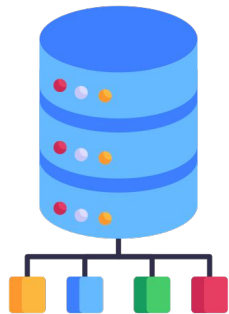




Conjunto de Dados

Para resolvermos o problema proposto, precisamos trabalhar com o pilar fundamental dos projetos de Ciência de Dados: o famoso conjunto de dados (comumente chamado de **dataset**).

Ele é basicamente uma coleção de dados em formato tabular (tabela), onde em cada linha temos um “indivíduo” do conjunto e em cada coluna temos as características.



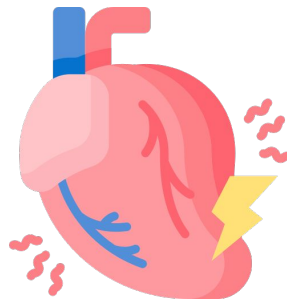
Problema



No nosso caso, temos um dataset relacionado com a área da saúde, onde em cada linha temos informações de um paciente e, com base nessas informações, nosso problema é basicamente prever se um determinado paciente pode ou não ser acometido com um problema cardiovascular, com base em informações médicas presente nas colunas do dataset.

“As doenças cardiovasculares são a causa número 1 de morte em todo o mundo, levando cerca de 17,9 milhões de vidas a cada ano, o que representa 31% de todas as mortes em todo o mundo.

Pessoas com doença cardiovascular ou com alto risco cardiovascular (devido à presença de um ou mais fatores de risco, como hipertensão, diabetes, hiperlipidemia ou doença já estabelecida) precisam de detecção e gerenciamento precoces, onde um modelo de aprendizado de máquina pode ser de grande ajuda.” **dataset:** kaggle.com/datasets/fedesoriano/heart-failure-prediction

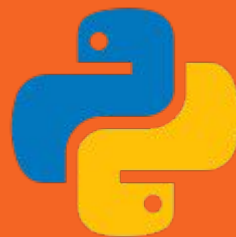




Ferramentas

Abaixo temos uma listagem das principais ferramentas que vamos utilizar durante o estudo do dataset:

- **Telegram:** Nosso canal de comunicação (como de costume).
- **Drive:** Local onde vamos colocar todos os arquivos (compartilhado).
- **Discord:** Local onde vamos nos reunir para discutir as coisas.
- **Python:** linguagem de programação utilizada para a resolução do problema.
- **Kaggle:** portal onde temos o nosso conjunto de dados.
- **Google Colaboratory:** ambiente onde vamos programar em Python.



Descrição Atributos



Age	Idade do paciente (em anos)
Sex	Sexo do paciente (M ou F)
ChestPainType	Tipo de dor no peito (4 valores)
RestingBP	Pressão sanguínea em repouso (mm Hg)
Cholesterol	Colesterol sérico (mg/dl)
FastingBS	Açúcar no sangue em jejum (0 ou 1)
RestingECG	resultados do eletrocardiograma repouso
MaxHR	Frequência cardíaca máxima alcançada
ExerciseAngina	Angina induzida por exercício (Y ou N)
OldPeak	Depressão de ST induzida por exercício em relação ao repouso.
ST_slope	Inclinação do segmento ST de pico do exercício.



Dataset

g18
observações
(linhas)

11 atributos
(colunas)



Observações Atributos

Angina

angina é um tipo de dor no peito causada por bloqueios nos vasos sanguíneos que levam ao seu coração

ChestPainType

TA: Typical Angina

ATA: Atypical Angina

NAP: Non-Anginal Pain

ASY: Asymptomatic

FastingBS

1: caso açúcar no sangue > 120 mg/dl

0: caso contrário

MaxHR

Valor numérico entre **60** e **202**

ST_Slope

Up: crescente

Flat: plano

Down: decrescente

HeartDisease

coluna do dataset que representa a classe de saída (1 para doença cardíaca e 0 normal)

RestingECG

- **Normal:** normal
- **ST:** com anormalidade da onda ST-T
- **LVH:** mostrando provável ou definitiva hipertrofia ventricular esquerda





Como foi falado anteriormente, o Python vai ser a linguagem em que estaremos utilizando para análise de dados e resolução do problema proposto.

Resumidamente, o Python é uma das linguagens mais populares, e isso se deve ao fato de ser uma linguagem que possui uma sintaxe relativamente simples e possuir inúmeras bibliotecas, fazendo com que seja uma linguagem altamente versátil e famosa em áreas como Engenharia, Matemática, dentre outros.

Links Úteis:

- O que é Python? - <https://kenzie.com.br/blog/o-que-e-python/>
- Learn Python (Kaggle) <https://www.kaggle.com/learn/python>
- Introdução à Ciência da Computação com Python: <https://pt.coursera.org/learn/ciencia-computacao-python-conceitos>
- Começando com o Google Colab: <https://colab.research.google.com>
<https://www.youtube.com/watch?v=inN8seMm7UI>

É de suma importância que vocês também participem do dojo do Hype, pois é lá que irão aprender desde o básico da linguagem até temas mais avançados.



**Vamos para o
Colab!**

