



Projeto: Prever Insuficiência Cardíaca Baseado em Fatores Clínicos e Comportamentais.

1. Objetivo

A Insuficiência Cardíaca (causada por doenças cardiovasculares) é a principal causa de morte de vidas todos os anos, representando um percentual de 31% de todas as mortes mundiais. Este conjunto de dados contém doze características divididas em fatores clínicos e comportamentais como, por exemplo, nível de plaquetas e sódio sérico no sangue, fumante ou não, hipertenso ou não e pode ser utilizado para prever a mortalidade por doenças cardiovasculares. Objetivo do trabalho é realizar uma análise exploratória sobre o conjunto de dados e descobrir uma série de insights para auxiliar a equipe médica para a tomada de ações preventivas.

2. Fonte de dados

Como entrada temos um arquivo no formato csv (base de dados pública) que está disponível em: <https://www.kaggle.com/andrewmvd/heart-failure-clinical-data>

3. Dicionário de Dados

Variável	Tipo	Descrição
age	inteiro	Representa a idade em anos
anaemia	lógico	Representa se possui anemia ou não.
creatinine_phosphokinase	inteiro	Nível da enzima CPK no sangue (mcg / L).
diabetes	lógico	Representa se possui diabetes ou não.
ejection_fraction	inteiro	Porcentagem de sangue saindo do coração a cada contração (porcentagem).
high_blood_pressure	lógico	Representa se é hipertenso ou não.
platelets	real	Nível de plaquetas no sangue (kiloplatelets/mL).
serum_creatinine	real	Nível de creatina sérica no sangue (mg/dL).
serum_sodium	inteiro	Nível de sódio sérico no sangue (mEq / L).
sex	caractere	Representa o sexo: M (masculino) ou F(feminino).
smoking	lógico	Representa se é fumante ou não.
time	inteiro	Tempo de acompanhamento em dias.
DEATH_EVENT	lógico	Representa se faleceu ou não.

4. Principais Indicadores

Será construído um dashboard com os indicadores abaixo:

1. Total de registros dividido entre o número de vivos e óbitos observados.
2. Contagem de óbitos categorizados por Idade.
3. Top 5 dos tempos maiores de acompanhamento em dias.
4. Total de fumantes e porcentagem de fumantes categorizados pelo sexo.
5. Total de anêmicos categorizados pelo sexo.

Outros indicadores.

6. Total de anêmicos categorizados pela idade.



7. Total de diabéticos categorizados pela idade.
8. Nível de plaquetas por anêmicos.
9. Tempo médio, tempo mínimo e tempo máximo de acompanhamento em Dias.
10. Nível de sódio sérico por pacientes que são diabéticos.
11. Nível de creatina sérica por hipertensos.

Há uma série de outros indicadores que podem auxiliar no acompanhamento dos registros listados no dashboard.

5. Público-Alvo.

Equipe-médica (médicos, enfermeiros, técnicos).

6. Possível Aplicação.

Implantação em um setor de internação ou UTI no hospital de modo a monitorar o pacientes quanto aos fatores críticos que podem influenciar uma doença cardiovascular. Monitorar em especial as taxas (ex: nível de sódio sérico) quanto aos níveis críticos que podem comprometer a saúde.

7. Possível Melhoria.

Aplicar Aprendizado de Máquina de modo a construir um modelo preditivo que ao assumir tais valores referentes aos fatores clínicos e comportamentais, diz a probabilidade de eventual morte causada por insuficiência cardíaca.

8. Tecnologias.

Foram usadas a ferramenta de visualização Microsoft PowerBI e linguagem de programação Python. Python foi escolhida por ter facilidade para codificar atividades de análise de dados, através de bibliotecas com muitas funções úteis. Principais bibliotecas usadas: pandas, matplotlib, seaborn e squarify.

9. Links de Acesso

Todos eles são públicos.

[Link do Dashboard](#)