Heart Failure Clinical Records Dataset

📌 Problem Statement

Heart failure is a leading cause of mortality worldwide, and early prediction of patient outcomes can significantly improve treatment strategies. This project aims to build a predictive model using clinical data to determine the likelihood of patient survival. The goal is to analyze key medical parameters and their impact on mortality, helping healthcare professionals make informed decisions.

📊 Dataset Overview

This dataset contains medical records of **299** patients who experienced heart failure. The data includes various clinical parameters that may impact patient survival.

📊 Dataset Information

- **Total Records:** 299
- **Total Variables:** 13
- **Target Variable:** DEATH_EVENT (0 = Survived, 1 = Died)

📂 Column Descriptions

Column Name	Description	Data Type
age	Age of the patient (years)	Float
anaemia	Decrease of red blood cells (0 = No, 1 = Yes)	Integer
creatinine_phosphokinase	CPK enzyme level (mcg/L)	Integer
diabetes	Whether the patient has diabetes (0 = No, 1 = Yes)	Integer
ejection_fraction	Percentage of blood leaving the heart each contraction	Integer
high_blood_pressure	Hypertension status (0 = No, 1 = Yes)	Integer
platelets	Platelet count (kiloplatelets/mL)	Float
serum_creatinine	Level of creatinine in blood (mg/dL)	Float
serum_sodium	Sodium level in blood (mEq/L)	Integer
sex	Gender (0 = Female, 1 = Male)	Integer
smoking	Whether the patient smokes (0 = No, 1 = Yes)	Integer
time	Follow-up period (days)	Integer
DEATH_EVENT	Survival status (0 = Alive, 1 = Deceased)	Integer

🔍 Exploratory Data Analysis (EDA)

Some key insights that can be explored:

- **Age vs Death Rate:** How survival rate changes with age.
- **Ejection Fraction vs Mortality:** Impact of heart pumping capacity on survival.
- **Serum Creatinine & Sodium:** Relationship with mortality.
- **Smoking & High Blood Pressure:** Influence on heart failure.
- **Correlation Analysis:** Identifying important features.

📈 **Suggested Visualizations**

- Boxplots for numerical variables vs DEATH_EVENT
- Countplots for categorical variables
- Heatmap for feature correlation
- Distribution plots for numerical features

⚙️ **Usage**

This dataset can be used for:

- **Predictive Modeling:** Classifying survival status using machine learning.
- **Feature Analysis:** Understanding which clinical features impact heart failure most.
- **Data Visualization:** Exploring patterns in patient data.

🔗 **Source**

This dataset is available publicly and is widely used in research for predictive modeling in healthcare.