**Heart Failure Prediction – logistic Regression, KNN, Decision Tree, Random Forest, XGBoost**

**Problem Statement:**

**Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Heart failure is a common event caused by CVDs and this dataset contains 12 features that can be used to predict mortality by heart failure. Most cardiovascular diseases can be prevented by addressing behavioural risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies. People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help.**

**Feature Description**

* **Age** : age [years]
* **anaemia** : Decrease of red blood cells or haemoglobin (Boolean)
* **creatinine phosphokinase** : Level of the CPK enzyme in the blood (mcg/L)
* **diabetes** : If the patient has diabetes (Boolean)
* **ejection fraction** : Percentage of blood leaving the heart at each contraction (percentage)
* **high\_blood\_pressure** : If the patient has hypertension (Boolean)
* **platelets** : Platelets in the blood (kilo platelets/mL)
* **serum\_creatinine** : Level of serum creatinine in the blood (mg/dL)
* **serum sodium** : Level of serum sodium in the blood (mEq/L)
* **sex** : Woman or man (binary)
* **smoking** : If the patient smokes or not (Boolean)
* **time** : Follow-up period (days)
* **DEATH\_EVENT** : If the patient deceased during the follow-up period (Boolean)