**PROBLEM STATEMENT -1**

Predict Heart Failure, create a model to assess the likelihood of a death by heart failure event from scratch.

EDA

The number of missing values in each column is checked. Dataset is checked with datatypes. As a result age, platelets, serum\_creatinine are in float and others are in integer and by analysing, the age and platelets column is converted to integer datatype whereas serum\_creatinine cannot be converted as it affects the whole dataset when it is converted. Regarding removal of outliers, creatinine phospokinase does not contains the outliers because there are case reports related to higher/abnormal CPK levels in blood.

DATA VISUALIZATION

Each column in dataset is visualized with count plot. For example, anaemia patients survival rate, death rate and non-anaemic patients death rate, survival rate is analysed in single graph to check the dependency of anaemia patients death rate. Feature elimination is done by analysing 2 methods.

1. Random forest classifier and importance of each attribute is retrieved as a graph.

2. Heatmap is done for correlation matrix.

By combining both methods and analysing, some features are eliminated for obtaining accuracy.

MODEL

The ML model have used is Logistic regression. Logistic regression is the go-to linear classification algorithm for two-class problems. Logistic regression uses an equation as the representation, very much like linear regression. Input values (**X**) are combined linearly using weights or coefficient values to predict an output value (**y**).A key difference from linear regression is that the output value being modelled is a binary value (0 or 1) rather than a numeric value. The confusion matrix is displayed as graph, to depict or show the wrongly predicted values.

**PROBLEM STATEMENT -2**

Perform Sentimental Analysis from scratch using python. In any Twitter post, there are various sentiments behind it. So perform sentimental analysis and find out the different sentiments in the dataset.

FINDING SENTIMENTS USING POLARITY

The tweet column is taken from dataset and tweet is cleaned, the overall sentiment is often inferred as positive, neutral or negative from the sign of the polarity score.