AIM - HEART DISEASE PREDICTION

DATASET DESCRPTION -

One datset csv file from kaggle

Heart.csv(1025\* 14)

It contain input and output variable(s). We need to predict the heart disease for the test data set.

Following columns are contained in csv file:-

ATRIBUTE INFORMATION

1. age
2. sex
3. chest pain type (4 values)
4. resting blood pressure
5. serum cholestoral in mg/dl
6. fasting blood sugar > 120 mg/dl
7. resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved
9. exercise induced angina
10. oldpeak = ST depression induced by exercise relative to rest
11. the slope of the peak exercise ST segment
12. number of major vessels (0-3) colored by flourosopy
13. thal: 0 = normal; 1 = fixed defect; 2 = reversable defect  
    The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

**Libraries Used --**

1.Numpy is used as np for linear algebra in our dataset.

2. Pandas is used as pd for data processing reading our csv files.

3. OrdinalEncoder from sklearn.preprocessing is used to interpret the ordinal datatypes.

4. train\_test\_split from sklearn.model is used for spliting the csv file into two different parts for training and testing and this traning and testing file is further divided into into two parts x and y value i.e. input and output values .

5. Logistic regression from sklearn.linear\_model is used as model for our data prediction.