**32.Heart disease prediction**

Problem Statement:

Now-a-days many people are suffering from heart diseases but due to the unawareness of their disease they are losing their lives.

In order to save their lives we build a model based on the clinical parameters of the patient and finds out whether the person is suffering from heart disease or not.

Variables:

1. age: continuous

2. sex: categorical, 2 values {0: female, 1: male}

3. cp (chest pain type): categorical, 4 values

{1: typical angina, 2: atypical angina, 3: non-angina, 4: asymptomatic angina}

4. restbp (resting blood pressure on admission to hospital): continuous (mmHg)

5. chol (serum cholesterol level): continuous (mg/dl)

6. fbs (fasting blood sugar): categorical, 2 values {0: <= 120 mg/dl, 1: > 120 mg/dl}

7. restecg (resting electrocardiography): categorical, 3 values

{0: normal, 1: ST-T wave abnormality, 2: left ventricular hypertrophy}

8. thalach (maximum heart rate achieved): continuous

9. exang (exercise induced angina): categorical, 2 values {0: no, 1: yes}

10. oldpeak (ST depression induced by exercise relative to rest): continuous

11. slope (slope of peak exercise ST segment): categorical, 3 values

{1: upsloping, 2: flat, 3: downsloping}

12. ca (number of major vessels colored by fluoroscopy): discrete (0,1,2,3)

13. thal: categorical, 3 values {3: normal, 6: fixed defect, 7: reversible defect}

14. target: (diagnosis of heart disease):

{ 0:not having heart disease,1:having heart disease}

Removed variables:

fbs and chol variables are removed because there is less correlation between those input and target variables.

Approach:

Step 1: Data is imported and analyzed by using some methods like info, corr and isnull etc..

Step 2: There are no null values present in the data so no changes are done.

Step 3: There is less correlation between fbs,chol and target variable so they are droped.

Step 4: Data is visualized by using some plots like pairplot,hist and countplot .

Step 5: Data is splitted into train set(70%) and test set(30%) by using train-test-split module.

Step 6: After preprocessing the data totally 8 machine-learning algorithms are applied to it

Step 6.1: 1)Linear Regression with accuracy of 60.79

Step 6.2: 2)Logistic Regression with accuracy of 92.307

Step 6.3: 3) Rondom Forest Regression with accuracy of 66.114

Step 6.4: 4) KNN classifier with accuracy of 79.125

Step 6.5: 5) decision tree with accuracy of 83.516

Step 6.6: 6) XGBoost with accuracy of 24.144

Step 6.7: 7) naive-bayes with accuracy of 90.109

Step 6.8: 8) support vactor machine with accuracy of 92.307