**HEART DISEASE PREDICTION AND IDENTIFICATION**

**with MACHINE LEARNING**

Heart-related diseases are among the most widely recognized reasons for death around the world. Patients are frequently asymptomatic until a deadly occasion occurs, and in any event, when they are under perception, a prepared workforce is required to recognize a heart abnormality. Somewhat recently, there has been expanding proof of how Machine Learning can be utilized to recognize such oddities, because of the accessibility of Electrocardiograms (ECG) in computerized design. New advancements in innovation have permitted us to take advantage of such information to fabricate models ready to dissect the examples in the event of heartbeats, and spot abnormalities from them. In this work, master cardiologists across various clinics and nations, and can recognize 7 sorts of signs: Typical, AF, Tachycardia, Bradycardia, Arrhythmia, Other, or Boisterous.

The determination of heart illness has turned into a troublesome clinical errand in the current clinical examination. This finding relies upon the point-by-point and exact examination of the patient's clinical test information on a singular's well-being history. The huge improvements in the field of profound learning look to make astute mechanized frameworks that help specialists both to anticipate and to decide the illness with the web of things with Machine learning. Certainty adjustment is an especially important issue in a medical care setting like the one tended to in this composition: when a neural network model makes an expectation, it is vital that this result can be relied upon.

The Machine Intelligence algorithm will be applied in the prediction of heart disease with different types of attributes with the dataset.

**Problem definition:**

Predicting and identifying the rate of intricate ongoing circumstances, for example, cardiovascular breakdown is still a challenging task. observational proof of strong expectation of mind-boggling ongoing circumstances, for example, Heart Failure is restricted. Noticeably the existing designs have shown humble execution in huge scope, complex heart datasets for risk expectation of different circumstances including Heart failure. capacity to show brings about a language justifiable by people. This has restricted their trustfulness and commitment to take a chance with factor disclosures and more extensive clinical reception.

**Project Execution Plan:**

The goal of machine learning-based cardiac disease diagnosis is to identify the illness itself in its earliest stages using the available qualities. This study will take into account the dataset that contains information about heart illness. The dataset will undergo pre-processing, and the noisy and null value data will be eliminated from the dataset. To facilitate further processing, the data will be evaluated and displayed. It will be decided to use a machine intelligence algorithm to produce the forecast.

Two portions of the dataset will be analyzed. The initial dataset is divided into two parts: training and testing. The first half of the dataset is used to give 70% of the training for the machine learning algorithm.

**Contribution :**

The Machine Learning algorithm will be a python-based application that contributes to finding out the heart disease's early stage. It will be helpful for humans to detect it early and to take the necessary treatments at the correct time.

**Evaluation:**

The project evaluation can be tested with the Machine Learning algorithm prediction results. Since the Machine Learning algorithm will be used to predict the disease, the accuracy of the algorithm result will be helpful to evaluate the results. The accuracy score of the algorithm in heart disease identification helps to evaluate the dataset.

The application will be developed with Google Colab Python Tool as the project can be directly executed in any type of computer system with an internet connection. There is no need for any specific software to be installed in the user system. The Colab Tool helps to develop and run the application directly inside the cloud server where the Python library files are installed. The Machine Learning algorithm libraries are built inside the Colab. It helps the project to use the Machine Learning algorithm in the finding of heart disease.

**Techniques Applied:**

The task included examination of the heart disease patient dataset with appropriate information handling. Then, at that point, various models were prepared and expectations are made with various calculations Logistic Regression, KNN, Random Forest ,SVM model. The machine learning library and machine learning libraries Sklearn and Keras' are applied into the application.

**Dataset:**

Most of the columns in a dataset are noisy and contain lots of information. But with feature engineering do, will get more good results. The first step is to import the libraries and load data. After that will take a basic understanding of data like its shape, sample, is there are any NULL values present in the dataset. Understanding the data is an important step for prediction or any Artificial Learninglearning project. It is good that there are no NULL values.

**Detailed Design of Features:**

This dataset contain the fields needed for the analyzing of the heart disease dataset.Exploratory examination is a cycle to investigate and comprehend the information and information relationship in a total profundity with the goal that it makes highlight designing and demonstrating steps smooth and smoothed out for expectation. Exploratory examination assists with validating our presumptions or misleading.

**Analysis of Heart Disease Prediction:**

It will begin from the principal segment and investigate every section and comprehend what influence it makes on the objective segment. At the necessary step, we will likewise perform preprocessing and include designing undertakings. The point in acting top to bottom exploratory examination is to get ready and clean information for better demonstrating to accomplish elite execution and summed up models. So it should begin with breaking down and setting up the dataset for expectation.

**Modules:**

1) Dataset collection

2) Data cleaning

3) Exploratory Data Analysis

4) Machine learning Modeling

5) Report

1) Dataset collection:

The information about the heart disease dataset with different types of attributes are collected from different type of patients.

2) Data Cleaning:

The large dataset contains more noisy and improper data which have to be pre-processed to produce the quality dataset for further pruning. The data is cleaned and processed with initial stage of removing the null values.

3) Exploratory Data Analysis

Exploratory analysis is a process to explore and understand the data and data relationship in complete depth so that it makes feature engineering and machine learning modeling steps smooth and streamlined for prediction. EDA helps to prove our assumptions true or false. In other words, it helps to perform hypothesis testing.

4) Machine Learning Modeling

Machine Learning modeling helps to find the best algorithm with the best hyperparameters to achieve maximum accuracy. The dataset is split into 2 variants. 70% of records are taken as training data and used to train the machine learning algorithm. The remaining 30% of dataset is applied to testing which helps to predict the process.

5) Report:

The Data is visualized based on the output of the Machine Learning algorithm and the data is mapped with different types of graphs to analyze and visualize the exact data to the user for the prediction. Marplot libraries are implemented to map the results based on the user requirements.

**SYSTEM SPECIFICATION**

**HARDWARE REQUIREMENTS**

* Processor Intel(R) Pentium(R) CPU G2010 @
* Clock Speed 2.80GHz
* RAM 2.00 GB
* Hard Disk 1 TB HDD
* Monitor 15.6 Inches
* Mouse Logitech B100 Wired Optical Mouse
* Keyboard Full-size island-style keyboard with number keypad
* Display Card Super Video Graphics Adapter

**SOFTWARE REQUIREMENTS**

* Operating System : Windows 10
* Front-End Tool : Python in Google Colab

**SYSTEM ARCHITECTURE DIAGRAM**

Heart Disease prediction

Training, test data

Data Cleaning

Data preprocess

ML Algorithm implementation

Heart disease dataset

ARCHITECTURE DIAGRAM

The heart disease dataset is given as input to the application and the pre-processing is applied , next the data cleaning is performed after the training and test data are splitted down and passed into machine algorithm and the heart disease prediction is done.

**Load Packages:**

First step have to import the necessary packages to the application:

//--- Libray files

from keras.models import Sequential

from keras.layers import Dense

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

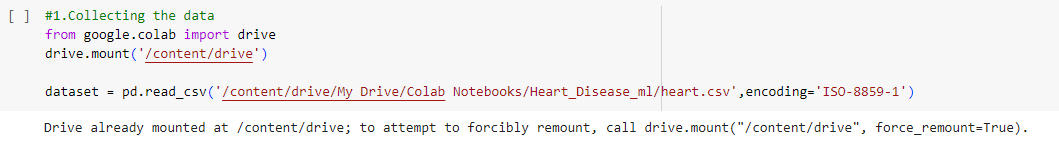
import os

print(os.listdir())

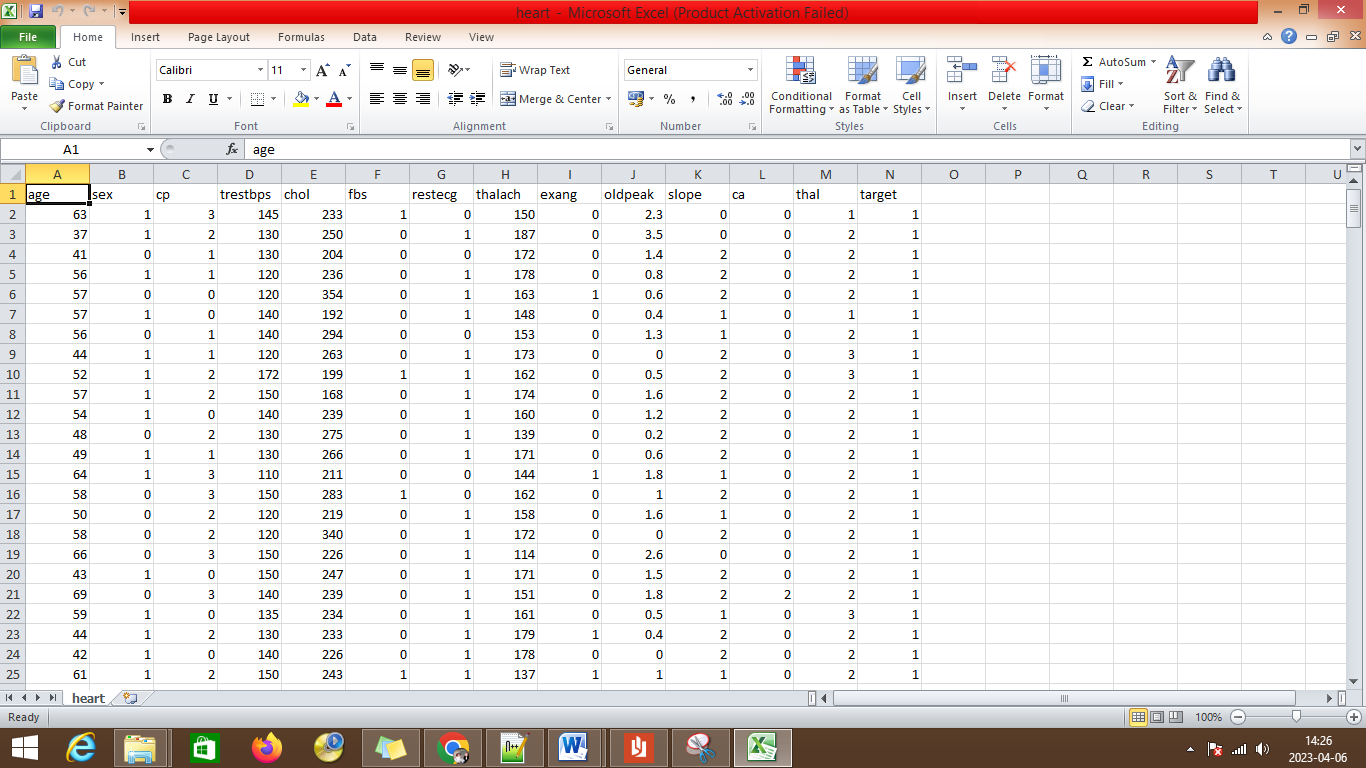
import warnings

warnings.filter warnings('ignore')

Next the dataset would be connected from the Google colab. Initially the dataset is uploaded into the Google colab folder. Then the Python file should connect to the path from the Google colab folder.

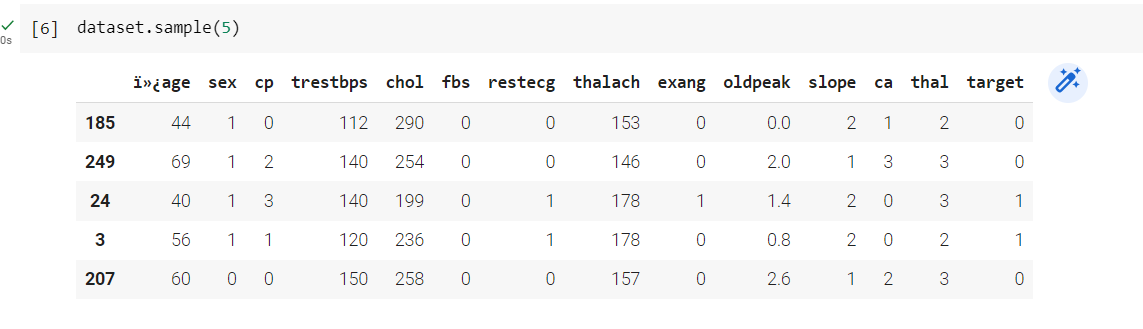
****

The information has an extremely straightforward design with elements. Each column is related with the heart disease attributes.



The label has been set with different label description features of:

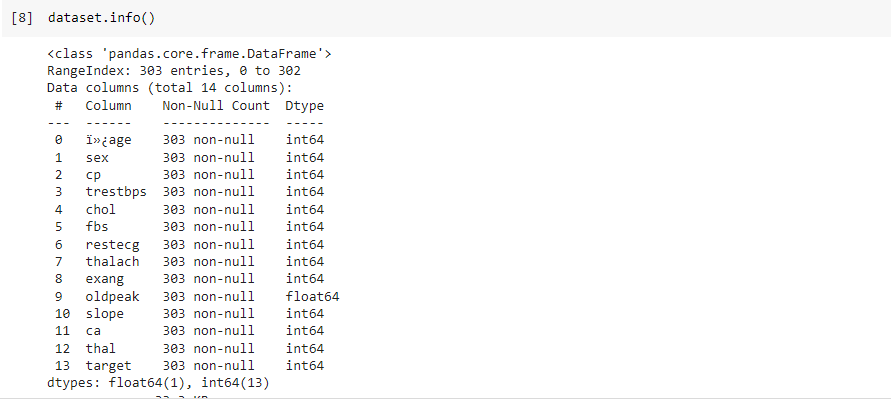
* age
* sex
* cp
* trestbps
* cholesterol
* fbs
* restecg
* oldpeak
* slope
* ca
* thal
* target



The dictionary shows the records displayed with head values of first 5 records from the dataset. The bar plot can be used to address the heart disease attributes.

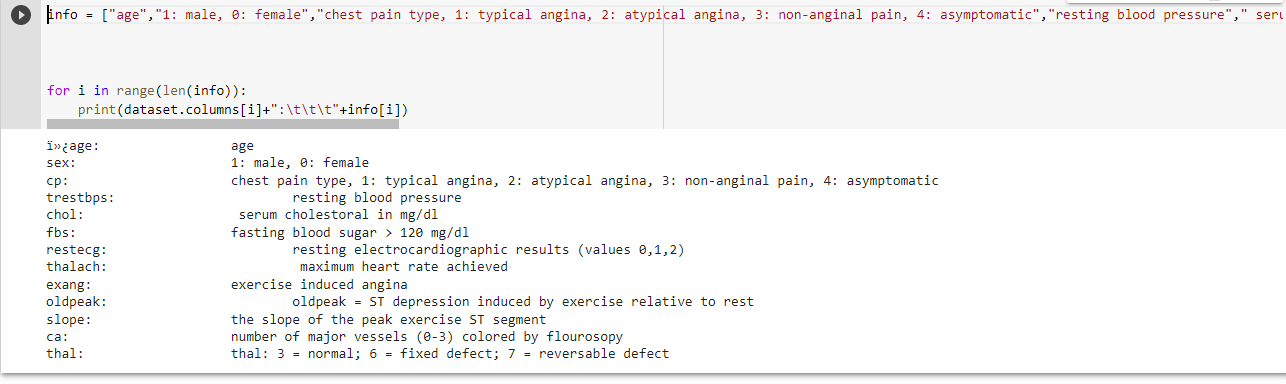


The describe function describes the coun, mean,max,min statistical reports for the coumns of age,sex,cp, chol and other fields.



The information about the data is displayed with the data type.

**Attributes Exploration:**



The heart disease dataset attributes are explored for further analyzing.

Below are the values:

sex: 1: male, 0: female

cp: chest pain type, 1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic

trestbps: resting blood pressure

chol: serum cholestoral in mg/dl

fbs: fasting blood sugar > 120 mg/dl

restecg: resting electrocardiographic results (values 0,1,2)

thalach: maximum heart rate achieved

exang: exercise induced angina

oldpeak: oldpeak = ST depression induced by exercise relative to rest

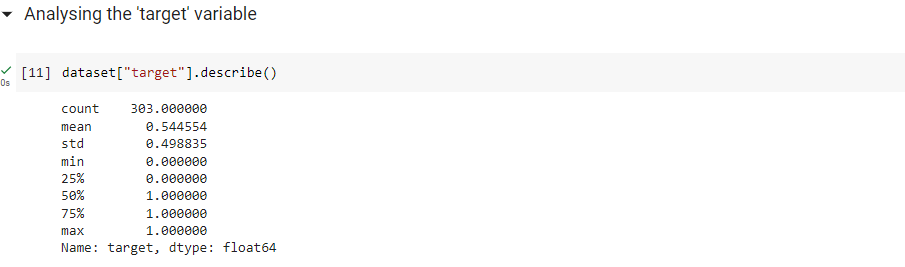
slope: the slope of the peak exercise ST segment

ca: number of major vessels (0-3) colored by flourosopy

thal: thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

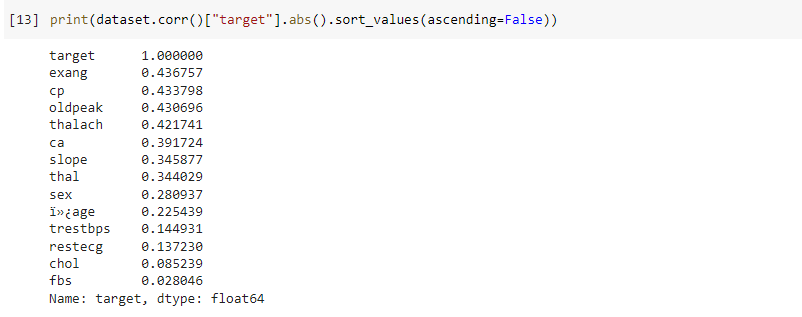
**Analysing :**

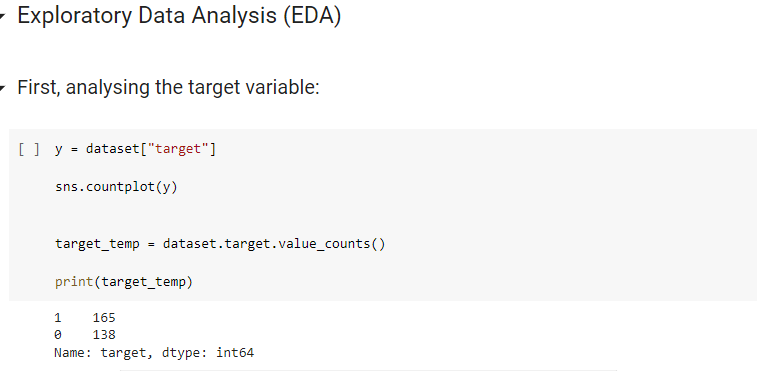
The target variables are analyzed



**Corrlelation of data:**

The correlation of data with respect to the attributes are linked and displayed.





The target variable is assigned to the y variable.

**Implemenation with Algorithms:**

The testing and training variables are split and passed into the algorithm for the heart disease prediction.



The testing size is 20% and the training size is 80% of the dataset.

**Evaluation with Logistric Regression Algorithm:**

Using logistic regression, the outcome of a dependent variable is predicted. As a consequence, the outcome should have a distinct or downright worth. In general, it tends to be either Yes or No, 0 or 1, and so on, but instead of providing a precise value between 0 and 1, it provides probabilistic attributes that fall between 0 and 1.

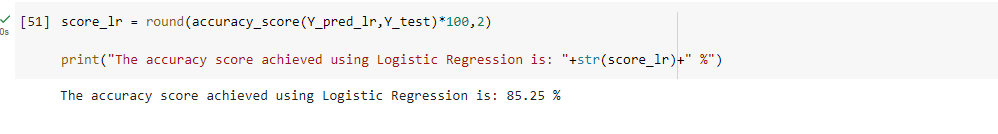
With the exception of how they are applied, linear regression and logistic regression are very similar. Regression concerns are dealt with using Direct Relapse, whereas arrangement difficulties are dealt with using Calculated Relapse.

Instead of fitting a relapse line in Calculated Relapse, we fit a "S" shaped strategic capacity, which forecasts the two best characteristics (0 or 1).



**Results:**

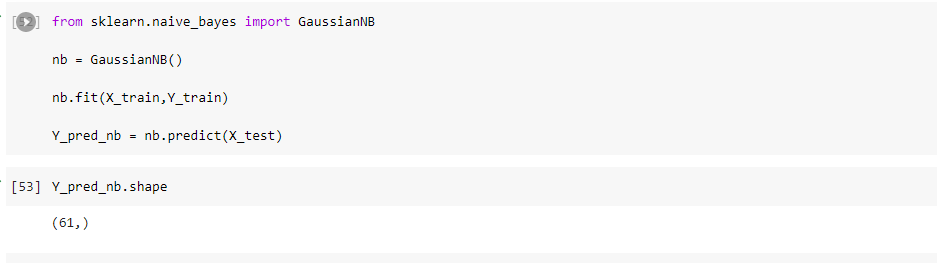
The results of the logistic regressin algorithm shows the accuracy of 85.25% with good performance in prediction.



Evaluation with Naive Bayes:

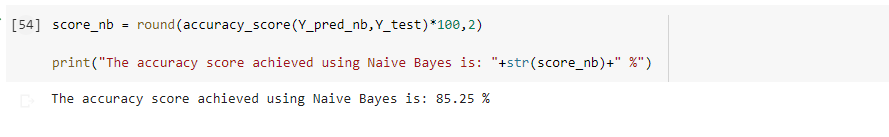
Naive Bayes is an AI algorithm that is involved by information researchers for characterization. The naive Bayes works in view of the Bayes hypothesis. The Naive Bayes is a grouping calculation that depends on Bayes' hypothesis, which is an approach to computing the likelihood of an occasion in light of its earlier information. The calculation is designated "innocent" since it makes an improving on suspicion that the highlights are restrictively free of one another given the class mark.

The Naive Bayes calculation can be utilized for twofold as well as multi-class order issues. It is normally utilized in message characterization errands, for example, spam sifting or opinion examination, however it can likewise be utilized in different applications where there are numerous classes and various highlights.

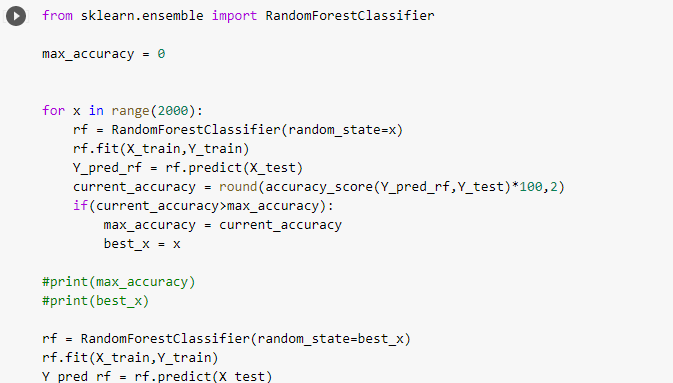


Results:

The results of the Naive Bayes algorithm shows the accuracy of 85.25% with good performance in prediction.

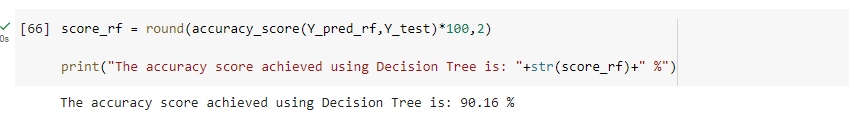


**Evaluation with Random Forest Algorithm:**

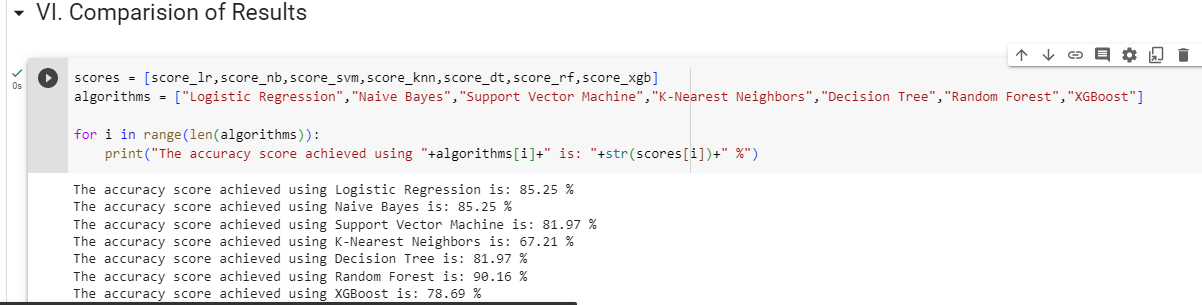


The random forest algorithm helps in prediction of the heart disease with the available dataset.

Results:



The results of the random forest algorithm shows the accuracy of 90% with good performance in prediction.



**Conclusion and Future Enhancement:**

The heart disease with different types of algorithms is compared and the accuracy levels were identified and compared. The Machine learning algorithm provides better accuracy levels for the prediction. In the future the large dataset can be used for the prediction which will help to increase the efficiency of the algorithm in predictions.