Lab Project Submission

Project Title: Heart Disease Prediction

Student's Name:

Showmen Sarker (19301188)

Sadman Fardin   (19301068)

Golam Sifat   (20301478)

Department: Computer Science and Engineering

Course Title: Artificial Intelligence

Lab Section: 06

Submitted To:

Nazibur Rahman

Sumaiya Akter

Date of Submission: 18-12-2022

Introduction:

Machine learning to predict heart disease is the name of our project. We used some pre-processing techniques to get our dataset ready for use. A dataset needs to be pre-processed because it might have some data that isn't needed or some null values. By preprocessing a dataset, we can get a good accuracy score and make sure our project runs smoothly. To get the accuracy score, we used about three algorithms or models.

Dataset description :

There are about 14 columns and 217 rows. There is one object type column. Others are float and integer type column. We have 13 features in our dataset which includes Age, sex, chest pain, BP, cholesterol, FBS over 120,Max HR, exercise angina, slope of ST,ST depression, EKG results Number of vessels fluro, Thallium. The target variable is Heart disease which tells us whether a person has heart disease or not.

Pre processing techniques:

• NULL VALUE:
First, we checked to see if the dataset had any null values. In our dataset, there were no null values. Because of this, we don't need to remove or add any columns or rows or add any values to the dataset.

• Feature scaling:
We scaled the features of our dataset since we discovered that certain columns included enormous numbers and others contained modest numbers. This is why we scaled features. Using these methods, we can obtain all values within a comparable border level. To do this, MIN MAX scalar was used. Min max scalar method can accept all numbers between 0 and 1.

• Categorical features:
We performed categorical feature encoding since one column included string values while the rest contained integer or float values. This is why we did it. We utilized the level encoder and map function.

Model Implementation:

- **LOGISTIC REGRESSION:**
  One of the machine learning algorithms that goes best with our dataset is logistic regression. We know that is it is a statistical analysis method through which we can predict binary outcome. In our dataset, the labels which will be predicted is also a binary outcome. A person has heart disease or not is a kind of binary outcome which could be predicted through this algorithm.

- **DECISION TREE:**
  The second algorithm that we used is decision tree. It is a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered. Our heart disease prediction problem is a classification problem. So, decision tree is a suitable algorithm for classification based problems.

- **RANDOM FOREST:**
  The third algorithm that we used to train our model is random forest. It is also a supervised learning algorithm which can be used for both classification and regression problems. It works by building decision trees on different samples and takes their majority vote for classification and average in case of regression. It gave better accuracy for our dataset.

ACCURACY:

Random forest: 0.85

Logistic regression: 0.85

Decision tree: 0.76

References:

Damarla, R. (2020)Heart disease prediction, Kaggle. Available at:
https://www.kaggle.com/datasets/rishidamarla/heart-disease-prediction
(Accessed: December 12, 2022).