IMLA Mini Project: Cardiovascular Disease Prediction

Problem Statement

Various studies are present which focuses on heart disease prediction where the diagnosis is done by using different techniques such as data mining, analysis, machine learning, Heart disease prediction can be done using machine learning where the required parameters are Age, Sex, Blood Pressure, Heart Rate, Diabetes, Hyper cholesterol.

Abstract

Cardiovascular diseases are one of the most vital causes of fatality. Cardiovascular disease prediction is a critical challenge in the area of clinical data analysis. Machine learning and Neural Networks are more promising in assisting decide and predict from the massive data produced by healthcare. We have noted different features had used in recent developments of the machine learning model. In this paper, we proposed machine learning techniques to predict cardiovascular disease using features. BMI is one of the highlighting features we used for prediction. BMI is important in predicting cardiovascular disease.

Methodology

Feature Selection and Adding a new feature:
BMI is computed by using the height data and weight data for each patient.   Formula: $[bmi] = [weight(kg)] / [height(m)]^2$

Experimental setup for evaluation:
1. The dataset is splitted into an 80:20 ratio.
2. The training set size is 80%, and the testing set size is 20% of the entire dataset.
3. The first model (model A) was developed with the given dataset and did not make any changes
4. For Second model (model B), a new column BMI was added.

Classification Models

1. Logistic Regression
2. K-Nearest Neighbors
3. Naive Bayes Classifier
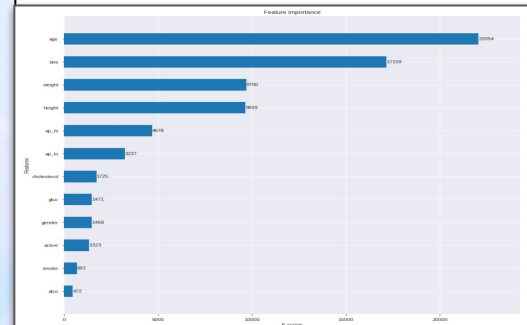4. Decision Tree Classifier
5. XGB Classifier

Proposed Method
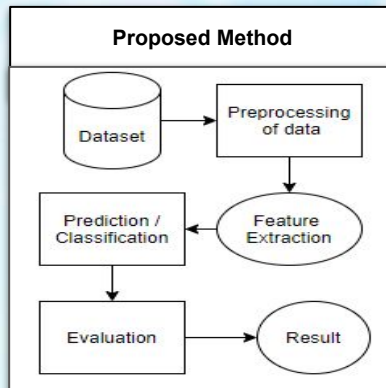
Accuracy

	Model	Score_train	Score_test
4	XGBClassifier	78.459755	73.014493
3	Decision Tree Classifier	72.811102	72.637681
1	k-Nearest Neighbors	72.845523	71.144928
2	Naive Bayes	70.791138	70.376812
0	Logistic Regression	62.776500	62.992754

Feature Importance

Conclusion

Heart disease prediction is essential as well as challenging work in the medical Field. The mortality rate can be reduced if the disease is recognized at the initial stages, and precautions and proper treatment are possible.
XGB classifier used to identify the importance of each feature in the prognostication of heart disease.
Adding feature BMI improved the accuracy of prediction. Thus, by assessing the results, the suggested approach generates a more precise prediction of cardiovascular diseases.
We conclude that BMI is a significant factor while predicting cardiovascular disease.

Presenters

1. PA-10   1032180297   Atharv Nikam
2. PE-11   1032180099   Aditya Mhaske
3. PE-23   1032180464   Vipul Dunde