

Heart Attack Prediction

Ganpat University, ICT

Name: Javal patel, Jagir Shastri, Manav Mehta



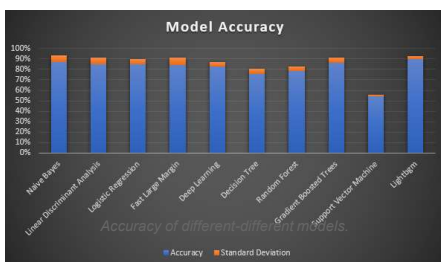
Introduction

Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used 5 different algorithms of machine learning to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by 99.6% accuracy.

Methods

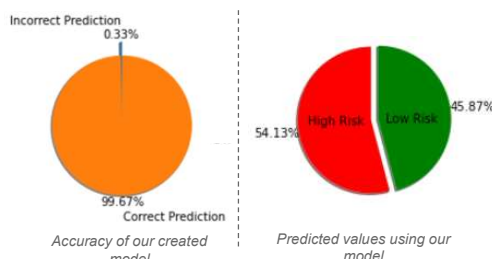
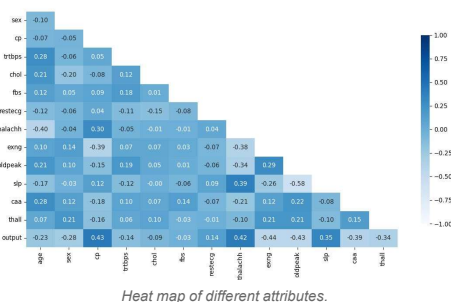
Firstly, we chose to carry out our own primary research to complement pre-existing papers on this topic. Both quantitative and qualitative surveys were implemented to allow for better data representation on this topic. After that we checked over 10 methods for better accuracy and finalized Gaussian Naive Bayes, Random Forest Classifier, Linear Discriminant Analysis, Logistic Regression and Light Gradient Boosting Machine.

Then we short listed 5 models with high accuracy and good reliability. Following, we developed our own algorithm to combine all this models.



Data Analysis

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyper lipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help



By applying different machine learning algorithms and then using deep learning to see what difference comes when it is applied to the data, three approaches were used. In the first approach, normal dataset which is acquired is directly used for classification, and in the second approach, the data with feature selection are taken care of and there is no outliers detection. The results which are achieved are quite promising and then in the third approach the dataset was normalized taking care of the outliers and feature selection; the results achieved are much better than the previous techniques (we have 99.67% accuracy), and when compared with other research accuracies, our results are quite promising.

Results

In order to predict the probability of patients having heart disease, a confusion matrix (Table 1) was created, where A denotes patients with heart disease, and B denotes patients with no heart disease. Also we have achieved 99.67% precision. Team also identified a need to work together with other players in medical field to diversify our strategies and generate a greater impact. The creation of local and more wide-spread networking organizations focused on this theme of heart attack prediction is filling this need and gaining momentum in the medical field.

Conclusion

The conclusion which we found is that machine learning algorithms performed better in this analysis. The computational time was also reduced which is helpful when deploying a model. It was also found out that the dataset should be normalized; otherwise, the training model gets overfitted sometimes and the accuracy achieved is not sufficient when a model is evaluated for real-world data problems which can vary drastically to the dataset on which the model was trained.

References

1. Dataset <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>
2. UCI Machine Learning Repository, "Heart disease data set," 2020, <http://archive.ics.uci.edu/ml/datasets/heart+disease>
3. Random forest classifier <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
4. Lightgbm <https://lightgbm.readthedocs.io/en/latest/Python-API.html>
5. Gaussian Naïve Bays http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html
6. Logistic Regression http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
7. Linear Discriminant Analysis http://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html

Contact Information

Javal Patel javalpatel19@gnu.ac.in
Jagir Shastri shastrijagir19@gnu.ac.in
Manav Mehta manavmehta19@gnu.ac.in