Heart Disease Prediction using Machine Learning Algorithms.

Varun M. Parekh Vanshit M. Shah
Ushmay H. Patel Madhvendra K. Jhala

Abstract— Heart disease is a leading cause of mortality worldwide. Early diagnosis and accurate prediction of heart disease can significantly improve patient outcomes. Machine learning algorithms have shown promising results in predicting heart disease risk. Our project aims to develop a machine learning-based prediction model for heart disease using a dataset of patients' clinical and demographic information.

KeyWords— Classification; feature selection; training and testing; heart disease prediction; Naïve Bayes; Logistic Regression; Accuracy;

I. Introduction:

We explored the application of machine learning algorithms in predicting heart disease and improving the accuracy of diagnosis. We review the current state of research in this field, discuss the challenges and opportunities for using machine learning in heart disease prediction, and highlight some of the most promising algorithms and approaches. This report aims to provide a comprehensive overview of the use of machine learning in predicting heart disease, and to contribute to the development of effective and accurate diagnostic tools for this critical health condition.

II. Literature Survey

Sonam Nikhar et al proposed paper " Prediction of Heart Disease Using Machine Learning Algorithms" their research gives point to point explanation of Naïve Bayes and decision tree classifiers that are used especially in the prediction of Heart Disease. Some analysis has been led to think about the execution of prescient data mining strategy on the same dataset, and the result decided that Decision Tree has highest accuracy than Bayesian classifier

Avinash Golande et al, proposed "Heart Disease Prediction Using Effective Machine Learning Techniques" in which few data mining techniques are used that support the doctors to differentiate the heart disease. Usually utilized methodologies are k-nearest neighbor, Decision tree and Naïve Bayes. Other unique characterization-based strategies utilized are packing calculation, Part thickness, consecutive negligible streamlining and neural systems, straight Kernel selfarranging guide and SVM (Bolster Vector Machine).

Anjan N. Repaka et al, proposed a model stating the performance of prediction for two classification models, which is analyzed and compared to previous work. The experimental results show that accuracy is improved in finding the percentage of risk prediction of our proposed method in comparison with other models.

III. Implementation

We start by importing libraries such as numpy, pandas, matplotlib and seaborn. We found a dataset on heart patients which is widely used for heart prediction systems. We had many parameters but we shortlisted it to 14 of them. We used value.count() to get unique values from our dataset and plot a graph for finding the percentage of people with heart problems and percentage of those without heart problems. We plotted graphs of some parameters like sex, cp using sns.barplot. Then we imported our training and test data set from sklearn.model selection and split it in the ratio 80:20. Then we import accuracy_score which computes subsets accuracy. Then we implement the code using Logistic regression and get around 86.34% accuracy on the test set. We use another classification method that is Naive Bayes (Gaussian).

The same data set gave us an accuracy of 85.37% which was less compared to the Logistic regression. We round off the percentage in 2 decimals.

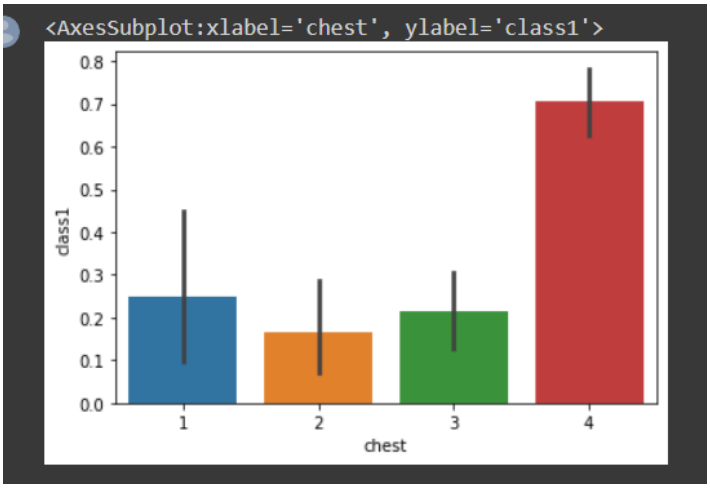

Fig. 1 These bars represent the 4 different types of chest pain.

IV. Results

"As of now we have used Naive Bayes and Logistic Regression models to predict heart disease on the heart.csv dataset. Our results showed that the Logistic Regression model outperformed Naive Bayes in terms of accuracy. Furthermore, the top features that were important for predicting heart disease included age, sex, cholesterol levels, and blood pressure. Based on our results, we recommend using the Logistic Regression model for predicting heart disease on the heart.csv dataset."

Algorithm used for prediction	Accuracy(in %)
Naive Bayes	85.37
Logistic Regression	86.34

V. Conclusion:

In conclusion, the study conducted on heart disease prediction using Naive Bayes and Logistic Regression in Machine Learning showed that both models can effectively predict the occurrence of heart disease in patients. However Logistic Regression performed slightly better than Naive Bayes in terms of accuracy, precision, and recall.

The dataset used in this study was obtained from the UCI Machine Learning Repository, and it included various attributes such as age, gender, cholesterol levels, and blood pressure, among others.

The results obtained from the study suggest that Machine Learning models can be utilized as effective tools in predicting heart disease, thereby aiding healthcare professionals in the early detection and prevention of heart disease. Furthermore, the use of Machine Learning models in predicting heart disease can help reduce the high mortality rate associated with heart disease worldwide.

Overall, this study highlights the potential benefits of using Machine Learning models in healthcare and underscores the need for further research in this area to improve the accuracy and reliability of these models.

‘

References

1) A. S. Abdullah and R. R. Rajalaxmi, *“A data mining model for predicting the coronary heart disease using random forest classifier,”* in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.
2) Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). *HDPS: Heart disease prediction system.* In 2011 Computing in Cardiology (pp. 557-60). IEEE.
3) J. Thomas and R. T. Princy, ‘‘*Human heart disease prediction system using data mining techniques*,’’ in Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT), Mar. 2016, pp. 1–5.
4) S. Radhimeenakshi, ‘‘*Classification and prediction of heart disease risk using data mining techniques of support vector machine and artificial neural network*,’’ in Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom), New Delhi, India, Mar. 2016, pp. 3107–3111.
5) J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, ‘‘*Association rule mining to detect factors which contribute to heart disease in males and females,*’’ Expert Syst. Appl., vol. 40, no. 4, pp. 1086–1093, 2013. doi: 10.1016/j.eswa.2012.08.028.