Heart Disease Prediction Using Machine Lerning

Anupama Yadav, Levish Gediya[1], Adnanuddin Kazi[2]

[1]Assistant professor, Dept. of Computer Science & Engineering, Parul University, Gujarat, India
[2]UG student, Dept. of Computer Science & Engineering, Parul University, Gujarat, India

---***---

Abstract - *This paper studies the prediction of heart disease using machine learning algorithms and python programming. Over the past decades, heart disease is a common and dangerous disease caused by fat suppression. This disease occurs due to overpressure in the human body. We can predict cardiac disease using a variety of parameters in the dataset. We observed a dataset cover of 12 parameters and 70000 different data values to evaluate patient performance. The major objective of this paper is to obtain improved accuracy for detecting heart disease using algorithms in which the target output calculates whether the person has heart disease or not.*

*Key Words***: Heart Disease, symptoms, Machine Learning, Prediction, Health**

1.INTRODUCTION

The core topic is prediction using machine learning techniques. Machine learning is extensively used nowadays in numerous business applications like e-commerce and many more. Prediction is one of an area where this machine learning used, our subject is about the prediction of heart disease by processing patient's dataset and a data of patients to whom we need to predict the chance of occurrence of heart disease. Python is the most usual programming language having several libraries which are used in this project with a machine learning model. Machine learning is a subset model of an artificial intelligence network in which uses complex algorithms and deep learning neural networks. The human body is made up of numerous organs, all of which have their functions. The heart is one such organ that pumps blood all over the body and if it does not do so, the human body can have fatal conditions. One of the key reasons for mortality today is having heart disease. Therefore, it becomes necessary to ensure that our cardiovascular system or any other system in the human body for that matter should remain healthy. Unfortunately, people around the world are suffering from cardiovascular diseases. Any technology that can help detect these diseases before much damage is done will shows that it is helpful in saving people's money and further significantly their life. Data mining techniques can be beneficial in predicting heart diseases. Analytical models can be made by finding earlier unknown patterns and trends in databases and using the obtained information. Machine learning is a technology that can help to accomplish a diagnosis of heart disease before much damage occurs to a person. As a developing field in science and technology, machine learning can categorize whether a person might be suffering from heart disease or not.

2.LITERATURE REVIEW

- Heart Disease prediction by using Naive Bayes Approach, Artificial Intelligence (AI) networks, Supports Vector Machine (SVM), Random forest algorithm, Simple regression method (Shadman Nashif).
- The paper represented and it is equated with KNN, SVM, Random classifier, decision tree classifier had given a precise result for Heart Disease Prediction System- HDPS. With the help of this the prediction was made with the improved accuracy of 98.85% by decision tree machine learning method as compare to other methods (O.E. Taylor).
- Improved data mining techniques when predicting heart disease (Animesh Hazra). In this paper, c4.5, Kmeans, decision tree, SVM, naïve Bayes, and all additional machine learning algorithms are compared to get an improved accuracy of heart disease.
- Instead, Praveen Kumar Reddy, 2019, attempt to decrease the occurrences of heart disease using a decision tree algorithm. In this, the Support Vector Machine algorithm categorizes the data values by using hyperplane, and the decision tree is applied by the Gini index method in which the maximum gain of the attributes gives an improved representation of the decision tree algorithm.
- M.A.Nishara Banu and B.Gomathy both had given a paper by the name Disease Predicting system using data mining techniques. In this paper, they had discussed about MAFIA (Maximal Frequent Itemset algorithm) and K-Means clustering. As classification is significant for the prediction of disease. The classification is based on MAFIA and K-Means results imprecision.
- Both Wiharto and Hari Kusnanto had presented a paper called The Intelligence System for the Diagnosis Level of Coronary Heart Disease with K-Star Algorithm. In this paper, they outline the expectation for heart infection using the Learning Vector Quantification Neural System Calculation. The neural system in this structure identifies 13 clinical

including information and predicts whether the patient has near or absent coronary disease, with various execution measures.

3.Methodology and Analysis

- **Data collection**

The general process of predicting heart disease conveys the following method:

1) Cardiovascular Disease Dataset by Svetlana Ulianova, 2019. The dataset contains 70,000 records of patient data made up of 11 features and the dataset is the information or tool that is important for any type of study or project.

- **Data Preprocessing**

1) Separation of target data and feature data as training and test data.

2) Scrambling the values in the data to be valued between 0 and 1 and scale all the values before training the Machine Learning models.

- **Applying Algorithms**

1) Equating 4-machine learning algorithms such as SVM, Decision tree, Random forest classifier, and K- nearest neighbor to get the improved accurateness to which highest parameter may cause disease.

2) For every algorithm, there is a pseudo code helpful to grow any kind of programming language. In python, there is a simple way to create any kind of algorithm in which simple and shortcode is easier to predict accuracy.

3.1 MACHINE LEARNING ALGORITHM :

The algorithms used in this project are highly helpful to predict the correct result to detect heart disease in which factors that cause disease can be detected. The following algorithms have been assembled in this project.

1) K-Nearest Neighbor algorithm:

KNN is a managed classifier that carries out an observation from within a test set to predict classification labels. KNN is one of the classification methods used whenever there is a classification. It has a several assumptions includes the dataset has slight noise, is labeled and it should covers relevant features. By applying KNN in huge datasets takes a long time to process. The accuracy grown with this algorithm is 63.4%.

2) Random Forest Classier:

A random forest classifier is a powerful tool in the machine learning library. With this classifier, we will be able to increase accuracy, and training time should be a smaller amount. Primarily, we have to build a model and by splitting variables into training and test set. After that, train the dependent variables and predict the answer. By using the random forest classifier, the accuracy predicted result is approximately 71% but actually it is 71.4%.

3) Decision tree classier:

In this algorithm, pre-processing is initially created by dividing the data into training and test data. Feature scaling can be done by normalizing the values before the forecast. Import decision tree classification into training sets of determined and independent variables using Guinea-index criteria to predict the accuracy or response of a test set. The accuracy obtained with this algorithm is 68.4%.

4) Support Vector Machine (SVM):

SVM is also one of the classification algorithms in machine learning in which improved accuracy can be predicted. As compare to other algorithms, it is much better for expectedly predicting accuracy.

In our prediction, the predicted highest accuracy is 72.6% using linear SVM kernel.

In our prediction, the predicted highest accuracy is 86.2% using the Gaussian SVM kernel.

4.PROPOSED SYSTEM
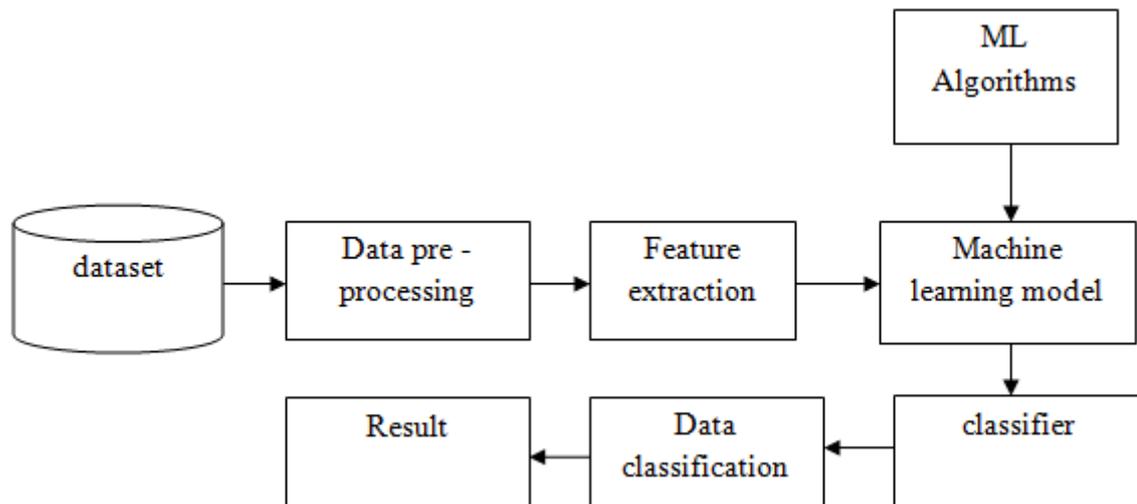
By this experiment whatever we say is as Naïve Bayes results and decision tree results may change so for each prediction we do not have a comparison of both the algorithms to get precise results and in the same way if we use only a single algorithm which cannot pre-process data we even can't get good accuracy so its better to have a combination of algorithms like k-means, ID3 and k-means, and Naïve Bayes.

After confirming that the data is balanced, the connection between the data is found out and is plotted as a heat map using the Seaborn library.

The heat map clearly displays that the attributes like cp (chest pain) and thalack (maximum heart rate achieved) have a positive connection with the target attribute. Now that the connection has been checked, we need to convert categorical variables like sex, cp, FBS, restecg, exang, slope, ca, and thal into dummy variables. This can be done by using the get_dummies method of the Pandas library. After creating dummy variables, the data in columns like age, trestbps, chol, thali, and old peak needs to be standard scaled, because they have much varied quantities and units. This can be done using the Scikit-learn library in Python.

5.CONCLUSIONS

In this what we found is during small datasets in some other cases most of the time decision trees direct us to a solution that is not accurate, but if we see Naïve Bayes results we are getting more precise results with chances of all additional possibilities but due to guidance to only a single solution decision trees may miss lead. Finally, we can tell by this experiment that Naïve Bayes is more accurate even if the input data is clean and well maintained even though ID3 can clean itself every time it gives an accurate result, and similarly Naïve Bayes is accurate. Will not give result. Each time we need to consider the results of different algorithms and with all its results it would be more accurate if predicted. But we can use Naïve Bayes to consider variables as individuals we can use a combination of algorithms like Naïve Bayes and K-means to get accuracy. After applying various algorithms, it can be said that machine learning is proving to be extremely valuable in predicting heart disease which is one of the most prominent problems of society in today's world. As more and more work is being done in the field of machine learning, new methods may soon come to make machine learning more helpful in the field of healthcare. The algorithms used in this experiment have performed well using the available attributes. The conclusion can be finally drawn that machine learning can reduce the damage done to a person physically and mentally, by predicting heart disease.

REFERENCES

[1] Patel, J., Tejal Upadhyay, D., and Patel, S. (2015). Predicting heart condition using machine learning and data mining techniques. Heart disease, 7 (1), 129-137.

[2] Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), 684- 687.

[3] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Intelligent and effective heart disease prediction system using weighted associative classifiers. International Journal on Computer Science and Engineering, 3(6), 2385-2392.

[4] Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dhavachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.

[5] Alapatt, B.P., Kavitha, A., Amudhavel, J., "A novel encryption algorithm for an end to end secured fiber optic communication", (2017) International Journal of Pure and Applied Mathematics, 117 (19 Special Issue), pp. 269-275.

[6] Amudhavel, J., Inbavalli, P., Bhuvaneswari, B., Anandaraj, B., Vengattaraman, T., Premkumar, K., "An effective analysis on harmony search optimization approaches", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2035-2038.

[7] Amudhavel, J., Kathavate, P., Reddy, L.S.S., Bhuvaneswari Aadharshini, A., "Assessment on authentication mechanisms in a distributed system: A case study", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1437-1448.

[8] Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattatraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2058-2061.

[9] Amudhavel, J., Kathavate, P., Reddy, L.S.S., Satyanarayana, K.V.V., "Effects, challenges, opportunities and analysis on security-based cloud resource virtualization", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1458-1463.

[10] Svetlana Ulianova, Cardiovascular Disease dataset. The dataset consists of 70 000 records of patient data, 11 features + target.

[11] Kaur, B., & Singh, W. (2014). Review on heart disease prediction system using data mining techniques. International journal on recent and innovation trends in computing and communication, 2(10), 3003-3008.

BIOGRAPHIES

Anupama Yadav
Assistant Professor
Computer Science & Engineering
Parul University

Levish Gediya
Student
Computer Science & Engineering
Parul University

Adnanuddin Kazi
Student
Computer Science & Engineering
Parul University