International Journal on Applications in Information and Communication Engineering
Volume 8 : Issue 2 : June 2022, pp 32 – 36 www.aetsjournal.com ISSN (Online) : 2394 - 6237

PREDICTION OF HEART DISEASE USING DATA WAREHOUSE

KALIYAMOORTHI P, MUNIKEERTHANA K , FARHEEN TAJ A , MURALI KRISHNA P , MADHUKAR NAIK S

Abstract — We suggest in this study that a system be developed to determine whether or not a patient has cardiac disease. The data warehouse aids in the correlation of clinical and financial records so that the cost-effectiveness of care can be estimated. Data mining techniques aid in the identification of data trends that may help predict future individual cardiac problems. Heart disease prediction necessitates a large amount of data that is too complex and enormous to collect and analyze using traditional methods. Our goal is to develop the most appropriate data mining algorithms for predicting cardiac disease that are both computationally efficient and accurate. Two data mining classification techniques, Decision Tree and KNN, are addressed in this study and used to construct a prediction system to analyse and predict the outcome. When compared to various algorithms for heart disease prediction, it was discovered that the Decision tree method works best with 91 percent precision, is computationally efficient, and is accurate.

Keywords— Data mining; decision tree classifier; logistic regression; k- nearest neighbour.

I. INTRODUCTION

A recision, is computationally efficient, and is accurate. person's life experiences determine the health of their heart, which is fully based on professional and personal behaviors. There may also be a number of hereditary factors that pass a type of heart disease down from generation to generation. More than 12 million people die each year as a result of various types of heart disease, often known as cardiovascular disease, according to the World Health Organization. The term "heart disease" refers to a variety of ailments that affect the heart and arteries of a person. The risk of acquiring heart disease has risen. Obesity, poor diet, family history, high blood pressure, high cholesterol, idle behavior, family history, smoking, and hypertension may all raise the risk of heart disease in young individuals.

The heart is a critical organ in our bodies, as we all know. When a person's heart stops working properly, it affects other body organs such as the brain, kidneys, and so on. Organs such as the brain suffer from insufficient blood circulation, and if the heart stops working completely, death occurs within minutes. The heart to function effectively is critical for survival.. Problems with the heart and its blood arteries are referred to as "heart disease." Heart disease is caused by a number of factors, including:

- Hypertension
- Physical inactivity
- Poor diet
- High Blood pressure
- High blood cholesterol
- Obesity
- Cholesterol
- Resting ECG
- Thal
- Peak exercise ST segment
- Diagnosis of heart disease
- Angina

II. LITERATUREREVIEW

In [1], According to Carlos Ordonez's paper "Improving Heart Disease Prediction Using Constrained Association Rules," the most difficult part is discovering and predicting heart disease association rules. The data studied in this paper includes medical records of persons with heart disease who have risk factor features, heart perfusion tests, and constricted arteries. 1) The

Kaliyamoorthi P, Assistant Professor , Department of CSE , Madanapalle Institute of Technology & Science, A.P., INDIA.
(Email: kaliyamoorthi@mits.ac.in)
Munikeerthana K , Department of CSE , Madanapalle Institute of Technology & Science, A.P., INDIA.
Farheen Taj A , Department of CSE , Madanapalle Institute of Technology & Science, A.P., INDIA.
Murali Krishna P, Department of CSE , Madanapalle Institute of Technology & Science, A.P., INDIA.
Madhukar Naik S Department of CSE , Madanapalle Institute of Technology & Science, A.P., INDIA.

qualities are only found on one side of the rule. 2) The rule splits the attributes into uninteresting categories. 3) Medical records of persons with heart disease control the number of qualities from the rule.. As indicated in the experiments, limiting the running duration further reduces the number of limits. The number of constraints of showing regulations has been considerably decreased by further reducing the running duration, as indicated in the trials.

In [2], Boleslaw Szymanski and colleagues published "Using Efficient Supnaova Kernel For Heart Disease Diagnosis." In this paper, we present a novel heuristic for computing sparse kernels in SUPANOVA. It was used to analyze a conventional Boston housing market dataset and to find heart disorders in the general population using a novel, The attractive field generated by the human heart is used to measure cardiac activity in a noninvasive manner. The results were correctly predicted 83.7 percent of the time, outperforming Support Vector Machine and similar kernel results..

In [3]. "Heart Disease Prediction System Using Data Mining Classification Techniques" was published by Chaitrai. This research looked at algorithms for predicting cardiac disease from a broader perspective. number of input variables. The method combines medical terminology like sex, blood pressure, cholesterol, and 13 other characteristics to assess the chance of a patient having heart disease. In this study, obesity and smoking were introduced as novel variables. The heart disease database is used to data mining algorithms such as Decision Trees, Nave Bayes, and neither Networks. On the basis of accuracy, the performance of several procedures is checked. The accuracy of Neural Networks, Decision Trees, and Naive Bayes percent, respectively..

In [4], SVM, and ANN, were advocated in [4]. (Beyene & Kamat, 2018). In 2018, a proposal for a Heart Disease Prediction System Using Data Mining Techniques was made (Beyene&Kamat). In healthcare facilities, WEKA software is utilised to deliver automatic illness diagnosis and service quality evaluationsand Decision Tree were among the algorithms utilised in the study. According to the analysis, when compared to other data mining approaches, SVM is more effective and accurate. According to Chala Beyene, data mining techniques could be utilised to forecast and assess the occurrence of cardiac illness. The primary goal is to predict the start of cardiac disease so that an early automatic diagnosis may be made with a quick turnaround time. The recommended methodology is equally important in healthcare organizations with professionals lacking in knowledge and experience. It looks at blood sugar and heart rate, as well as age and sex, to see if someone has heart disease.

In [5], "A Neuro-Fuzzy to Classification of ECG Signals for Ischemic Heart Disease Diagnosis" was published by Victor-Emil. The Fuzzy-Gaussian Neural Network (FGNN), a neuro-fuzzy classifier that can distinguish ECG data for Ischemic Heart Disease (IHD) diagnosis, was the subject of this study. i. the Discrete Cosine Transform (DCT) in addition to the FGNN; ii. Pattern classification for IHD diagnosis using the FGNN. The proposed neuro-fuzzy model for IHD diagnosis was built and tested in this work utilising an ECG database of 40 persons, 20 of whom had IHD and the remaining 20 were healthy. The best outcome was a perfect IHD recognition score of 100 percent.. The result is moving as much as using only one lead (V5) of ECG records as input data because existing diagnosis algorithms require a collection of 12 lead ECG signals.

In [6], Niti Guru et al. [9], "Decision Support System for Heart Disease Diagnosis Using Neural Network." The use of neural networks to predict heart disease, blood pressure, and sugar levels was advocated in this research. The recommended methodology is equally important in healthcare organisations with professionals lacking in knowledge and experience. It looks at blood sugar and heart rate, as well as age and sex, to see if someone has heart disease.

In [7],(A & Naik) recommended establishing a prediction system based on the patient's medical data to diagnose heart illness. The system was designed with 13 input quality risk factors in mind. To forecast heart illness, he employed k-

means and nave bayes. The purpose of this project is to create a system that provides diagnostics based on a historical cardiac database and the view of the themes.

The purpose of this project is to create a system that provides diagnostics based on a historical cardiac database. The system was created with 13 features in mind. Data mining techniques such as clustering and can be used to extract knowledge from databases. The Cleveland Heart Database included 43 attributes in its 300 records. Based on the values, this model will forecast if the patient has heart disease or not.

III. PROPOSED SYSTEM

The major purpose of the research is to create a heart disease prediction model that can be used to predict when heart disease will strike. Furthermore, the goal of this research is to figure out which categorization method is the most successful for recognising heart disease in patients. This research is backed up by a comparative study and analysis that compares and contrasts three classification algorithms. Despite the fact that these machine learning methods are widely utilized, heart disease prediction is a significant task that necessitates the highest level of precision. As a result, the three algorithms are assessed at various levels and using various methods. As a result, researchers and medical practitioners will have a greater grasp of the problem and possible solutions. The Bayes rule is with a strong assumption that the attributes are independent, given the class, to construct the Nave Bayes algorithm. Despite the fact that this criteria for independence is sometimes broken in practise, the accuracy of naive Bayes classification is generally competitive Each internal node represents a "test" .(for example, whether a coin flip would land heads or tails), each indicates the test's conclusion, and each node provides a class label."K-Nearest Neighbour" is abbreviated as . It's a machine learning algorithm that's supervised. Both classification and regression problems can be solved using this strategy.

- Analyze the general properties
- Import required libraries packages
- Load the given data set

A. Classification using decision tree :

The (DT) is a easy and simple classifier. Only Decision Trees have the ability to retrieve in-depth patient profiles via the bit through capability. The structure of a decision tree makes classification and models easier to debug and operate. Both category data can be handled by decision trees.. By determining the information gain of the qualities and deleting the attributes, the approach separates the branches into threes.

B. K-Nearest Neighbor(KNN) Algorithm:

The method is one of the most basic ML algorithms. It is based on the Supervised Learning technique. The K-NN algorithm thinks that the new case/data and existing cases. Despite the fact that these machine learning methods are widely utilized, heart disease prediction is a significant task that necessitates the highest level of precision. As a result, the three algorithms are assessed at various levels and using various methods. The ratio of the heart in the memberane is the ability that the nearest algorthim that which help us to slove and identify the disease which comes to us before. This the main one that we are using to predict the system in the responsible of the structure and images that are ehich the membrane.

The K-NN approach saves all available data and categorises new data points depending on how similar they are to the current data.

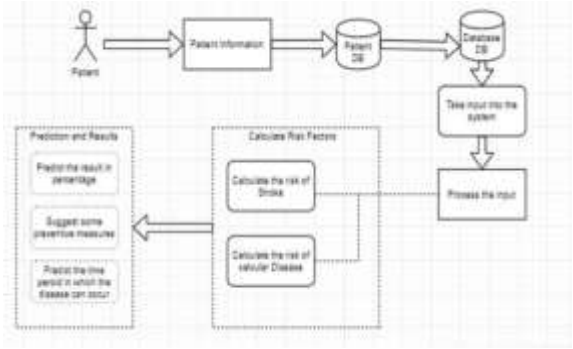

Figure 3.1: *Prediction of heart disease block diagram*

IV. RESULTS AND DISCUSSIONS

The Resting electrographic results i.e,. fig 2.

0 - venticular hypertrophy.

1 - normal

2 - ST devation or depression

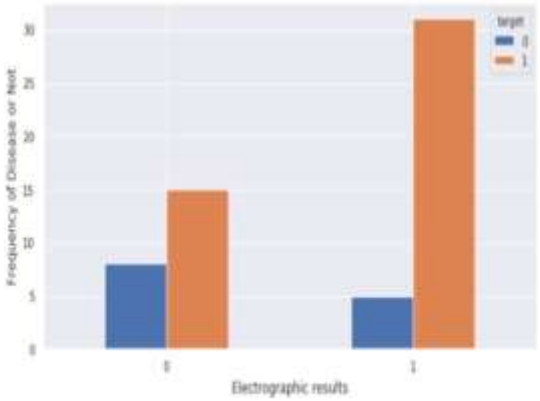

Figure 2: Electrographic Results

The Trestbps graph of a patient is as shown in below graph i.e., fig 3.

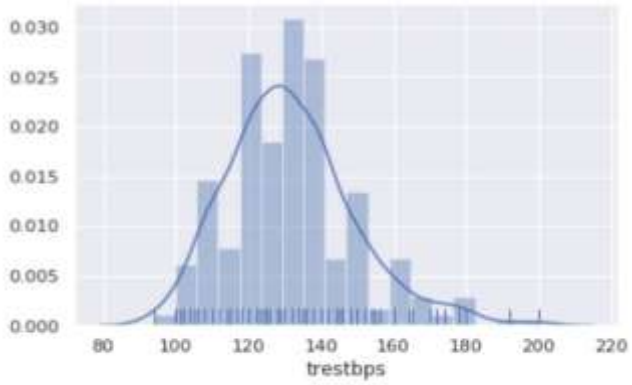

Figure 3: Trestbps Graph

Based on oldpeak data gender bias graph has plot on this paper and shown as fig 4.

The ST depression caused by exercise compared to rest is depicted in this graph.

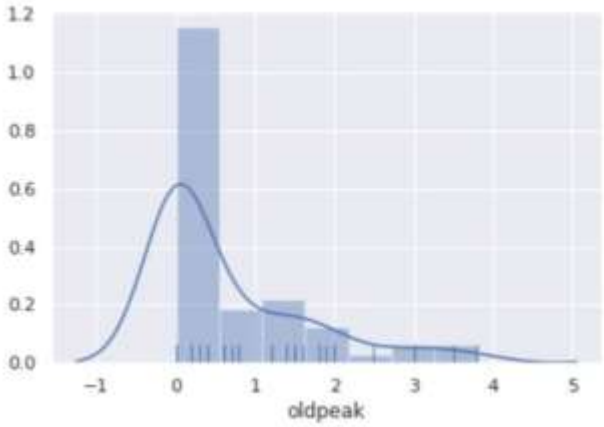

Figure 4 : Oldpeak results

The below fig 5 explains that chest pain type distribution of patient.

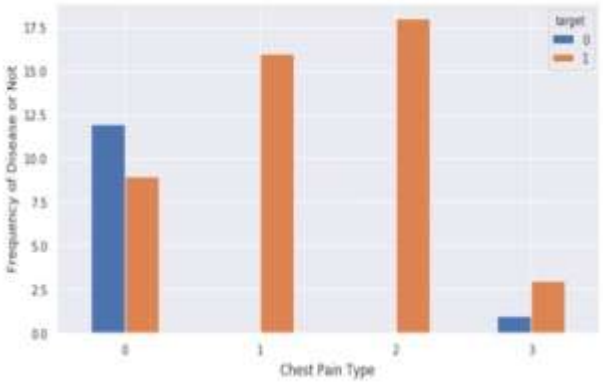

Figure 5 : Chest pain type distribution

Heart rate is depicted in Figure 6.

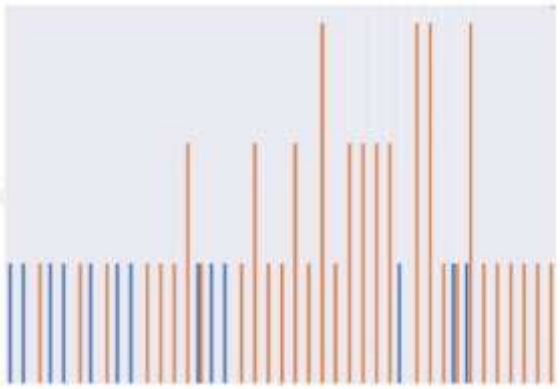

Figure 6 : Maximum heart rate achieved

The fig 7 shows the classifier creates a decision tree based on which, it assigns the class values to each data point.

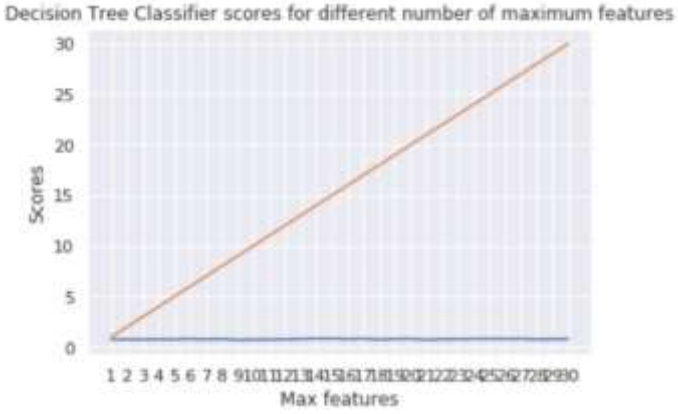

Figure 7: Decision Tree Classifier

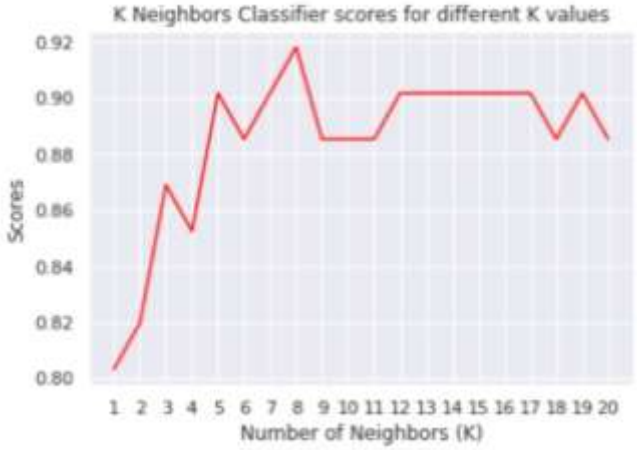

Figure 8 : K Neighbors Classifier

The classifier in Figure 8 looks for the classes and gives a class to that data point based on the class.

V. CONCLUSION

Several suggestions and classification algorithms are applied to cardiac datasets in attempt to anticipate heart issues. Classification algorithms are used to anticipate a limited number of correlations between attributes in databases in order to construct a strong classifier. The current study's key contribution is the attainment of high computation accuracy for early heart disease diagnosis. We analysed the most popular and effective heart disease prediction methods using a literature review, and then selected the most effective decision tree and logistic regression methods for heart disease prediction performance analysis. A heart ailments based on assumed factors and the best associative classification algorithm. There are a lot of them, according to the trial's findings. between the fraud detection model, the user interface, and the data inciting warehousing.

FUTURE WORK

An ensemble approach-based prototype smart heart disease prediction model including RF trees, SVM Nave Bayesian, neural networks, and logistic regression analysis-based classifiers has been offered as a system. The suggested system is a graphical user interface (GUI)-based, user-friendly, scalable, stable, and expandable system built on the WEKA platform. By delivering timely initial diagnosis, the proposed working paradigm can also help to reduce treatment expenses. The model can also be used as a teaching tool for medical students, as well as a soft diagnostic tool for doctors and cardiologists. There are other changes that might be investigated to improve the predictability and scalability of this system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets.

REFERENCES

[1] Shtake and Sanap, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," International Journal of Healthcare and Biomedical Research, vol. 1, no. 3, pp. 94-101, 2013.
[2] V. Chaurasia, "Data Mining for Early Heart Disease Prediction," Caribbean Journal of Science and Technology, vol. 1, pp. 208-217, 2013.
[3] D. S. Chaitrali and A. S. Sulabha, "A Data Mining Approach for Heart Disease Prediction Using Neural Networks," International Journal of Computer Engineering & Technology (IJCET), vol. 3, no. 3, pp. 30-40, 2012..
[4] R. Rao, "A Study on Heart Disease Prediction Using Data Mining Techniques," International Journal of Data Mining & Knowledge Management Process (IJDKP), vol. 1, no. 3, pp. 14-34, 2011..
[5] S. Vijiyarani and S. Sudha, "Disease Prediction in Data Mining Technique – A Survey," International Journal of Computer Applications and Information Technology, vol. ll, no. l, pp. 17-21, 2013.
[6] T. J. Peter and K. Somasundaram, "An Empirical Study on Heart Disease Prediction Using Classification Data Mining Techniques," 2012.
[7] Nidhi Bhatla (nine) International Journal of Engineering Research & Technology (IJERT), Kiran Jyothi, An Analysis of Heart Disease Prediction Using Different Data Mining Techniques, 2012..
[8] N. A. Sundar, P.P. Latha, and M. R. Chandra, "Performance Analysis of Classification Data Mining Techniques over Heart Disease Data Base," International Journal of Engineering Science and Advanced Technology, vol. 2, no. 2, pp. 470-478, 2012..
[9] K. Sudhakar, "Data Mining Study of Heart Disease Prediction," vol. 4, no. 1, pp. 1157-1160, 2014.
[10] R.SenthamilSelvan" A Novel Watchdog Timer for Real-Time Intensive Applications" on EAI Innovating Research-Springer , Proceedings of the First International Conference on Computing, Communication and Control System, I3CAC 2021, 7-8 June 2021, Bharath University, Chennai, India, June 2021.
[11] R. Chitra and V. Seenivasagam, "Review of Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques," Journal Soft Computing (ICTACT), vol. 3, issue 4, pp. 605-609, 2012.