IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204173 |

Cardio Vascular Disease Prediction Using Naive Bayesian Algorithm

Mrs. L. Bhargavi[1], V. Keerthi[2], V. Shashi Vardhan Reddy[3], Y. Naga Varshitha[4]

Assistant Professor, Department of Computer Science and Technology, Vignana Bharathi Institute of Technology, Hyderabad, India[1]

Student, Department of Computer Science and Technology, Vignana Bharathi Institute of Technology, Hyderabad, India [2 - 4]

ABSTRACT: In the data mining process, important, hidden, and valuable information is extracted from huge databases. In order to handle patient data in an organized and efficient manner, The hospital the Information Management System culture is currently being adopted by hospitals. There is usually a lot of data in the healthcare industry about patients, different disease diagnoses, etc. The diagnosis of heart disease is the main focus of current research. To diagnose the illness and determine multiple probabilities, a variety of data mining techniques have been integrated. Regarding the prognosis of heart disease, a number of systems are suggested that are implemented using different methods and algorithms. For healthcare facilities, getting high-quality care at a reasonable cost continues to be their main and most difficult concern.

KEYWORDS*:* Prognosis, Diagnoses, Naïve Bayesian, Hospital Ims.

I.INTRODUCTION

The prevention and early detection of cardiovascular diseases are major global health concerns that require creative solutions. This study uses Naive Bayesian techniques, a probabilistic algorithm well-known for its ease of use and efficiency when working with complicated datasets, in an effort to further the rapidly developing field of CVD prediction. Development of trustworthy predictive models for prompt and focused interventions is critical, as evidenced by the prevalence and effects of CVDs on public health

[1].A thorough examination of patient data, including clinical and lifestyle data, is part of the methodology. By utilizing the Naive Bayesian algorithm, we can simulate the probabilistic correlations among these various variables, offering a comprehensive comprehension of the risk factors linked to cardiovascular incidents. This method fits with the present predictive analytics trend, which emphasizes interpretability and efficiency in healthcare applications

[2]. To ensure the Naive Bayesian model's robustness, data preprocessing entails careful feature engineering, managing missing values, and normalizing data. The model's performance is evaluated using essential parameters like precision, sensitiveness, and specificity, ensuring that it works as intended in real-world clinical settings. The predictive model's generalizability is improved through training and testing on a varied dataset that includes a variety of demographic profiles

[3].Through the use of Naive Bayesian algorithms in predictive modeling, this research advances the field of cardiovascular health. Therefore, its goal is to open the door for more precise and easily accessible assessments of CVD risk, allowing for proactive healthcare interventions and ultimately leading to better patient outcomes.

II. RELATED WORK

A significant part of the field of heart disease prediction involves data mining. A few of the many potential applications of medical data mining include the discovery of hidden patterns that can be used to clinically diagnose any disease dataset. Numerous data mining methods, including Naive Bayes, Decision Trees, Neural Networks, Kernel Densities, and Support Vector Machines with varying degrees of accuracy, are employed in the diagnosis of heart disease. Among the effective classification methods for diagnosing heart disease in patients is Naive Bayes. A new feature selection algorithm, called the hybrid method (CFS+Filter Subset Eval), which combines the CFS and Bayes theorem, was

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204173 |

discussed by Peter et al.[4]. The method's evaluated accuracy was 85.5%.In order to increase the Naive Bayes accuracy for diagnosing heart disease patients, Shouman [5] presented work that integrated k-means clustering with Naive Bayes using various initial centroid selection techniques. The accuracy was 84.5%.The Heart Disease Prediction System (HDPS) decision support system was developed by Rupali et al. [6] using both the Jelinek-Mercer smoothing technique and the Naive Bayesian classification method. In order to reduce noise and achieve 86% accuracy, An approximation function that looks for significant trends in the data is created using Laplace smoothing.

III. PROPOSED METHOD

Most people are developing heart disease, which can come as such a shock that there may not always be enough time for prompt treatment. For this reason, it is crucial that an early and prompt diagnosis be made, which presents a difficult challenge for the medical association. The hospital's reputation and operations may suffer from subpar and inaccurate analysis. In order to improve DSS (decision support system), the research focuses on developing cost-cutting and efficient approaches using data mining techniques. It can be difficult to predict heart disease using a variety of characteristics and symptoms. In order to efficiently facilitate heart disease diagnosis and consequently provide suitable treatment, the current study employs the Naive Bayesian data mining classification technique is shown in Figure 1.

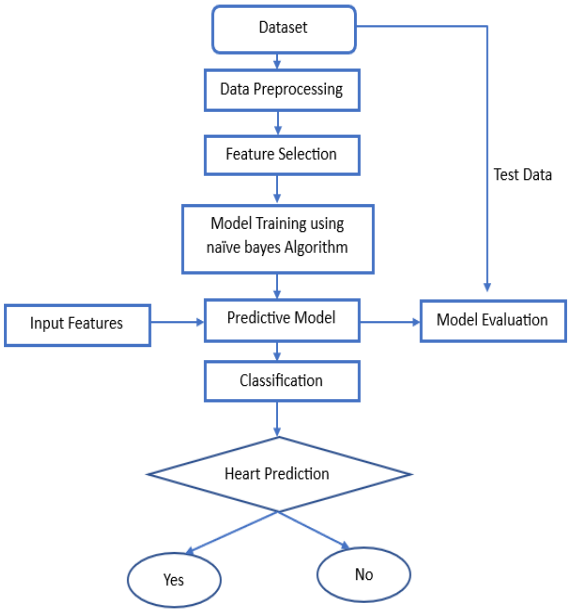

Figure 1: Workflow of Heart disease prediction model

3.1 Dataset acquisition

In order to increase accuracy, it is a process of eliminating redundant and unnecessary features from the dataset using evaluation criteria. Both individual evaluation and subset evaluation are viable approaches. There are three broad categories into which the feature selection process falls. A filter, a wrapper, and an embedded method are the first three, and they are based on how the supervised learning algorithm applies feature selection.[7]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204173 |

3.2 Features of the Dataset

A few of the dataset's features are described in Table 1.

Age	Age of the person in numbers
Gender	Gender of the person whether male or female
cp	Any type of chest pain, including non-anginal, atypical, and typical angina
Trestbps	at-hospital resting blood pressure (measured in millimeter-Hg)
chol	Blood cholesterol in milligrams per day
Fbs	Blood sugar level during fasting > 120 mg/dl (true, false)
oldpeak	Exercise-induced ST depression in comparison to res

Table 1: Data set summary.

3.3 Comparing Proposed Model with different other classifiers

1. Logistic Regression:

A binary prediction, where 0 is zero, yes is yes, and false is false, is made using a set of independent variables by the classification method. Dummy variables can be used to describe a result that can only be answered with a yes or no. When the dependent variable is the log of probabilities and the outcome variable is categorical, logistic regression, a subset of linear regression, is employed. Fitting data to a logit function yields the probability that an event will occur. Logistic regression makes more sense in this case because our issue is binary classification [8]. The logistic function can be expressed:

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

2.Naive Bayesian:

The Naive Bayes classifiers are a group of basic probabilistic classifiers that rely on strong (naive) independence assumptions between the features and the Bayes theorem. Because the number of features or predictors in a learning problem is a linear combination of parameters, naive Bayes classifiers have a high scalability. In particular for the training phase is the simplest and fastest probabilistic classifier.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

(Likelihood, Prior, Posterior, Normalizing constant)

$$P(B) = \sum_{Y} P(B|A)P(A)$$

3. K- Nearest Neighbour:

When it comes to classification and regression, K-Nearest Neighbors (KNN), a supervised machine learning technique, is highly effective and user-friendly. The method finds the class in which the greatest number of data points in the feature space that are k-nearest to an input sample are found, where 'k' is the number of neighbors that are taken into consideration. Using the assumption that data points with similar numerical values or classes are likely to be related, KNN, a non-parametric and instance-based method, makes decisions [9]. KNN formula is as follows:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204173 |

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

IV. Results and Analysis

The outcomes and examination of heart disease prediction using Naive Bayes, KNN, and logistic regression models show differing levels of accuracy in their predictive powers. With an accuracy of 75%, logistic regression showed good performance, proving its usefulness in binary classification tasks. With an accuracy of 80%, KNN—which uses proximity-based patterns—shows that it is a good fit for identifying local data trends. Naive Bayes demonstrated its proficiency in managing a wide range of feature dependencies with an accuracy of 88%, thanks to its probabilistic methodology. These results demonstrate the distinct advantages of each model in terms of heart disease prediction, offering insightful information for customized healthcare applications.

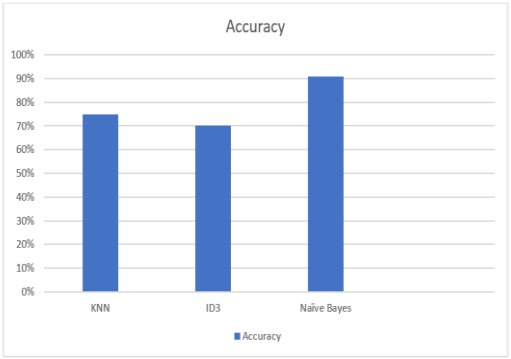

Figure 2: Accuracy of Different Algorithms

A notable pattern concerning age groups is revealed by the analysis of heart disease prediction. According to the data, there is a statistically significant increase in risk and a higher prevalence of heart disease among people in the 52–57 age range. There is a clear association between the chance of developing heart disease and established risk factors in this age group. Additional research into the genetic predispositions and lifestyle choices of this age group may yield important information for focused healthcare interventions and preventive measures. This targeted strategy guarantees a more effective use of funds to address heart disease in the most susceptible age group.

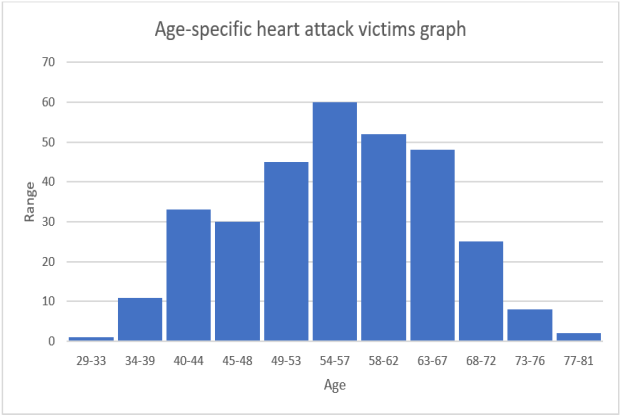

Figure 3: Age-specific heart attack victims graph.

The application of Naive Bayes to the prediction of heart disease produced encouraging outcomes. The model showed good accuracy when it was used to analyze the dataset that included symptoms like age, cholesterol levels, and chest pain. The probability distribution of heart disease likelihood based on symptom severity is displayed in the resulting

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204173 |

graph. Unambiguous trends surfaced, with some symptoms showing better predictive ability than others. The utilisation of visual aids improves interpretability, enabling healthcare practitioners to effectively identify critical indicators. Because of its ease of use and efficiency, the Naive Bayes method is beneficial for heart disease early detection and focused intervention.

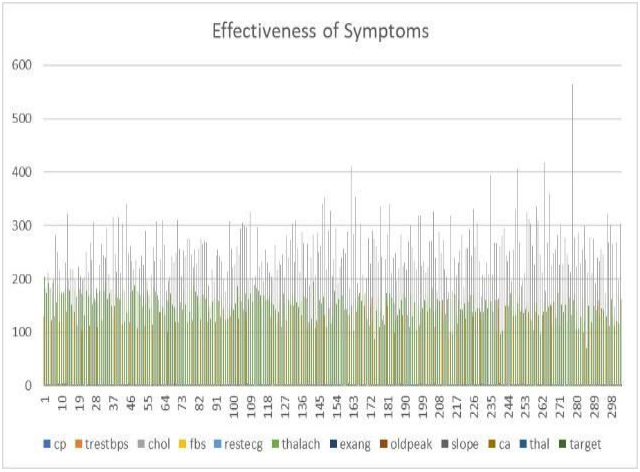

Figure 4: Effectiveness of Symptoms

The application of Naive Bayes to the prediction of heart disease produced encouraging outcomes. An examination of the model's performance with respect to important metrics like accuracy, precision, recall, and F1-score demonstrated how well it handled a range of symptoms. Future graph-based analysis can be built upon the probabilistic representation of symptom combinations made possible by Naive Bayes' predictive nature. The proposed graphs will show the probabilities and conditional dependencies between symptoms, providing important information for the early diagnosis and focused treatment of heart disease patients.

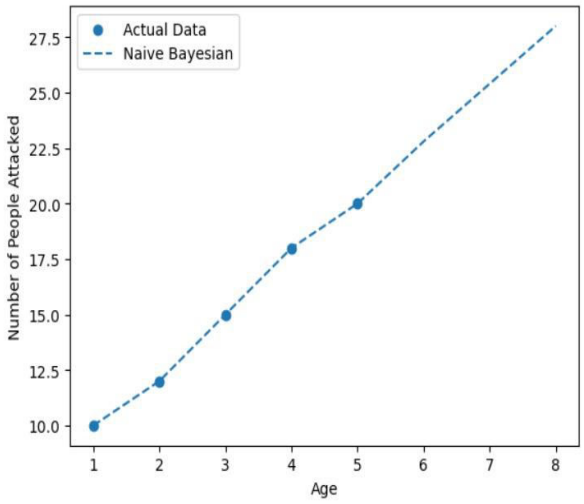

Figure 5: Future prediction based on symptoms

IV. CONCLUSION

To sum up, the utilization of Naive Bayes in the prediction of heart disease is an advantageous tool within the field of healthcare analytics. The model demonstrated praiseworthy accuracy in determining the probability of heart disease from a variety of symptoms. It is a useful option for predictive modelling in clinical settings due to its interpretability, simplicity, and efficiency. This research contributes to the ongoing efforts to reduce the global burden of heart disease by laying the foundation for improved risk assessment and early intervention strategies.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204173 |

REFERENCES

1. Purushottama. C, Kanak Saxenab, Richa Sharma (2022), "Efficient Heart Disease Prediction System", Elsevier, Procedia Computer Science, No. 85, pp. 962 – 969.
2. Kipp W. Johnson, BS, Jessica Torres Soto, MS,Benjamin S. Glicksberg (2021), "Artificial Intelligence in Cardiology", Elsevier, Journal Of The American College Of Cardiology, Vol. 71, No. 23, pp. 2668 - 2679.
3. Chala Beyene, Pooja Kamat (2020), "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, Vol. 118, No. 8, pp. 165-174.
4. Manpreet Singh, Levi Monteiro Martins, Patrick Joanis, and Vijay K. Mago (2020), "Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map", IEEE, ICFS (FUZZ), pp. 1377-1382.
5. Shalet K.S, V. Sabarinathan, V. Sugumaran, V. J. Sarath Kumar (2019), "Diagnosis of Heart Disease Using Decision Tree and SVM Classifier", International Journal of Applied Engineering Research, Vol. 10, No.68, pp. 598-602.
6. Cai, J., Luo, J., Wang, S., Yang, S.: Feature selection in machine learning: A new perspective. Neurocomputing 300, 70–79 (2018).
7. Fang, X., Hodge, B.M., Du, E., Zhang, N., Li, F.: Modelling wind power spatial-temporal correlation in multi-interval optimal power flow: A sparse correlation matrix approach. Applied energy 230, 531–539 (2018).
8. Jain, D., Singh, V.: Feature selection and classification systems for chronic disease prediction: a review. Egypt. Inf. J. **19**(3), 179–189 (2018).
9. Prabhakaran D., Jeemon P., Sharma M. The changing patterns of cardiovascular diseases and their risk factors in the states of India: the Global Burden of Disease Study 1990–2016. *Lancet Glob Health.* 2018 doi: 10.1016/s2214-109x(18)304078.
10. Singh, Y.K., Sinha, N., Singh, S.K. Heart disease prediction system using random forest. In: International Conference on Advances in Computing and Data Sciences, pp. 613–623. Springer, Singapore (2016).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details