IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11455

Heart Disease Prediction Using Naive Bayes Classifier

Sudhanshu Memane[1], Aakash Patel[2], Anjal Patel[3], Omkar Dive[4], Uzma Shaikh[5]

Student, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India[1,2,3,4]

Assistant Professor, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India[5]

Abstract: It might have happened such a lot of times thatyou just or somebody yours would like doctors facilitate right away, however they're not obtainable thanks to some reason.The heart disease Prediction application is a user support and on-line consultation project. Here, we tend to propose a web application that enables users to induce instant steerage on their cardiopathy through an intelligent system online. The application is fed with numerous details and also the heart disease related to those details. The application permits userto share their heart connected problems. It then processes user specific details to see for numerous health problem that mightbe related to it. Here we tend to use some intelligent data processing techniques to guess the foremost correct illness that might be related to patient's details. Supported result, they will contact doctor consequently for any treatment. The system permits user to look at doctor's details too. The systemmay be used without charge heart disease consulting online.

Keywords: Naive Bayes classifier, heart disease prediction, Python, Machine learning

I. INTRODUCTION

A major challenge facing health care organizations (hospitals, medical centres) is that the provision of quality services at reasonable prices. Quality service implies diagnosing patients properly and administering treatmentsthat are effective. Poor clinical selections will result in fatefulconsequences that are so unacceptable. Hospitals should additionally reduce the price of clinical tests. They'll come through these results by using applicable computer-based info and/or call support systems. Most hospitals today use some form of hospital information systems to manage theirhealth care or patient knowledge. These systems usually generate huge amounts of knowledge that take the shape ofnumbers, text, charts and pictures. Sadly, these knowledgesare rarely accustomed support clinical decision-making. Naive Bayes or Bayes' Rule is that the basis for several machine-learning and data processing ways. The rule (algorithm) is employed to make models with predictive capabilities. It provides new ways that of exploring and understanding knowledge. It learns from the "evidence" bycalculating the correlation between the target (i.e., dependent) and different (i.e., independent) variables.

II. LITERATURE SURVEY

1. "Polaraju, Durga Prasad, & Tech Scholar" (2017)

This paper proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing.

2. "Deepika & Seema" (2017)

This study focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine (SVM) and Artificial Neural Network (ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

3. "Beyene & Kamat" (2018)

recommended different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms suggested Heart Disease Prediction System using Data Mining Techniques. WEKA software used for automatic diagnosis of disease and to give qualities of services in healthcare centres. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11455

4. "A & Naik" (2016)

This study recommended to develop the prediction system which will diagnosis the heart disease from patient's medical data set. 13 risk factors of input attributes have considered to build the system. After analysis of the data from the dataset, He used k-means and naïve Bayes to predict heart disease. To extract knowledge from database, data mining techniques such as clustering, classification methods can be used. 13 attributes with total of 300 records were used from the Cleveland Heart Database. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes.

5. "Sultana, Haider, & Uddin" (2017)

This study proposed an analysis of cardiovascular disease. This paper proposed data mining techniques to predict the disease. It is intended to provide the survey of current techniques to extract information from dataset and it will useful for healthcare practitioners. The performance can be obtained based on the time taken to build the decision tree for the system. The primary objective is to predict the disease with a smaller number of attributes.

6. "Sai & Reddy" (2017)

This study proposed heart disease prediction using ANN algorithm in data mining. Due to increasing expenses of heart disease diagnosis disease, there was a need to develop new system which can predict heart disease. Prediction model is used to predict the condition of the patient after evaluation on the basis of various parameters like heart beat rate, blood pressure, cholesterol etc.

III. EXISTING SYSTEM

In the existing system the data set is typically small, for patients and diseases with specific conditions. These systems are mostly designed for the more casual symptoms such as mild pain, etc. The pre-selected characteristics may sometimes not satisfy the changes in the disease and its influencing factors which could lead to inaccuracy in results. As we live in continuously evolving world, the symptoms of diseases also evolve over a course of time. Also, most of the current systems make the users wait for long periods by making them answer lengthy questionnaires.

IV. PROPOSED SYSTEM

In this system we are implementing effective heart attack prediction system using Naïve Bayes algorithm. We can give the input as in CSV file or manual entry to the system. After taking input the algorithms apply on that input that is Naïve Bayes**.** After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.

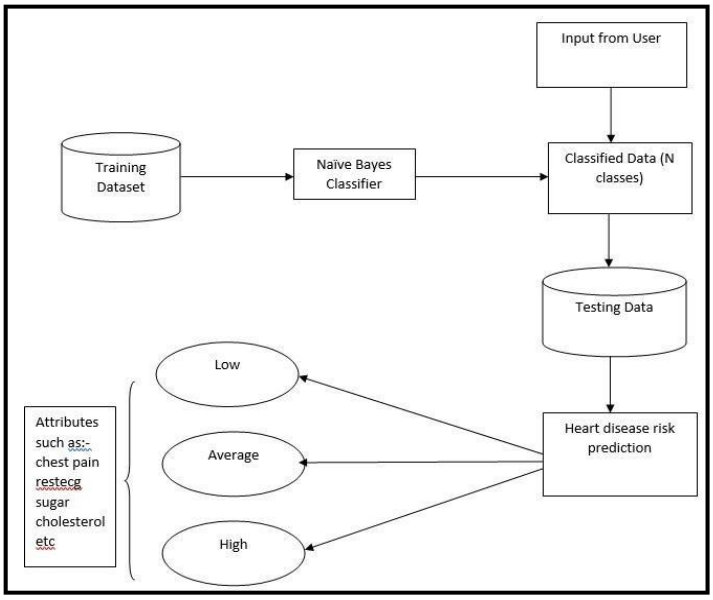

Fig. 1 System Architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11455

Here, we propose a web application that allows users to getinstant guidance on their heart disease through an intelligent system online. The application is fed with various details and the heart disease associated with those details. The application allows user to share their heart related issues. Itthen processes user specific details to check for various illness that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's details.

Table I Input Attributes

No.	Attributes	Description	Values
1	Age	Age	Patient's age
2	Sex	Gender	1=Male, 0=Female
3	Cp	Chest pain	1=Typical angina 2=Atypical angina 3=Non-typical angina 4=Asymptomatic
4	Trestbps	Resting blood pressure	Continuous value in mg/dl
5	Chol	Serum cholesterol	Continuous value in mg/dl
6	Fbs	Fasting blood sugar	1 ≥ 120 mg/dl 1 ≤ 120 mg/dl
7	Restcg	Resting ECG	0=Normal 1=ST wave 2=Left hypertrophy
8	Thalach	Maximum heart rate achieved	Continuous value
9	Exang	Exercise induced angina	0=No, 1=Yes
10	Oldpeak	ST depression induced by exercise relative to set	Continuous value
11	Slope	Slope of the peak exercise ST segment	1=Upsloping 2=Flat 3=Down sloping
12	Ca	Number of major vessels colored by fluoroscopy	0-3 value
13	Thal	Defect type	3=Normal 6=Fixed defect 7=Reversible defect

Questionnaires have advantages over some other types of medical symptoms that they are cheap, do not require as much effort from the questioner as verbal or telephone surveys, and often have standardized answers that make it simple to compile data. With the help of the dataset, the patterns important to the heart attack prediction areextracted. The records were split equally into 2 datasets:training dataset and testing dataset. To avoid bias, the records for every set were selected indiscriminately.

V. NAIVE BAYES CLASSIFIER

A Naive Bayes' classifier may be a term addressing a simpleprobabilistic classification supported applying Bayes' theorem. In easy terms, a Naïve Bayes classifier assumes that the presence (or absence) of a specific feature of a category is unrelated to the presence (or absence) of the other feature.As an example, a fruit could also be thought of to be an apple if it's red, round, and regarding 4" in diameter. Even supposing these options rely on the existence of the opposite options, a Naive Bayes' classifier considers all of those properties to independently contribute to the likelihood that this fruit is an apple.
Given training data X, posterior probability of a hypothesis H, P(H|X), follows the Bayes theorem,
P(H|X) = P(X|H)P(H)/P(X)
The Naive Bayes algorithm is based on Bayesian theorem asgiven by equation
Steps in algorithm are as follows:
a) Each data sample is represented by an n dimensionalfeature vector, X = (x1, x2….. xn), depicting n measurementsmade on the sample from n attributes, respectively A1, A2,An.
b) Suppose that there are m classes, C1, C2……Cm. Givenan unknown data sample, X (i.e., having no class label), the

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022
DOI: 10.17148/IJARCCE.2022.11455

classifier will predict that X belongs to the class having the highest posterior probability, conditioned if and only if, P(Ci/X)>P(Cj/X) for all 1< = j< = m and j!= i

Thus, we maximize P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes theorem,

c) As P(X) is constant for all classes, only P(X|Ci)P(Ci) need bemaximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, i.e., P(C1) = P(C2) = P(Cm), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci). Note that the class prior probabilities may be estimated byP(Ci) = si/s , where Si is the number of training samples of class Ci, and s is the total number of training samples. on X.That is, the naive probability assigns an unknown sample X to the class Ci

VI. CONCLUSION

This paper gives research of multiple researches done in this field. Our Proposed System aims at bridging gap between Doctors and Patients which will help both classes of users in achieving their goals. The proposed system scalable, reliable and an expandable system. The proposed working model can also help in reducing treatment costs by providing Initial diagnostics in time. The model can also serve the purpose of training tool for medical students and will be a soft diagnostic tool available for physician and cardiologist. As we have developed a generalized system, in future we can use this system for the analysis of different data sets. The performance of the health's diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research.

REFERENCES

[1]. Meenakshi, K., Maragatham, G., Agarwal, N., & Ghosh, I. (2018). A Data mining Technique for Analyzing and Predicting the success of Movie.
[2]. Desai, S. D., Giraddi, S., Narayankar, P., Pudakalakatti, N. R., & Sulegaon, S."Back-propagation neural network versus logistic regression in heart disease classification" (2017)
[3]. A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk prole of heart failure" (2011)
[4]. M. Durairaj and N. Ramasamy, "A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate" (2016)
[5]. Nandhini, S., Debnath, M., Sharma, A. and Pushkar 2018. Heart Disease Prediction using Machine Learning.
[6]. Garate-Escamila, A. K., Hassani, A. H. and Andres, E. 2020. Classification models for heart disease prediction using feature selection and PCA. Informatics in Medicine Unlocked 19:1-11.
[7]. Kavitha, S., Baskaran, K. R. and Sathyavathi, S. 2018. Heart Disease with Risk Prediction using Machine Learning Algorithms. International Journal of Recent Technology and Engineering 7(48):314-317.
[8]. Polaraju, K., Durga Prasad, D., & Tech Scholar, M. (2017). Prediction of Heart Disease using Multiple Linear Regression Model.
[9]. Obenshain, M.K," Application of Data Mining Techniquesto Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.
[10]. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms" Dinesh Kumar G., Santhosh Kumar D.(2018).
[11]. "Heart Disease Prediction Using Data Mining Techniques" H. Benjamin Fredrick David and S. Antony Belcy(2018).
[12]. "Multi Disease Prediction Using Data Mining Techniques" K. Gomathi, Dr. D. Shanmuga Priyaa (2017).
[13]. "Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning" Reddy Prasad Pidaparthi Anjali, S. Adil, N. Deepa (2019)
[14]. Jaymin Patel, Prof.Tejal Upadhyay, Dr. Samir Patel "Heart disease prediction using Machine learning and Data Mining Technique" Volume7.Number1 Sept 2015March 2016