IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Volume 4, PP 64-67

Heart Disease Prediction Using Data Mining

Palash Ukey[1], Vinit More[2],Pravin Patil[3], Amruta Sankhe[4].
[1,2,3]*Department of Computer Engineering, Atharva College of Engineering/Mumbai University, India*
[4] *Assistant Professor Amruta Sankhe of Computer Engineering, Atharva College of Engineering/Mumbai University, India*

Abstract:-*In today's modern world we have various heart related problems that require immediate diagnosis and remedy. So, to overcome these diseases, we are designing a Web application with the help of Data Mining Techniques that will help us identify various heart related problems and provide solutions to it. We decided to build a project which we name as HEART DISEASE PERDICTION. In this, we are giving the user quick guidance about heart disease through an Intelligent System Online. By adding basic details about the heart disease, patient can diagnose the condition or problem he/she is having with this Web based application. The application is fed with various details and the heart disease associated with those details. Through this application we allows user to share their heart related issues. In order to guess the most accurate illness that could be associated with patient's details, Data mining techniques are being used. Based on result, system automatically shows the result along with the specific doctors for further treatment.*

Keywords: *Disease Predication, Data Mining, Naive Bayes Algorithm, Neural Network, Decision tree.*

I Introduction

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable .Hospitals must also minimize the cost of clinical tests. This result can be achieved by Hospital's by employing appropriate decision support systems and/or computer-based information [1]. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. Huge amount of data is been typically generated by these systems which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped [3].

How can we turn these types of data into useful information which can help to make intelligent clinical decisions and can enable healthcare practitioners? Although data mining has been around for more than two decades, its potential is only being realized now [4].

II Existing System

Problem with current scenario

The information systems at various hospitals are designed in such a way so that it supports patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited [2].

They can answer simple queries like "What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer[4]."

However the answers to various queries such as "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", "Identify the important Preoperative predictors that increase the length of hospital stay" and "Given patient records, predict the probability of patients getting a heart disease" are not answered[5].

Drawbacks of the existing system

1.) Maintenance of the system is very difficult.
2.) There is a possibility for getting inaccurate results.
3.) User friendliness is very less.
4.) It consumes more time for processing the activities.

III Literature Survey:

Palaniappan and Awang decided to built a Heart Disease Prediction System using Data mining techniques like Decision Trees, Naïve Bayes and Neural Network. The results demonstrated the strange strength

of each of the methodologies in realizing the objectives of the specified mining objectives. Conventional decision support systems were not able to answering queries that Heart Disease Prediction System was able to. Extraction of interesting patterns from the dataset using vital parameters is the aim of this paper to employ and analyze different data mining techniques for the prediction of heart disease in a patient. It facilitated the establishment of vital knowledge, e.g. patterns, relationships amid medical factors connected with heart disease.[1]

This paper strives to bring out the methodology and implementation of these techniques- Decision Tree and Naive Bayes and stress upon the results and conclusion induced on the basis of accuracy and time complexity .Different types of studies have been done to focus on prediction of heart disease. Various data mining techniques are used for diagnosis and achieved different accuracy level for different methods [2].

The Naive Bayes algorithm uses conditional independence; it believes that an attribute value of a given class is independent of the values of other attributes. The preprocessed data has been considered as the training set. Two phase namely classification and prediction was discussed in that work. Preprocessing is done in the classification phase. The Heart Disease Prediction can be completely automated through an inefficient online software program or software application. The preprocessing includes cleaning of data, normalization and reduction of data, etc. In the prediction phase the disease types are classified and predicted, i.e. a training set is formed based on the disease type and the test set is formed based on the questions. The predicted results are sent to the doctor [4].

Clinical choices are frequently made focused around doctors' instinct and experience instead of on the knowledge rich information covered up in the database. This practice prompts undesirable biases, blunders and unnecessary medicinal expenses which influence the quality of services given to the patients. Wu, et al proposed that combination of clinical choice backing with computer based patient records could decrease medical errors, enhance safety of patients, lessening undesirable practice variety, and enhance patient outcome The benefits of implementing this technology touch everyone involved in the scheduling process, as administrators and users can conduct their tasks more efficiently and accurately. The system brings out of various knowledge from a historical heart disease database. This system can be further enhanced and expanded [5].

ANN, often just called a "neural network", is a mathematical model or computational model used for a biological purpose. In other words, it is an emulation of biological neural system. It has mainly three layers, i.e. the input layer, hidden layer and the output layer. The input is given to the input layer and the result is obtained in the output layer [6].

Naive Bayes classifier mainly pre assumes the effect of a variable value on predefined class that is not dependent on value of other variable. This is called as property of class conditional independence. It is particularly suited when the dimensionality of the inputs is high. Naïve Bayesian is mainly used to form models with predictive capabilities. Bayes' Theorem: Probability (B given A) = (Probability (A and B)/Probability (A))Assume X as a data tuple. Let H be any hypothesis. P (H|X) be posterior probability of the H that is conditioned on X. In the same way, P (X|H) is the posterior probability of X condition on H.P (H|X) = (P(X|H)P(H)/ P (X))P(H) is prior probability of H[7].

Data Mining Extension (DMX) query language was used for model creation, model training, model prediction and model content access. All parameters were set to the default setting except for parameters "Minimum Support = 1" for Decision Tree and "Minimum Dependency Probability = 0.005" for Naïve Bayes [10]. The trained models were evaluated against the test datasets for accuracy and effectiveness before they were deployed in IHDPS. The models were validated using Lift Chart and Classification Matrix[8].

The top green line shows the ideal model; it captured 100% of the target population for patients with heart disease using 46% of the test dataset. The bottom blue line shows the random line which is always a 45-degree line across the chart. It shows that if we randomly guess the result for each case, 50% of the target population would be captured using 50% of the test dataset. All three model lines (purple, yellow and red) fall between the random-guess and ideal model lines, showing that all three have sufficient information to learn patterns in response to the predictable state [9].

The study done by Kharya highlighted the fact that artificial neural network is thefrequently used technique for prediction in

the medical field. The paper also demonstrated the merits & de merits of the machine learning techniques like Decision

tree, Naïve Bayes, Neural network & SVM [10]

IV Purposed:

1.) Considering the anomalies in the existing system computerization of the whole activity is being suggested after initial analysis.

2.) It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason.

3.) The Heart Disease Prediction application is an end user support and online consultation project.
4.) Here, we propose a web application that allows users to get instant guidance on their heart disease through an intelligent system online.
5.) Various details are fed in the application and the heart disease associated with those details. Users can share their heart related issues with this application.
6.) It then processes user specific details to check for various illness that could be associated with it.
7.) After getting the result from the system, system suggests various doctors for treatment. The system allows user to view doctor's details.
8.) The system can be use in case of emergency.

DIAGRAM:

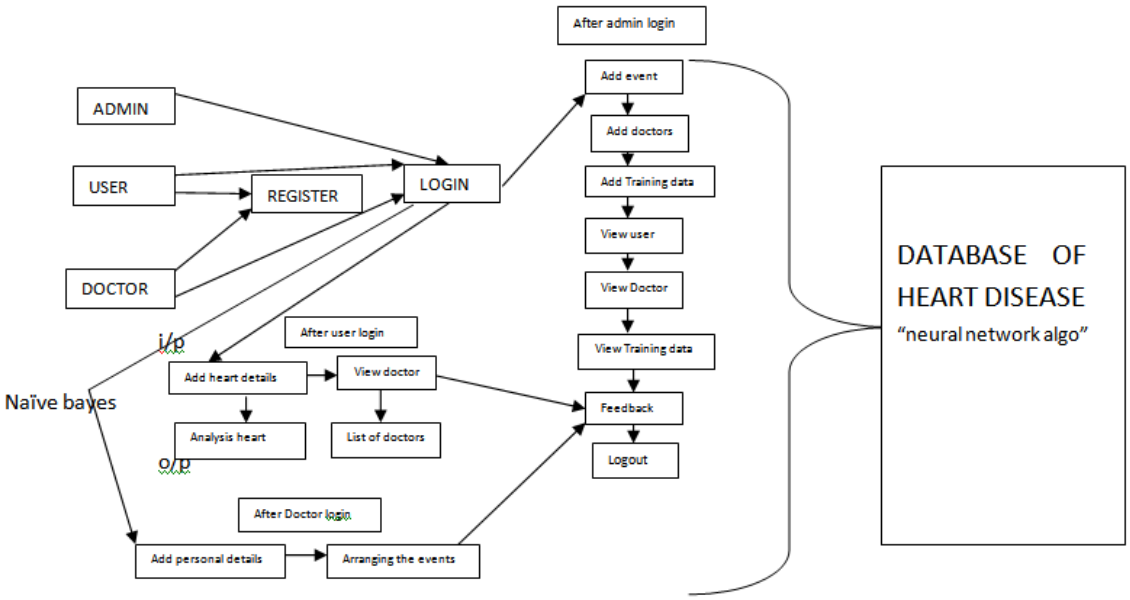

Fig 1.Block Diagram

Explanation: In the above diagram is the flow of our project where there are three modules are been provided by us so in each modules there are some sub modules in each modules we have provide a security for each modules so that one can easily identify by the email id and password in this is the user is new for this application then he/she need to register in this application after registration user can login it with the particular email id and password after login each module has to perform his own work he don't need to interfere in the others details in such a way the Admin, User, Doctor will perform his task so let's take one module Admin module will login with his email id and password in this module there are sub modules are been provided by us in which admin can Add event, Add Doctor, Add Training Data, View the user, View Doctor , View the Training Data, Feedback and Logout. The second module is User in which if the user is new for this application then he will register it and then he will login with email id and password after there will be sub module in which user will go for the entering the details of heart after entering it will analysis the heart prediction in which it will show which type of the disease is it in this the user can view the doctor for taking a quick guidance along with this we are providing the doctors details and the last modules is Doctor in this the doctor will add his personal details and Event which are been organize by their institute along with this there is feedback form for your review how was your experience through which we can improve are project at a higher end user. In the above flow we are using the naives bayes algorithm in which we are using for the good performance base on through which we can get accurate result in which it will show us which type of heart disease is it. In the database system we are using the neural network in which we can manage all the training data through which we can select the appropriate data set.

V Conclusion

The Heart Disease Prediction can be completely automated through an inefficient online software program. The benefits of implementing this technology touch everyone involved in the scheduling process, as administrators and users can conduct their tasks more efficiently and accurately. The system gets deep knowledge from a historical heart disease database. This system can be further enhanced and expanded.

Acknowledgements

We are pleased to present **"Heart Disease Prediction"** project and take this opportunity to express our profound gratitude to all those people who helped us in completion of this project.

We thank our college for providing us with excellent facilities that helped us to complete and present this project. We would also like to thank the staff members and lab assistants for permitting us to use computers in the lab as and when required.

We express our deepest gratitude towards our project guide for her valuable and timely advice during the various phases in our project. We would also like to thank her for providing us with all proper facilities and support as the project co-coordinator. We would like to thank her for support, patience and faith in our capabilities and for giving us flexibility in terms of working and reporting schedules.

We would like to thank all our friends for their smiles and friendship making the college life enjoyable and memorable and family members who always stood beside us and provided the utmost important moral support. Finally, we would like to thank everyone who has helped us directly or indirectly in our project.

Refrences

[1]. Heart Disease Prediction System Using Data Mining Technique Abhishek Taneja **,**Department of Computer Science, S.A. Jain College , Ambala City ,India.(Received:November15, 2013; Accepted: November25, 2013).

[2]. Intelligent Heart Disease Prediction System Using Data Mining Techniques, IJCSNS International Journal of Computer Science and Network Security, VOL.8No.8 August2013.

[3]. Prediction of Heart Disease using Data Mining Techniques Indian Journal of Science and Technology**,** *Vol 9(39),DOI: 10.17485/ijst/2016/v9i39/102078,October2015* http://www.sanfoundry.com/best-reference-books-neural-network-fuzzy-logic/.

[4]. Fayyad,UDataMiningandKnowledgeDiscoveryinDatabases:ImplicationsfromscientificdatabasesProc.ofthe9thInt.Conf.onScientificandStatisticalDatabaseManagement,Olympia,Washington,USA,2-11,2013 http://ieeexplore.ieee.org/document/4493524/DOI-10.1109/AICCSA.2013.4493524.

[5]. Human Heart Disease Prediction System using Data Mining Techniques, Theresa Princy. R and J. Thomas Department of Computer Science and Engineering Christ University faculty of engineering DOI-10.1109/ICCPCT.2016.7530265

[6]. L. Burke, J. Ma, K. Azar, G. Bennett, E.Peterson,Y. Zheng, W. Riley,J. Stephens, S. Shah, B.Suffoletto, T. Turan, B. Spring, J. Steinbergerand C. Quinn,"Current Science on Consumer Use of Mobile Healthfor Cardiovascular Disease Prevention", Circulation, vol. 132, no. 12,pp. 1157-1213, 2015 http://ieeexplore.ieee.org/document/7154917/DOI- 10.1109/ABLAZE.2015.7154917.

[7]. M. Raihan1, Saikat Mondal2, Arun More3, Md. Omar Faruqe Sagor4, Gopal Sikder5, Mahbub Arab Majumder5, Mohammad Abdullah Al Manjur5 and Kushal Ghosh ,"Smartphone Based Ischemic Heart Disease (Heart Attack) Risk Prediction using Clinical Data and Data Mining Approaches, a Prototype Design", International Conference on Computer and Information Technology, December 18-20, 2016.

[8]. Ms.Rupali R.Patil," Heart Disease Prediction System using Naive Bayes and Jelinek mercer smoothing", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2014.

[9]. Usage of Data mining techniques in predicting the Heart diseases – Naïve Bayes & Decision tree Priyanka N B.E., (M.Tech) 1 CS & E dept Dr.Pushpa RaviKumar B.E., M.Tech., Ph.D 2 AIT College, CS & E dept Chikkamagaluru, Karnataka state,India . AIT College, Chikkamagaluru, India. http://ieeexplore.ieee.org/document/8074215/ DOI- 10.1109/ICCPCT.2017.8074215