Heart Disease Prediction Using Naive Bayes

[1]Akanksha Toley
Department of Electronics, V.E.S.I.T., Mumbai University Sindhi Society, Chembur, Maharashtra, India
Email: akanksha.toley@ves.ac.in
[2] Shrutika Jambhale
Department of Electronics, V.E.S.I.T., Mumbai University Sindhi Society, Chembur, Maharashtra, India
Email: shrutika.jambhale@ves.ac.in
[3]Sukanya Rawool
Department of Electronics, V.E.S.I.T., Mumbai University Sindhi Society, Chembur, Maharashtra, India
Email: sukanya.rawool@ves.ac.in
[4]Vinayak Sridhar
Department of Electronics, V.E.S.I.T., Mumbai University Sindhi Society, Chembur, Maharashtra, India
Email: vinayak.sridhar@ves.ac.in
[5]Anushree Prabhu
Asst. Prof, Department of Electronics, V.E.S.I.T., Mumbai University Sindhi Society, Chembur, Maharashtra, India
Email: anushree.prabhu@ves.ac.in

Abstract—The main objective of this research is to develop a decision support in Heart Disease Prediction System (HDPS) using one Data Modelling technique, namely, Naive Bayes. Using medical attributes such as age, sex, blood pressure, ECG graph, etc. It can predict the likelihood of patients getting a heart disease. HDPS is implemented as an application in MATLAB which can answer user queries, it can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot.

I. INTRODUCTION

As large amount of data is generated in medical organi-zations (hospitals, medical centers) but as this data is not properly used. There is a wealth of hidden information present in the datasets. This unused data can be converted into useful data. For this purpose we can use different data mining tech-niques. This paper presents a classifier approach for detection of heart disease and shows how Naive Bayes can be used for classification purpose. In our system, we will categories medical data into five categories namely no, low, average, high and very high. Also, if unknown sample comes then the system will predict the class label of that sample. Hence two basic functions namely classification (training) and prediction (test-ing) will be performed. Accuracy of the system is depends on algorithm and database used. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not mined to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. This research has developed a prototype Heart Disease Prediction System (HDPS) using data mining techniques, namely, Decision Trees, Nave Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals. HDPS can answer complex what if queries which traditional decision support systems cannot. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. It enables signif-icant knowledge, e.g. patterns, relationships between medical factors related to heart disease, to be established. HDPS is Web-based, user-friendly, scalable, reliable and expandable. In our project the main objective is to develop an Intelligent System using data mining modeling technique, namely, Naive Bayes. It is implemented as web based application in this user answers the predefined questions. It retrieves hidden data from stored database and compares the user values with trained data set. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent

clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce cost of treatment. 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 con-sequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data 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. This raises an important question: How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions? This is the main motivation for this research.

A. Review of Literature

Most hospitals today employ sort of hospital information systems to manage their healthcare or patient data. These systems typically produce large amounts of data. There is a wealth of unknown information in these data that is largely not accessed. So how this data can be converted into useful infor-mation that can enable healthcare systems and practitioners to make intelligent clinical decisions. The main objective of this research work is to develop a Decision Support system in Heart Disease Prediction System (HDPS) using one data mining modeling technique, namely, Nave Bayes and another one is the smoothing to improve performance. HDPS is implemented as an application in MATLAB which can answer user queries, it can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. We provide the report of the patient which indicates whether that particular patient having the heart disease or not. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have an ability to generate a knowledge rich environment which can help to significantly improve the quality of clinical decisions. The diagnosis of diseases is a significant and complex task in medicine. To detect heart disease from various factors or symptoms is a multi -layered issue which is not free from false presumptions often accompanied by effects that are not predictable. Thus the attempt to utilize knowledge and experience of number of specialists and clinical screening data of patients which collected in databases to facilitate the diagnosis process is considered a valuable option. We apply several smoothing models to Naive Bayes for improving results, and study their performance. The experimental results on a large database show that the smoothing methods are able to significantly improve performance of Nave Bayes.

B. Data Mining

Data mining is a powerful new technology with great poten-tial to help companies focus on the most important information in the data they have collected about the behavior of their customers and potential customers. It discovers information within the data that queries and reports can't effectively reveal It is the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most impor-tant information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding

3) The Future of Data Mining: In the short-term, the results of data mining will be in profitable, if mundane, business related areas. Micro-marketing campaigns will explore new niches. Advertising will target potential customers with new precision. In the medium term, data mining may be as common and easy to use as e-mail. We may use these tools to find the best airfare to New York, root out a phone number of a long-lost classmate, or find the best prices on lawn mowers. The long-term prospects are truly exciting. Imagine intelligent agents turned loose on medical research data or on sub-atomic particle data. Computers may reveal new treatments for diseases or new insights into the nature of the universe. There are potential dangers, though, as discussed below.

C. Naive Bayes Technique

1) Introduction: Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness and diameter features. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the appar-ently implausible efficacy of naive Bayes classifiers. Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests. An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters necessary for classification.

2) System Architecture : The diagram given is basically the System architecture. Different important blocks are shown in the diagram.

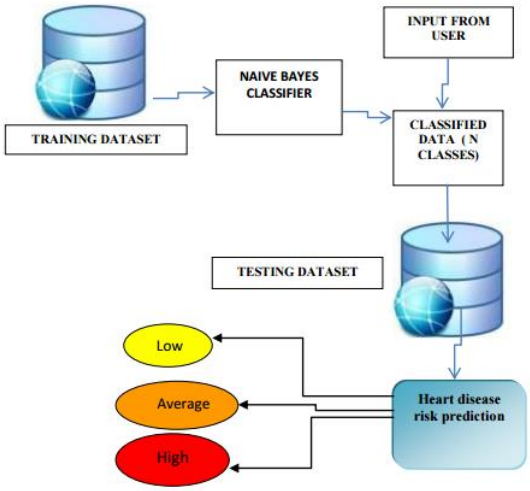
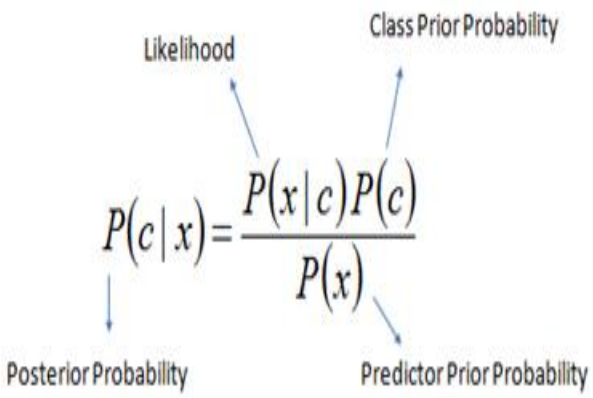


Fig. 1. System Architecture



$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

Fig. 2. Naive Bayes Model for Classification

D. Naive Bayes Model for Classification

Naive Bayes is a simple but important probabilistic model. It will be used as a running example in this note. In particular, we will first consider maximum-likelihood estimation in the case where the data is fully observed; we will then consider the expectation maximization (EM) algorithm for the case where the data is partially observed, in the sense that the labels for examples are missing. The Naive Bayesian classifier is based on Bayes theorem with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it par-ticularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

Naive Bayes Algorithm Given the intractable sample com-plexity for learning Bayesian classifiers, we must look for ways to reduce this complexity. The Naive Bayes classifier does to reduce the complexity by making a conditional inde-pendence assumption that dramatically reduces the number of parameters to be estimated when modeling P(X—Y), from our original 2(2n-1) to just 2n.

E. Naive Bayes Algorithm

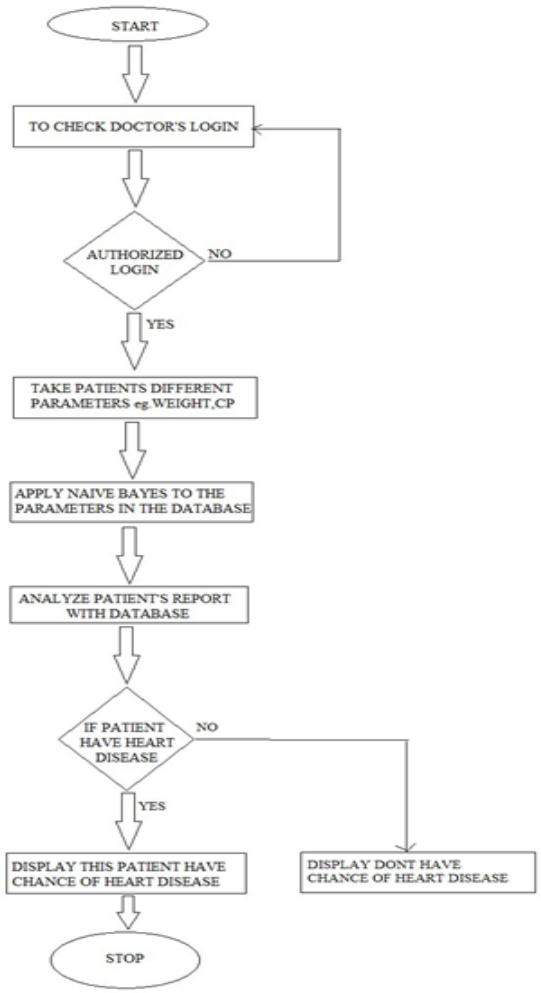


Fig. 3. Algorithm

II. CONCLUSION

In this paper, development of an effective heart disease prediction system was limned along with the techniques used for development.The main techniques in our system are chosen after exhaustive research and comparison with other available methods and algorithms . Taking into account every known method used currently, we came up with Naive Bayes algo-rithmic method.Predictive analysis used for medical purposes is one of the most integrating areas of machine learning.To use this method to predict the likelihood of heart disease as accurately as possible is our main goal.

REFERENCES

[1] Heart disease prediction system using Naive bayes- International Journal of Advanced Research in Computer and Communication Engineering.
[2] Blake,C.L., Mertz : UCI Machine Learning Database- shttp:/mlearn.ics.uci.edu/databases/heartdisease/
[3] Sellappan Palaniappan, Rafiah Awang, ”Intelligent Heart Disease Pre-diction System Using Data Mining Techniques”, IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008