WEB BASED HEALTH CARE DETECTION USING NAIVE BAYES ALGORITHM

S.Indhumathi,
Research Scholar,
Periyar University,
PGP College of Arts & Science,
Dept of CS& Applications,
TamilNadu, India
Contact No. 9965525880

Mr.G.Vijaybaskar,
Asst. Professor,
Periyar University,
PGP College of Arts & Science,
Dept of CS& Applications,
TamilNadu, India,
Contact No. 9791623336

Abstract—**A major challenge facing healthcare organizations (hospitals, medical centres) 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. 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 are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The main objective of this research is to develop a Intelligent Heart Disease Prediction System using the data mining modelling technique, namely, Naïve Bayes. It is implemented as web based questionnaire application .Based on the user answers, 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. By providing effective treatments, it also helps to reduce treatment costs.**

Keywords-**Data Set, Heart Disease, Input Attributes Naive Bayes.**

I. INTRODUCTION

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. 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."

However, they cannot answer complex queries like "Identify the important preoperative predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", and "Given patient records, predict the probability of patients getting a heart disease."

Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. IHDPS can be further enhanced and expanded. For example, it can incorporate other medical attributes besides the 15 listed and suggestions such as doctor details and medications to patients could also be provided.

II. RELATED WORK

Tremendous works in literature related with heart disease diagnosis using data mining techniques have motivated our work. The researchers in the medical field diagnose and predict the diseases in addition to providing effective care for patients [3] by employing the data mining techniques. The data mining techniques have been employed by numerous works in the literature to diagnose diverse diseases, for instance: Diabetes, Hepatitis, Cancer, Heart diseases and more [4]. A model Intelligent Heart Disease Prediction System (IHDPS) built with the aid of data mining techniques like Decision Trees, Naïve Bayes and Neural Network was proposed in [5]. The problem of identifying constrained association rules for heart disease prediction was studied by Carlos Ordonez [6]. The assessed data set encompassed medical records of people having heart disease with attributes for risk factors, heart perfusion measurements and artery narrowing association rule mining is a major data mining technique, and is a most commonly used pattern discovery method. It retrieves all frequent patterns in a data set and forms interesting rules among frequent patterns. The term Heart disease encompasses the diverse diseases that affect the heart. Coronary heart disease, Cardiomyopathy and Cardiovascular disease are some categories of heart diseases. The term "cardiovascular disease" includes a wide range of conditions that affect the heart and the blood

ISSN: 2278 – 1323

vessels and the manner in which blood is pumped and circulated through the body. Cardiovascular disease (CVD) results in severe illness, disability, and death [10].

III. Health Care Detection

The proposed system uses data mining technique "Naïve Bayes classifier" for the construction of the prediction system. It can predict three diseases which are diabetes, cancer and heart attack. This system involves higher number of data sets and attributes which are directly collected from doctor's information for accurate prediction of the disease. Nearly 300 records for each disease will be collected and stored in the database. 16 medical attributes for heart disease, 9 medical attributes for diabetics and 10 medical attributes for cancer are taken. These datasets will be used for prediction of the heart disease using data mining technique. Since more attributes and more data sets are available, the datasets can be used to predict more accurate occurrence of the disease.

A. *Advantages:*

1. This causes the disease to be predicted more effectively.
2. Moreover this proposed system also consists of various suggestions such as doctor details and prescriptions.
3. Each disease will have different specialists for analyzing the disease. The details of each doctor along with their location for each disease will be given.
4. Cost of visiting the doctor in the initial stage could be avoided since the medications will be prescribed.

IV. Naïve Bayes Implementation

Naive Bayes or Bayes' Rule is the basis for many machine-learning and data mining methods. The rule (algorithm) is used to create models with predictive capabilities. It provides new ways of exploring and understanding data. It learns from the "evidence" by calculating the correlation between the target (i.e., dependent) and other (i.e., independent) variables.

Neural Networks consists of three layers: input, hidden and output units (variables). Connection between input units and hidden and output units are based on relevance of the assigned value (weight) of that particular input unit. The higher the weight the more important it is. Neural Network algorithms use Linear and sigmoid transfer functions. Neural Networks are suitable for training large amounts of data with few inputs. It is used when other techniques are unsatisfactory.

A. *Analyzing the Data Set*

A data set (or dataset) is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the data set in question. It lists values for each of

the variables, such as height and weight of an object or values of random numbers. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows.

The values may be numbers, such as real numbers or integers, for example representing a person's height in centimeters, but may also be nominal data (i.e., not consisting of numerical values), for example representing a person's ethnicity. More generally, values may be of any of the kinds described as a level of measurement. For each variable, the values will normally all be of the same kind. However, there may also be "missing values", which need to be indicated in some way.

A total of 500 records with 15 medical attributes (factors) were obtained from the Heart Disease database lists the attributes. The records were split equally into two datasets: training dataset (455 records) and testing dataset (454 records). To avoid bias, the records for each set were selected randomly.The attribute "Diagnosis" was identified as the predictable attribute with value "1" for patients with heart disease and value "0" for patients with no heart disease. The attribute "Patient ID" was used as the key; the rest are input attributes. It is assumed that problems such as missing data, inconsistent data, and duplicate data have all been resolved.

Here in our project we get a data set from .dat file as our file reader program will get the data from them for the input of Naïve Bayes based mining process.

V. Naive Bayes Implementation in Mining:

I recommend using Probability for Data Mining for a more in-depth introduction to Density estimation and general use of Bayes Classifiers, with Naive Bayes Classifiers as a special case. But if you just want the executive summary bottom line on learning and using Naive Bayes classifiers on categorical attributes then these are the slides for you.

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. If B represents the dependent event and a represents the prior event, Bayes' theorem can be stated as follows.

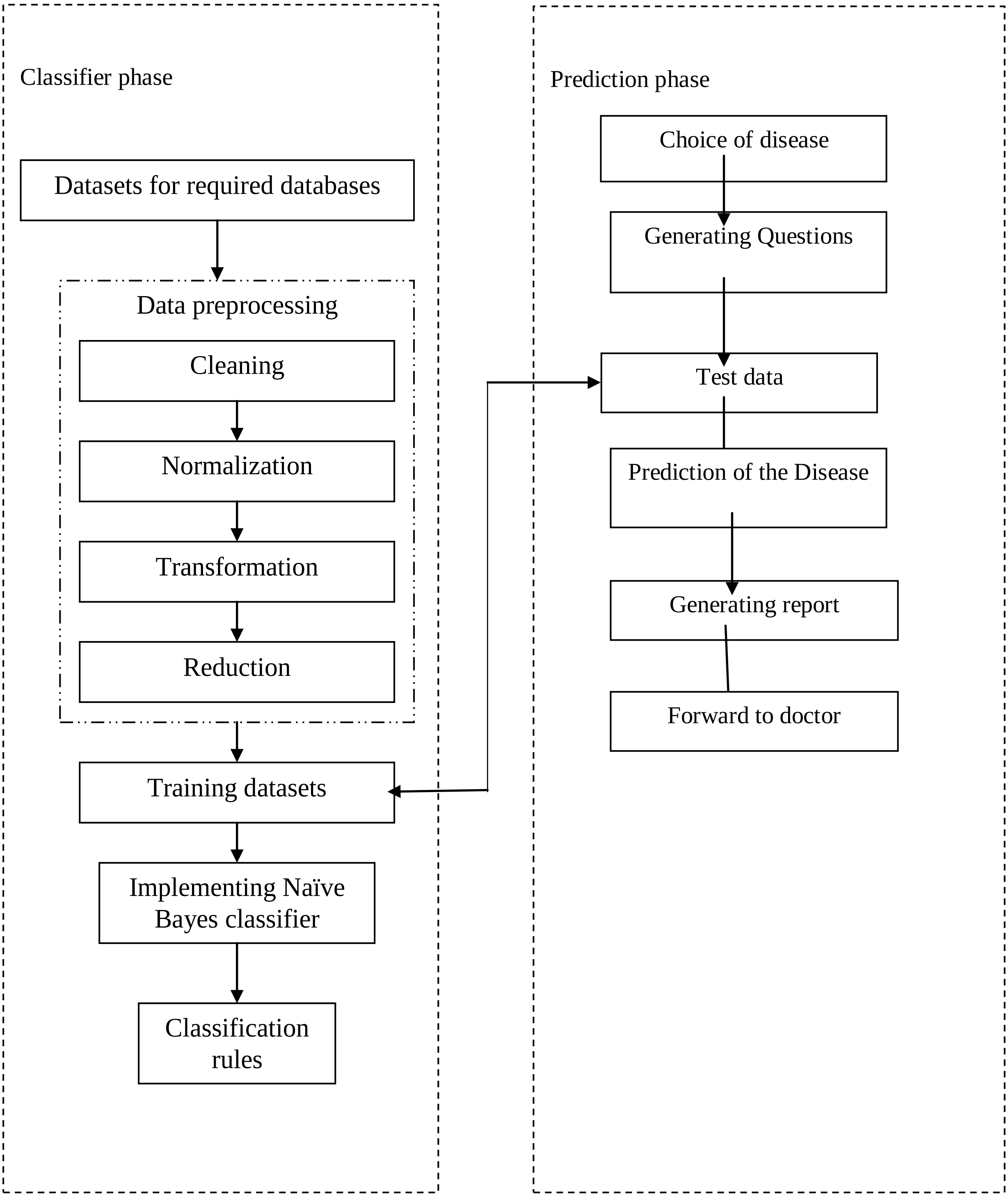
A. *Bayes' Theorem:*

ISSN: 2278 – 1323

Prob(B given A) = Prob(A and B)/Prob(A)

To calculate the probability of B given A, the algorithm counts the number of cases where A and B occur together and divides it by the number of cases where A occurs alone.

VI. SYSTEM DESIGN AND ARCHITECTURE



ISSN: 2278 – 1323

Applying Naïve Bayes to data with numerical attributes and using the Laplace correction (to be done at your own time, not in class)(data with some numerical attributes), predict the class of the following new example using Naïve Bayes classification: with some numerical attributes), predict the class of the following new example using Naïve Bayes classification:

$$\mu = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n-1}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A. Designing the Input attributes

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. However, such standardized answers may frustrate users. Questionnaires are also sharply limited by the fact that respondents must be able to read the questions and respond to them.

Here our questionnaire is based on the attribute given in the data set, so the our questionnaire contains :

1) Input attributes

1. Sex (value 1: Male; value 0 : Female)
2. Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic)
3. Fasting Blood Sugar (value 1: > 120 mg/dl; value 0:< 120 mg/dl)
4. Restecg – resting electrographic results (value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy)
5. Exang – exercise induced angina (value 1: yes; value 0: no)
6. Slope – the slope of the peak exercise ST segment (value 1: unsloping; value 2: flat; value 3: downsloping)
7. CA – number of major vessels colored by floursopy (value 0 – 3)
8. Thal (value 3: normal; value 6: fixed defect; value 7:reversible defect)
9. Trest Blood Pressure (mm Hg on admission to the hospital)
10. Serum Cholesterol (mg/dl)
11. Thalach – maximum heart rate achieved
12. Oldpeak – ST depression induced by exercise relative to rest
13. Age in Year
14. Height in cms
15. Weight in Kgs.

B. Designing Dynamic User Interface

In our Heart disease development the modeling and the standardized notations allow to express complex ideas in a precise way, facilitating the communication among the project participants that generally have different technical and cultural knowledge.

MVC architecture has had wide acceptance for corporation software development. It plans to divide the system in three different layers that are in charge of interface control logic and data access, this facilitates the maintenance and evolution of systems according to the independence of the present classes in each layer. With the purpose of illustrating a Successful application built under MVC, in this work we introduce different phases of analysis, design and implementation of a database and web application.

Ajax (asynchronous JavaScript and XML), or AJAX, is a group of interrelated web development techniques used for creating interactive web applications or rich Internet applications. With Ajax, web applications can retrieve data from the server asynchronously in the background without interfering with the display and behavior of the existing page.

ISSN: 2278 – 1323

In many cases, the pages on a website consist of much content that is common between them. Using traditional methods, that content would have to be reloaded on every request. However, using Ajax, a web application can request only the content that needs to be updated, thus drastically reducing bandwidth usage and load time. The use of asynchronous requests allows the client's Web browser UI to be more interactive and to respond quickly to inputs, and sections of pages can also be reloaded individually. Users may perceive the application to be faster or more responsive, even if the application has not changed on the server side. The use of Ajax can reduce connections to the server, since scripts and style sheets only have to be requested once.

VII. Conclusion and Future Work

Here in our project we get a data set from .dat file as our file reader program will get the data from them for the input of Naïve Bayes based mining process. A total of 500 records with 15 medical attributes (factors) were obtained from the Heart Disease database lists the attributes. The records were split equally into two datasets: training dataset (455 records) and testing dataset (454 records). To avoid bias, the records for each set were selected randomly.As we got a data set, we proceed our work with Implementing Naïve Bayes algorithms.

Prob(B given A) = Prob(A and B)/Prob(A)

To get input from client part we yet prepare a questionnaire (for the patient).In our Heart disease development the modeling and the standardized notations allow to express complex ideas in a precise way through a WEB.

VIII. References

[1] Tzung-I Tang, Gang Zheng, Yalou Huang, Guangfu Shu, Pengtao Wang, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and System Reconstruction Analysis", IEMS Vol. 4, No. 1, pp. 102-108,June 2005.

[2] Hian Chye Koh and Gerald Tan ,"Data Mining Applications in Healthcare", Journal of healthcare information management, Vol. 19, Issue 2, Pages 64-72, 2005.

[3] S Stilou, P D Bamidis, N Maglaveras, C Pappas , "Mining association rules from clinical databases: an intelligent diagnostic process in healthcare" .

[4] A. Bellaachia and Erhan Guven, "Predicting Breast Cancer Survivability using Data Mining Techniques", the Sixth SIAM International Conference on Data Mining , April 22, 2006.

[5] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS, Vol.8 No.8, August 2008.

[6] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules,"Seminar Presentation at University of Tokyo, 2004.

[7] Agrawal, R., Imielinski, T. and Swami, A, 'Mining association rules between sets of items in large databases' .

[8] Douglas Burdick, Manuel Calimlim, Johannes Gehrke, "MAFIA: A Maximal Frequent Itemset Algorithm for Transactional Databases",Proceedings of the 17th International Conference on Data Engineering.

[9] "Heart disease" from http://en.wikipedia.org/wiki/Heart_disease.

[10] "Hospitalization for Heart Attack, Stroke, or Congestive Heart Failure among Persons with Diabetes", Special report: 2001 – 2003, New Mexico.

[11] Cornell, D, "A Vertical Partitioning Algorithm for Relational Databases", Proceedings of the Third International Conference on Data Engineering, Los Angeles, CA, February.

ISSN: 2278 – 1323