International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 2 | ISSN : 2456-3307

Smart Health Prediction System Using Data Mining

Nikita Kamble*[1], Manjiri Harmalkar[2], Manali Bhoir[3], Supriya Chaudhary[4]

Information Technology, University of Mumbai, Mumbai, Maharashtra, India

ABSTRACT

The paper presents an overview of the data mining techniques with its applications, medical,and educational aspects of Clinical Predictions. In medical and health care areas, due to regulations and due to the availability of computers, a large amount of data is becoming available. Such a large amount of data cannot be processed by humans in a short time to make diagnosis, and treatment schedules. A major objective is to evaluate data mining techniques in clinical and health care applications to develop accurate decisions. It also gives a detailed discussion of medical data mining techniques which can improve various aspects of Clinical Predictions. It is a new powerful technology which is of high interest in computer world. It is a sub field of computer science that uses already existing data in different databases to transform it into new researches and results. It makes use of machine learning and database management to extract new patterns from large data sets and the knowledge associated with these patterns. The actual task is to extract data by automatic or semi-automatic means. The different parameters included in data mining include clustering, forecasting, path analysis and predictive analysis.

Keywords: Smart Health Prediction System, Data Mining, Clinical Predictions, Semi-Automatic Means, Clustering, Forecasting, Predictive Analysis.

I. INTRODUCTION

In recent years, with the development of computer technologies, the puzzles with respect to Bayesian statistics and posterior distribution have been better solved. Meanwhile, Bayesian statistics successfully apply to economic, sociology and some other fields. In medical fields, the foreign scholars have solved some medical problems that are hard to be settled in classic statistics by using Bayesian classification. Naive Bayes is one of the most popular classification technique introduced by Reverend Thomas Bayes. Without any additional data, classification rules are generated by the training samples themselves.

"Smart Health Prediction System" is the computerization of medical information to support and optimize

(1) administration of health services
(2) clinical care
(3) medical research
(4) training

It is the application of computing and communication technologies to optimize health information processing by collection, storage, effective retrieval (in due time and place).The proposed system is mainly used by the all the people where confidentiality and integrity of the data has utmost importance. Computer assisted information retrieval may help support quality decision making and to avoid human error. Imagine a doctor who has to examine 5 patient records; he or she will go through them with ease. If the number of records grows with a time constraint, it is almost certain that the accuracy with which the doctor delivers the results will not be as high as the ones obtained when he had only five records to be analyzed.

II. METHODS AND MATERIAL

1. Existing system

Everyone is a patient at some time or another, and we all want good medical care. We assume that doctors

are all medical experts and that there is good research behind all their decisions. However, that cannot always be the case. Nevertheless, they cannot possibly commit to memory all the knowledge they need for every situation, and they probably do not have it readily available.

Even if they did have access to the massive amounts of data needed to compare treatment outcomes for all the diseases they encounter, they would still need time and expertise to analyse that information and integrate it with the patient's own medical profile. But this kind of in-depth research and statistical analysis is beyond the scope of a physician's work. They want a doctor who will talk to them, listen to what they say and give them advice about how to get better and protect their health in the future. In many cases, the wish for a prescription is secondary to the wish of being cared for

Disadvantage of an existing system would be that the patients have to visit the doctor in person and still does not get proper treatment, as the doctors are unable to predict the exact disease. Human error can be avoided with the help of computer assisted quality decision making. It is poor when there are huge amounts of data to be classified. In addition, efficiency and accuracy of decisions will decrease when humans are put into stress and immense work. Imagine a doctor who has to examine five patient records; he or she will go through them with ease. However, if the number of records increases with a time constraint, it is almost certain that the accuracy with which the doctor delivers the results will not be as high as the ones obtained when he had only five records to be analysed.

2. Proposed System

To overcome the drawback of existing system we have developed smart health prediction System. The system design of the proposed system is shown in Figure:

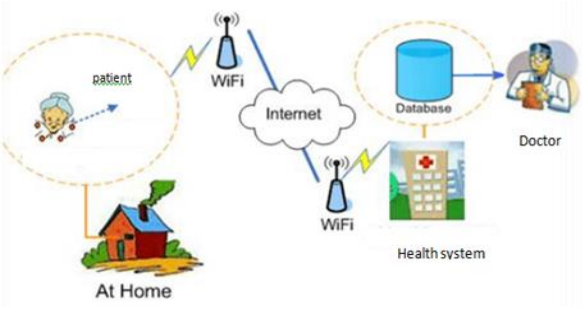


Figure 1. Proposed System Architecture

We have developed an expert system called Smart Health Prediction system, which is used for simplifying the task of doctors. A system checks a patient at initial level and suggests the possible diseases. It starts with asking about symptoms to the patient, if the system is able to identify the appropriate disease then it suggests a doctor available to the patient in the nearest possible area. If the system is not sure enough, it asks some queries to the patients, still if the system is not sure then it will display some tests to the patient. Based on available cumulative information, the system will display the result. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's symptoms and based on the database of several patients medical record, algorithm (Naïve Bayes) is applied for mapping the symptoms with possible diseases.

This system not only simplifies task of the doctors but also helps the patients by providing necessary help at an earliest stage possible.

3. Features of System

Patient Module

Patient Login: - Patient Login to the system using his ID and Password.

Patient Registration: -If Patient is a new user he will enter his personal details and he will get a user Id and password through which he can login to the system.

Disease Prediction: - Patient will specify the symptoms caused due to his illness. The system will ask certain questions regarding his illness and then predict the disease based on the symptoms specified by patient and the system will also suggest doctors based on the disease.

Search Doctor:- Patient can search for doctor by specifying name, address or type.

Feedback:-Patient will give feedback this will be reported to the admin.

Doctor Module

Doctor Login: - Doctor will access the system using his User ID and Password.

Patient Details : Doctor can view patient's personal details.

Notification: Doctor will get notification on how many people had accessed the system and what all are the diseases predicted by the system.

In order to implement the above-discussed Proposed System, the following diagrams helps us to understand the design of the proposed system.

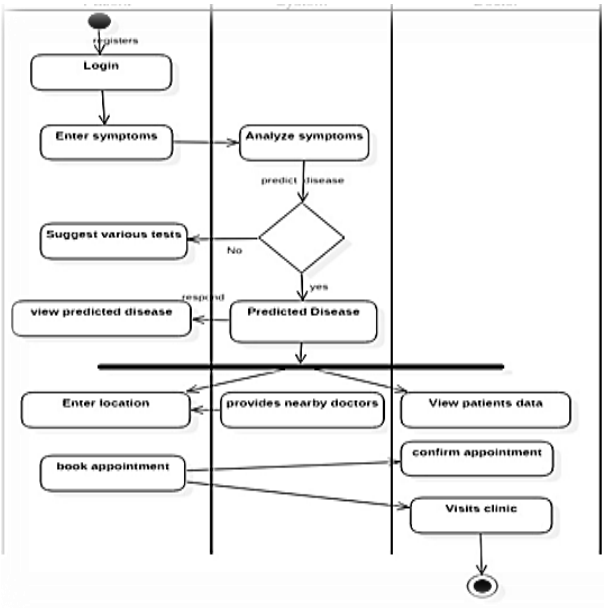


Figure 2. Activity Diagram for health prediction System

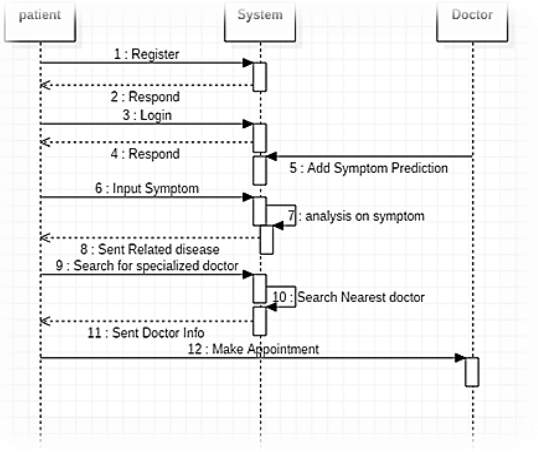
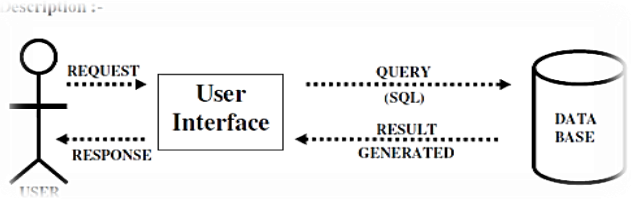


Figure 3. Sequence Diagram for smart health prediction System

With the help of these designs, the system is designed and implemented which helps in automation of the health prediction system.

4. Working of The System



- According to the diagrams, it is a two tier architecture. We provide a form that shows a list of symptoms. The user will input those symptoms that he experiences .
- On the basis of selected symptoms the system will generate related disease. The system will show another form that contain some queries if the information for the disease is not enough.
- On the basis of the information a query is generated and the data base responses to that query.

5. Data Mining Architecture

Data Mining used in the field of medical application can exploit the hidden patterns present in voluminous medical data which otherwise is left undiscovered.

The term Knowledge Discovery in Databases, or KDD for short, refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods.

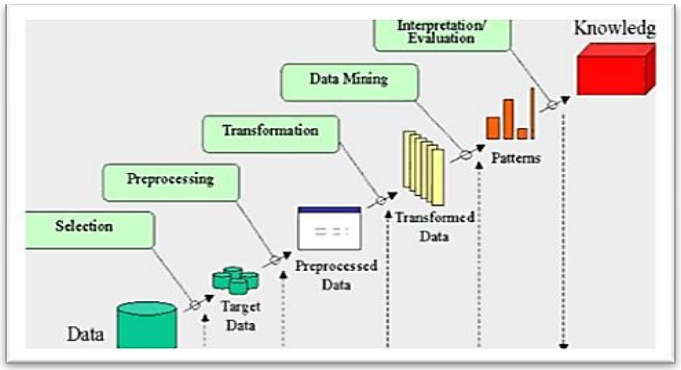


Figure 4. KDD Architecture

Naive Bayes Algorithm:

The proposed system uses data mining technique "Naïve Bayes classifier" for the construction of the prediction system. This system involves higher number of data sets and attributes which are directly collected from doctor's information for accurate prediction of the disease.

"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 learns from the "evidence" by calculating the correlation between the target (i.e., dependent) and other (i.e., independent) variables.

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) There is a specialist appointed for each disease predicted. The details of each doctor along with their location for each disease will be given.
4) Doctor's consultation cost can be avoided at an initial stage prescribed medicines are displayed in detail.

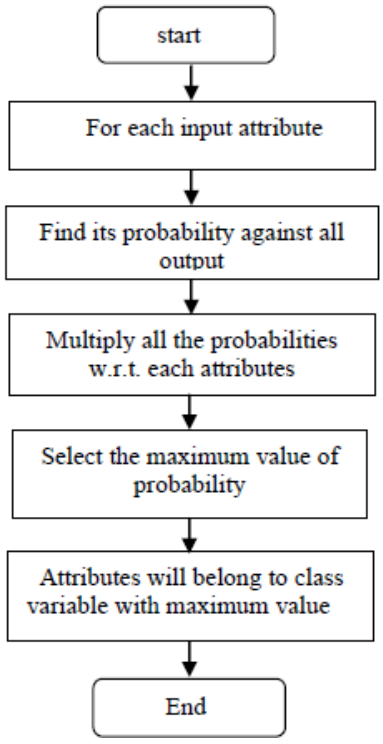


Figure 5. Flowchart for Naive Bayes

Following steps are implemented the naïve bayes algorithm:-

1. Baye's therom is P(c|x)=P(x|c)P(c)/P(x)
   - ✓ Where, P(c|x) is the posterior probability of class (target) given predictor (attribute).
   - ✓ P(c) is the prior probability of class.
   - ✓ P(x|c) is the probability of predictor given class likelihood.
   - ✓ P(x) is the prior probability of predictor.
2. A set of cases was taken in program was trained with the data sets such that the probabilities of all the classes with all the conditions were calculated.
3. Result was gather in database and when the test data was given we got the probabilities for the different classes for the given symptom values on the basis of which we implies that the patient fell into the class with the highest probability. Hence it is Naïve Bayes" classification.
4. By using this flow chart we can easily conclude that the patient has been suffered from particular disease or not .we will test this from the incoming attribute which belongs to class variable with maximum value .
5. We first compute all possible individual probabilities conditioned on the target attribute of particular disease contained all probabilities of attribute of that disease.
6. Compute the possible probabilities for all condition decide that the p has split up into two cases one for Y and second for N.
7. Hence, if the argument of probability of P1 is greater than P2 then patient is not having the disease.
8. In the following figure, steps of implementation of naïve bayes on the patient dataset are shown.

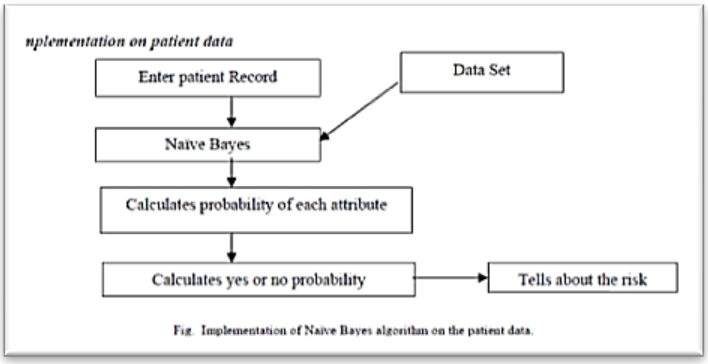


Figure 6. Implementation of naïve bayes on the patient dataset

6. Technologies Used

Eclipse IDE:

For user interface, Eclipse IDE(Integrated Development Environment) will be used for will be used for designing the Graphical User Interface (GUI).

Java:

Java will be used for connecting various components of user interface to database system.

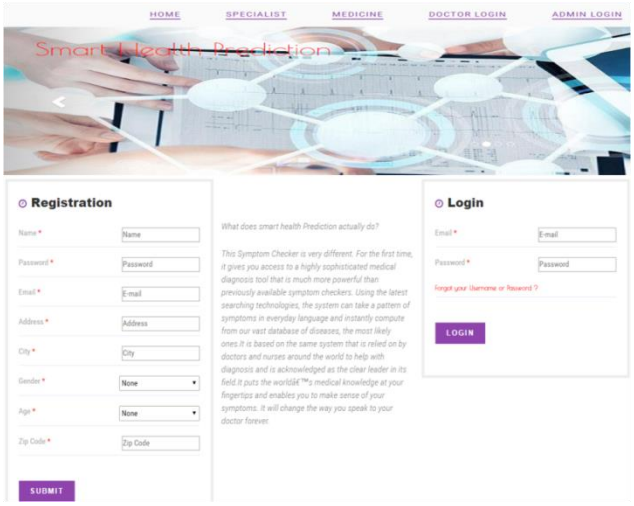
Navicat Mysql:

MYSQL is used as a database at the web server. In this system, server used is tomcat server.

Disease, Doctor and Patient database is created with the help of Navicat Mysql.

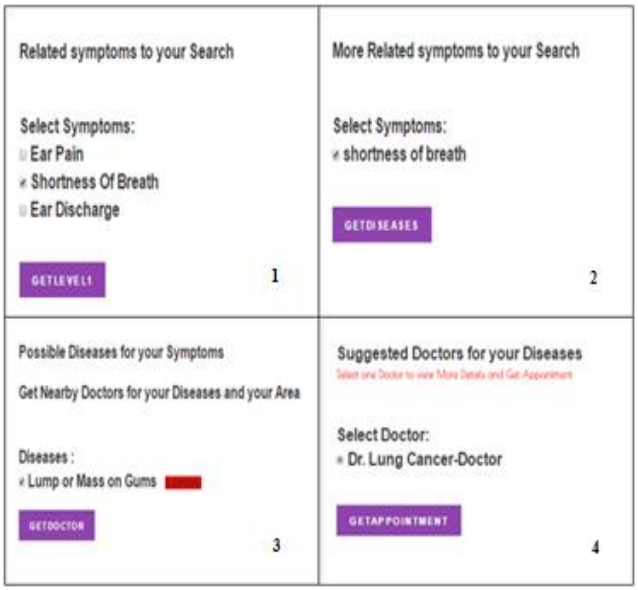
It provides an intuitive and powerful graphical interface for database management, development, and maintenance.

III. RESULTS AND DISCUSSION

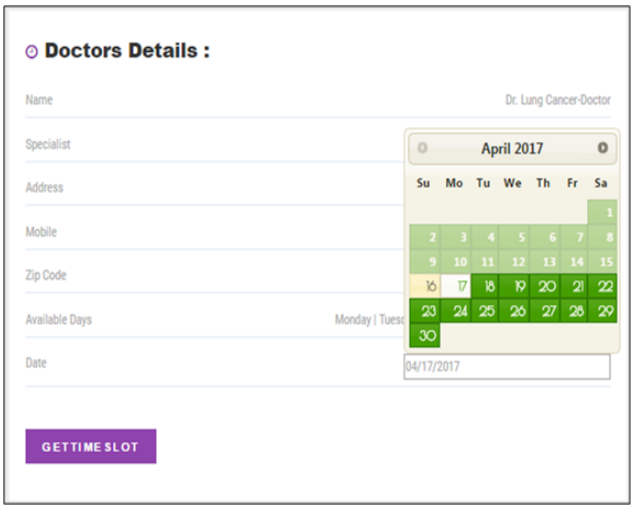


Patient Registration : - If Patient is a new user, he will enter his personal details, he will get a user Id, and password through which he can login to the system.

Patient Login: - Patient Login to the system using his ID and Password.



Disease Prediction: - Patient will specify the symptoms caused due to his illness. The system will ask certain symptoms regarding his illness and then predict the disease based on the symptoms specified by patient.



Search Doctor :- Patient can search for doctor when the disease is predicted by the system which will provide him with the address of any nearby doctor.

Get Appointment:-Patient will choose an appointment date for the selected doctor.

IV. FUTURE SCOPE

Hidden knowledge will be extracted from the historical data in the proposed system, by preparing datasets by applying Apriori algorithm. Predicting smart health can be done only if system responds that way. These datasets will be compared with the incoming queries and the final report will be generated using Association Rule Mining. Since this proposed methodology will work on real historical data, it will provide accurate and efficient results, which will help patients, get diagnosis instantly. More work can be done in the future by using more data set related to heart diseases and by using different data reduction methods to improve the classification. For better accuracy and prediction of heart diseases the datasets that will be used must be quality oriented and free from outliers, inconsistencies and missing values.

This web application can be further enhanced in an Android app. This will be available to users on mobile basis and its use can be further increased. Also, feature like getting the doctor online on chat so that patients can directly talk to the concerned doctors. The modules doing cancer analysis can be integrated to find how close the person associated with cancer is. This will make this web application predictable in true sense.

V. CONCLUSION

Data mining can be beneficial in the field of medical domain .However privacy, security and unable to log into the account are the big problems if they are not addressed and resolved properly. It describes about the proposal of hybrid data mining model to extract classification knowledge for aid of various diseases in clinical decision system and presents a framework of the tool various tools used for analysis.

Sometimes the situation occurs when you need the doctor's help immediately, but they are not available due to some reason. In our project, we have designed a new health prediction system, which is an online system, and various patients from any locations can view it. Our system comprises of main components such as patient login, enter symptoms in the System, and prescribe medicines, suggested nearby doctor. The application takes the input of various symptoms from the patient, does the analysis of the entered symptoms, and gives appropriate disease prediction. Our system allows the users to get analysis on the symptoms they give for predicting the disease they are suffering from.

Sometimes the situation occurs when you need the doctor's help immediately, but they are not available due to some reason. Thus, it allows the users to get analysis on the symptoms they give for predicting the disease they are suffering from.

VI. REFERENCES

[1]. MANASWINI PRADHAN (International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 12, December 2014 Data Mining& Health Care: Techniques Of Application

[2]. DR.B.SRINIVASAN, International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 03 | Mar-2016 A Study On Data Mining Prediction Techniques In Healthcare Sector K.PAVYA, International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 03 | Mar-2016 A Study On Data Mining Prediction Techniques In Healthcare Sector

[3]. K.VEMBANDASAMY, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015 Heart Diseases Detection Using Naive Bayes Algorithm

[4]. Kalyan Netti, International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, Special Issue 6, August 2015 A web Implementation of Naive Bayes Classifier

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

[6]. Y. Zhang and B. F. L. Xiao, "Web page classification based on a least square support vector machine with latent semantic analysis", in Proc. of the 5th Int. Conf. on Fuzzy Systems and Knowledge Discovery2008, Vol. 2, pp. 528-532,

[7]. O. Kwon and J. Lee, "Web page classification based on k-nearest neighbour approach", in Proc. of the 5th Int. Workshop on Information Retrieval with Asian languages, Hong Kong, China, 2000, pp. 9-15.

[8]. S. Dehghan and A. M. Rahmani, "Aclassifier-CMAC neural network model for web mining", in Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology2008, Vol. 1, pp. 427-431.

[9]. S. Dou, S. Jian-Tao , Y. Qiang and C. Zheng, "A comparison of implicit and explicit links for web page classification", in Proc. of the 15th International Conference on World Wide Web, Edinburgh, Scotland, 2006 , pp. 643–650.

[10]. S. Zhongzhi and L. Xiaoli, "Innovating web page classification through reducing noise", Journal of Computer Science and Technology, Vol. 17, no. 1 , pp. 9–17, Jan. 2002