SMART HEALTH CARE USING DATA MINING

Pinky Saikia Dutta[1], Shrabani Medhi[2], Sunayana Dutta[3], Tridisha Das[4], Sweety Buragohain[5]

Abstract

Data mining 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. In this data-rich world, people are running out of information. This can be a matter of risk for the person who needs immediate remedies regarding their poor health. To unfold this hurdle, the concept of data mining is the best suited. Here, the traditional approaches have been replaced by smart technologies. The main purpose of data mining application in healthcare system is to develop an automated tool for identifying and disseminating relevant healthcare information. In this system, we have presented a web based application for Predicting diseases based on user input symptoms. It predicts probable diseases by mining data sets and provides remedial solutions for Effective Treatment.

Keywords: Apriori Algorithm, Data mining, FP Growth algorithm, Web Application.

I. INTRODUCTION

Health is one of the most important assets of our life which directly reflects in any form of progress or development. In today's hustle and bustle of life, most of the people neglect this asset which may be due to lack of time and complexity in the vast data available over the Internet. Data mining has introduced various techniques which would overcome this problem and assist us to emphasize on both health and work simultaneously. In present era, Data Mining is becoming popular in healthcare field because there is a need of efficient analytical methodology for detecting unknown and valuable information in health data. In health industry, Data Mining provides several benefits such as detection of the fraud in health insurance, availability of medical solution to the patients at lower cost, detection of causes of diseases and identification of medical treatment methods. It also helps the healthcare researchers for making efficient health care policies, constructing drug recommendation systems, developing health profiles of individuals etc.

II. SURVEY OF RELATED WORK

M. Durairaj, V. Ranjani presented a paper that aims to make a detailed study report of different types of data mining applications in the healthcare sector and to reduce the complexity of the study of the healthcare data transactions. Also presents a comparative study of different data mining applications, techniques and different methodologies applied for extracting knowledge from database generated in the healthcare industry. Finally, the existing data mining techniques with data mining algorithms and its application tools which are more valuable for healthcare services are discussed in detail[1]. Divya Tomar and Sonali Agarwal have presented a brief introduction of data mining techniques such as classification, clustering, association, regression in health domain and their advantages and disadvantages. This survey also highlights applications, challenges and future issues of Data Mining in healthcare[2].

R.Karthiyayini , J.Jayaprakash have presented a paper which analyses the various results generated by implementing the Apriori algorithm of Association technique. The focus of this paper is to provide precise information about chronic diseases for public[3].

Priyanka Vijay Pawar, Megha Sakharam Walunj, and Pallavi Chitte presents a methodology to predict diseases based on user input symptoms. They have built a prototype to

demonstrate the efficiency of these methods which will inform users about the disease they are suffering from. It predicts probable diseases by mining data sets and provides suggested doctors and remedial solutions[4].

Gitanjali J, C.Ranichandra ,M.Pounambal has presented a method for identifying frequency of diseases in particular geographical location for a given period of time using Apriori data mining technique based on association rules is proposed[5].

Abdullah Saad Almalaise Alghamdi discussed about the importance of data mining using medical data then discussion of general data mining techniques has been presented. Furthermore, methodology describes the conceptual model for the extraction of rules on medical databases finally result can guide the relationship between the different attributes presented in the data. In this regard, they applied FP growth algorithm for extracting rules from the medical data[6].

III. METHODOLOGY

The core objective of our project is to develop a web application using data mining concept accompanied by JSP (Java Server pages) technology and MYSQL. The whole process can be termed as "knowledge discovery process, (KDD)". This is because here we need to predict the disease for user input symptoms where the predicted disease is in the form of information or knowledge. Following Fig:1 shows the steps carried out to predict the probable disease for inputted patient symptoms:
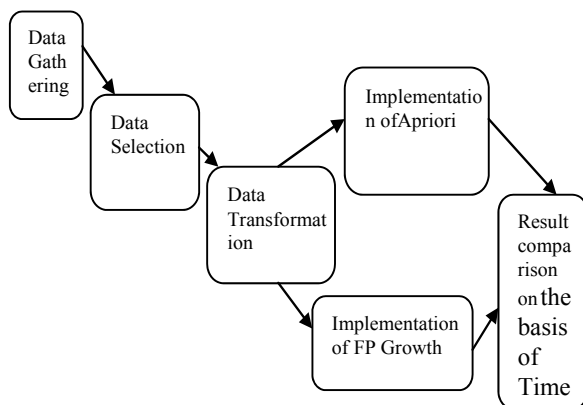
PREPROCESSING :

Data gathering and selection:
The raw data has been collected from the world wide web and data relevant to our purpose has been selected for further processing.

• Data transformation:
After the first step the data is transformed into xls file so as to form a standard database. This xls file is extracted or read using file handling concept and stored in a database using MYSQL . From the xls file only those information are read which are associated with the basic objective of our intended application. For example various stops words, verbs and adjective irrelevant to the application are kept behind and only the key meanings are read from the xls file so that it becomes easier for the application to implement the algorithms i.e Apriori and FP Growth over the disease-symptom database.

We have used Apriori and FP Growth algorithm for predicting the disease for a given set symptoms. These symptoms are provides by an user as inputs. On accepting these inputs the application executes these algorithms over them by accessing the database created using MYSQl in step 2 during preprocessing stage.

Apriori Algorithm:

The Apriori algorithm is an influencial algorithm for mining frequent item sets for Boolean association rules.

Apriori is a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation, and group of candidates are tested against the data).

Apriori is designed to operate on database containing transactions, (for example: collection of items bought by customers).

Key concepts:

Frequent item sets: All the sets which contain the item with the minimum support(denoted by Li for ith item set).

Apriori property: Any subset of frequent item set must be frequent.

Join operation: To find Lk a set of candidate k item set is generated by joining Lk-1 by with itself.

Here we have implemented the Apriori algorithm by generating only one candidate set. This is

because here our motive is to predict only one disease for a set of inputted symptoms.

FP Growth Algorithm:

FP Growth stands for frequent pattern growth.

It is a scalable technique for mining frequent patterns in a database. FP Growth is a two step procedure.

Step1: Build a compact data structure called the FP-Tree.

(Build using two passes over the data set).

Step 2: Extracts frequent item sets directly from the FP-Tree.

FP tree Generation: Item sets are considered in order of their descending value of support count. To facilitate tree traversal, an item header table is built so that each item points to its occurrences in the tree via a chain of node links.

Frequent item generation: Frequent items are directly extracted from the FP tree on the basis of maximum frequency

## IV. ARCHITECTURE OF THE APPLICATION
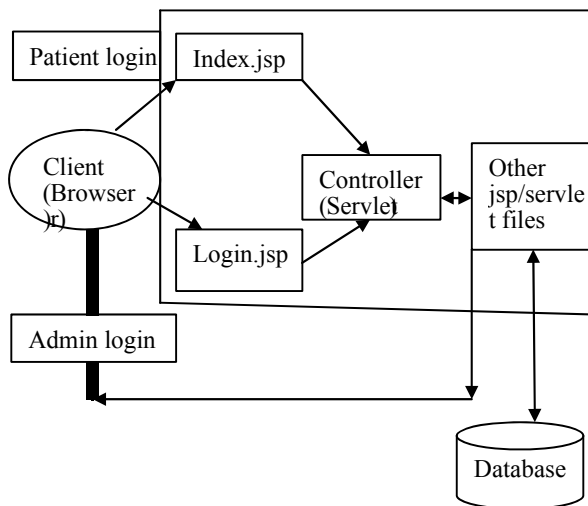
The application consists of two login pages:

- Patient login page i.e. the user interface for the patient.
- Admin login page i.e. the user interface for the administrator.

Here we have made use of softwares like eclipse which is an integrated development environment for any programming language and jsp which is server side programming technology. The sever type used here is Apache Tomcat server and JDBC for connecting java language with

relational database management system like MySql.

V. RESULT
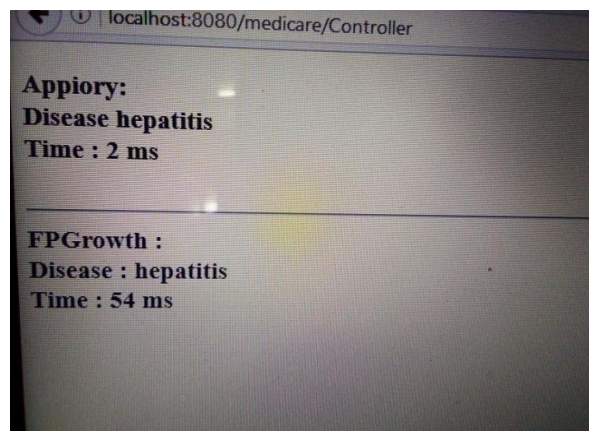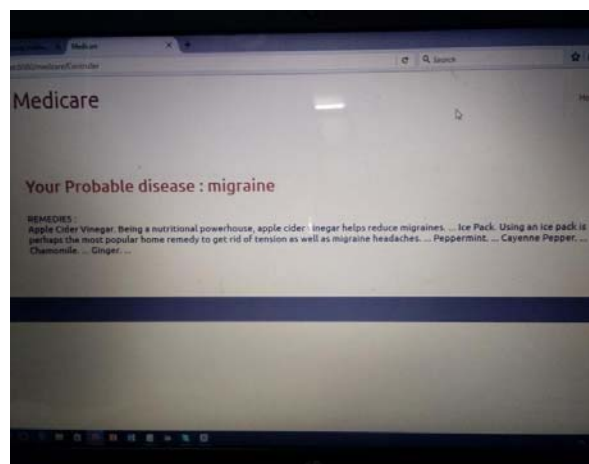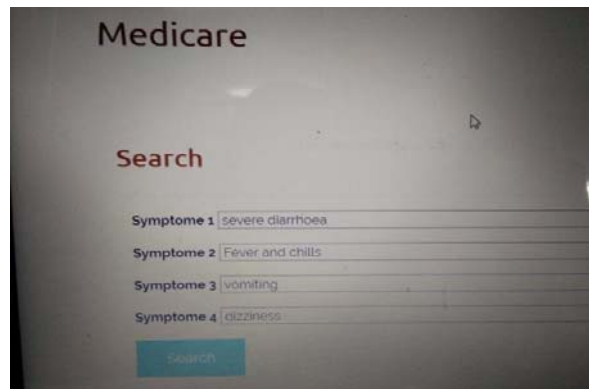
VI. SCOPE OF FUTURE STUDY

This system can be accessed by others in future to make android apps, or can be embedded in other applications locally available doctors could be referred to the patients. Further, the software we made may be extended, by adding a link, for buying medicines online, for predicted diseases, prescribed by doctors. Also, features like, detecting the causes of the detected diseases can be added. For more concern of the patients, the software can have features, saying what things

should be avoided by the patients, during the illness period.

VII.CONCLUSION

The system would drastically reduce the human effort, reduce the cost and time constraint in terms of human resources and expertise, and increase the diagnostic accuracy. The prediction of diseases using Data Mining applications is a challenging and risky task as the data found are noisy, irrelevant and massive too. In this scenario, data mining tools come in handy in exploring of knowledge of the medical data and it is quite interesting.

REFERENCES

[1]. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 10, OCTOBER 2013. "Data Mining Applications In Healthcare Sector: A Study".

[2]. Indian Institute of Information Technology, Allahabad, India. "A survey on Data Mining approaches for Healthcare".

[3]. Assistant Professor, Department of Computer Applications, Anna University (BIT Campus), Trichy, India P.G. Student, Department of Computer Applications, Anna University (BIT Campus), Trichy, India. "Association Technique on Prediction of Chronic Diseases Using Apriori Algorithm".

[4]. Department Of Computer Engineering, Ramrao Adik Institute of Technology Nerul,Navi Mumbai. "Estimation based on Data Mining Approach for Health Analysis".

[5]. School of Information Technology and Engineering, VIT UNIVERSITY, Vellore-632014, Tamil Nadu, India. "APRIORI algorithm based medical data mining for frequent disease identification".

[6]. Chairman of CIT Faculty of Computing and Information Technology King Abdulaziz University, Jeddah Kingdom of Saudi Arabia. "Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data".