

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

Association Technique on Prediction of Chronic Diseases Using Apriori Algorithm

R.Karthiyayini¹, J.Jayaprakash²

Assistant Professor, Department of Computer Applications, Anna University (BIT Campus), Trichy, India¹

P.G. Student, Department of Computer Applications, Anna University (BIT Campus), Trichy, India²

ABSTRACT: Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. Preventive Health care knowledge is essential for clinical and administrative decision making. Chronic diseases are the important cause of death in the world. Some most common chronic diseases are cancer, diabetes cardiovascular disease and chronic respiratory disease. In this paper proposed to discover the association between chronic diseases. Association rule mining greatly helps to identify trends and patterns from huge databases. This paper 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.

KEYWORDS: Chronic diseases, Apriori algorithm, Association technique, Data mining.

I. INTRODUCTION

Enormous gathered data in Health care Information society are scattered with different archive systems which are not connected with one another. This unorganized data leads to delay in monitoring, improper planning, defocus the analysis which leads to inaccuracy in decision making. The purpose of this study is to review the association between the causes for chronic diseases. This paper focuses on various models and techniques used in data mining for health care and its applications for better health policy-making and in decision making. It provides recommendation for future research in the application of data mining in chronic diseases.

Data mining is a technique that mines data effectively to get useful information. Data mining plays a crucial role in mining of healthcare data. Healthcare data can be collected from various hospitals. The assimilated data can be used to analyze the patient reports which help in identifying the patterns present in the databases which further helps to get information about various diseases present, their symptoms, causes, remedies and precautions that can help to prevent the occurrence of various diseases. Medical researchers have been interpreting data from time to time and data gets modified almost every day. The main concern is to be updated about the changes in patterns of data. Data mining techniques help in achieving this objective. Apriori association technique has been proven to be effective in finding various trends in healthcare databases. 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. Data mining is the essential process of discovering hidden and interesting patterns from massive amount of data where data is stored in data warehouse, OLAP (on line analytical process), databases and other repositories of information. This data may reach to more than terabytes. Data mining is also called (KDD) knowledge discovery in databases, and it includes an integration of techniques from many disciplines such as statistics, neural networks, database technology, machine learning and information retrieval, etc

II. DATA MINING IN HEALTH CARE SECTOR

Healthcare sector is a massive area which deals with data about hospitals, patients, doctors, medical devices and equipment's. The management of large health care data poses a great challenge to the researchers. The usage of data mining and machine learning techniques has revolutionized the healthcare organizations. The field of data mining

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

helps to discover hidden patterns by bringing a set of machine learning tools and techniques. It is useful in evaluating the effectiveness of medical treatments. Data mining techniques like classification, association, clustering are applied to healthcare datasets to analyses data to improve health policy-making, early detection of disease outbreaks and preventing the occurrence of various diseases. Data mining provides healthcare authorities an additional source of knowledge for effective decision-making. The information provided by data mining methods can help healthcare insurers detect fraud and abuse and healthcare organizations can make better customer relationship management decisions. Further physicians can identify effective treatments with best practices and patients will receive better and more affordable healthcare services. Data Mining techniques are also used to analyze the various factors that are responsible for diseases for example type of food, different working environment, education level, living conditions, availability of pure water, health care services, cultural ,environmental and agricultural factors.

III. LITRATURE REVIEW

[1]Prasanna et al., discussed healthcare management and an overview of how data mining helps in management of healthcare data. Authors tell about current trends and prominent models for detection of various diseases. Various types of data used in hospitals like HL7, EHR, EMR, ENR etc. are discussed efficiently. They draw attention towards the new challenges faced by data mining to aid in healthcare management. Data mining helps in the detection of fraud and abuse, healthcare resource management and diagnosis and treatment of various diseases. It also helps healthcare organizations in making better customer relationship management decisions.

[2]Ms.Shweta et al., focused on using Apriori Algorithm for frequent item set mining. Authors have discussed the problem of frequent itemset mining and have addressed the mining using Apriori Algorithm with an example. Authors have implemented Apriori Algorithm on a bank dataset in Weka and then a comparison is made between the results of Apriori Algorithm, Predictive Apriori Algorithm and Tertius Algorithm. In future, these algorithms can also be used in other domains. These algorithms can be combined to get improved algorithm which can be used in any real- time application. In association rule mining is used to generate strong association rules by executing Apriori Algorithm on real time datasets. In this paper, authors have exemplified AprioriAlgorithm on a dataset by finding frequent itemsets and then generating association rules from frequent sets. In each iteration, frequent patterns are identified by using the candidate generation and pruning steps. At every step, it is taken care that “Apriori Property” of the algorithm is not violated. With the implementation of the algorithm they predict the occurrence of different diseases in a particular area and what type of people are affected by a particular disease is found out.

IV. ARCHITECTURE

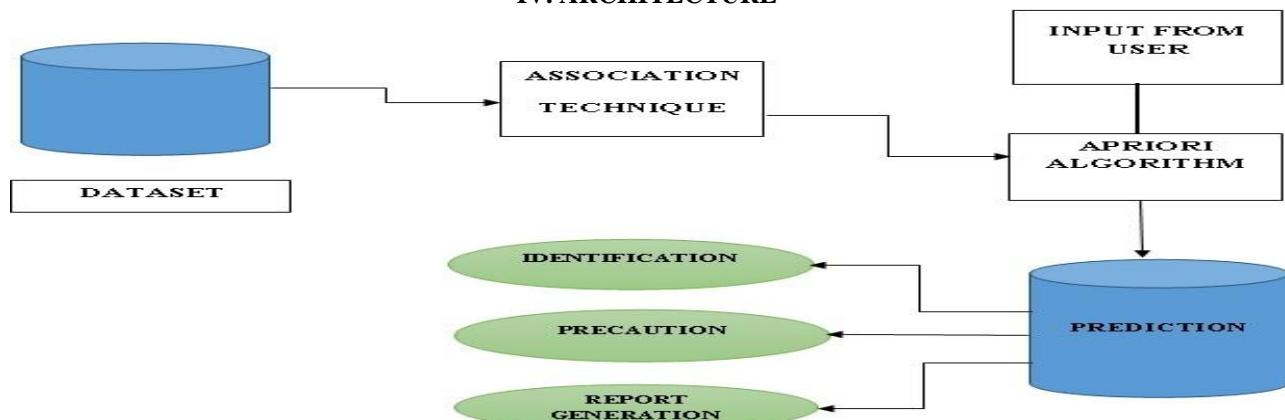


Fig 4.1 Architecture diagram of Proposed system

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

IV. APRIORI ALGORITHM

It is the fundamental and most important algorithm for mining frequent itemsets. It was first given by Agrawal and Srikant in 1994. It is a level wise algorithm which works in an iterative fashion to discover all frequent itemsets in a database. It uses prior knowledge of frequent itemsets properties. Frequent itemsets are the sets of items that satisfy minimum support threshold. This algorithm takes only categorical input and associates attributes present in the dataset. There is a property associated with this algorithm called "Apriori Property" which states that any subset of frequent itemsets is also a frequent itemset. For example, if $\{x,y,z\}$ is a frequent set then the sets $\{ \{x\}, \{y\}, \{z\} \}$, $\{ \{x,y\}, \{x,z\}, \{y,z\} \}$ must also be frequent. The execution of this algorithm is organized in two phases. In the first stage, the candidates are generated and in the next phase frequent itemsets are generated. The generated large itemsets are used to produce association rules from database.

Notation: L_k , set of large k -itemset

1. $L_1 = \{ \text{large 1-itemsets} \}$
2. for($k=2$; $L_{k-1} \neq \emptyset$; $k + +$) do begin
3. $C_k = \text{apriori-gen}(L_{k-1})$; // new candidates
4. forall records $r \in D$ do begin
5. $C_r = \text{subset}(C_k, r)$; // candidates contained in r
6. forall candidates $c \in C_r$ do
7. $c.\text{count} + +$
8. end
9. $L_k = \{ c \in C_k | (c.\text{count}/\text{size}(D)) > \text{minsup} \}$
10. end
11. Answer = $L = k L_k$

VI. ASSOCIATION MINING

Association Mining is one of the most important data mining's functionalities and it is the most popular technique has been studied by researchers. Extracting association rules is the core of data mining. It is mining for association rules in database of sales transactions between items which is important field of the research in dataset. The benefits of these rules are detecting unknown relationships, producing results which can perform basis for decision making and prediction. The discovery of association rules is divided into two phases [10, 5]: detection the frequent itemsets and generation of association rules. In the first phase, every set of items is called itemset, if they occurred together greater than the minimum support threshold, this itemset is called frequent itemset. Finding frequent itemsets is easy but costly so this phase is more important than second phase. In the second phase, it can generate many rules from one itemset as in form, if itemset $\{I_1, I_2, I_3\}$, its rules are $\{I_1, I_2, I_3\}$, $\{I_2, I_1, I_3\}$, $\{I_3, I_1, I_2\}$, $\{I_1, I_2, I_3\}$, $\{I_1, I_3, I_1\}$, $\{I_2, I_3, I_1\}$, number of those rules is n^{2-1} where $n = \text{number of items}$. To validate the rule (e.g. $X \rightarrow Y$), where X and Y are items, based on confidence threshold which determine the ratio of the transactions which contain X and Y to the transactions $A\%$ which contain X , this means that $A\%$ of the transactions which contain X also contain Y . minimum support and confidence is defined by the user which represents constraint of the rules. So the support and confidence thresholds should be applied for all the rules to prune the rules which it values less than thresholds values. The problem that is addressed into association mining is finding the correlation among different items from large set of transactions efficiency. The research of association rules is motivated by more applications such as telecommunication, banking, health care and manufacturing, etc.

VII. OBJECTIVES

The main goal of this paper is to provide people know the details about the treatment for chronic diseases. The good news is that chronic disease can be prevented or controlled through 1) regular participation in physical activity, 2) eating healthy, 3) not smoking, and 4) avoiding excessive alcohol consumption.

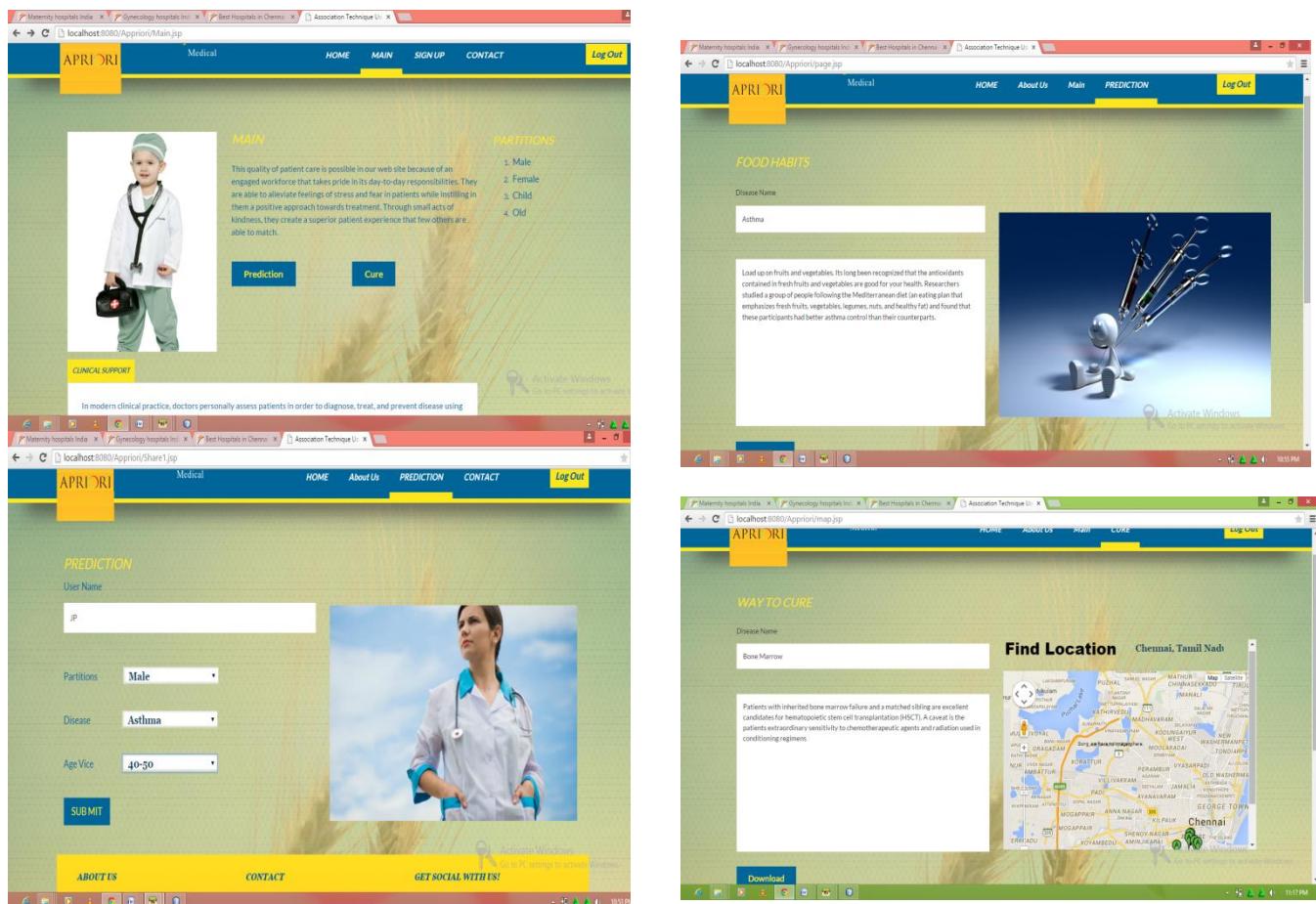
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

There are several advantages of using this system. A few of them are listed below: This system provides prevention about the chronic diseases. This system is easy to implement in a large scale. This system provides accurate information to the people. This system can make people waste less time in the hospital.

VIII.EXPERIMENTAL RESULT (screenshots)



VIII. CONCLUSION

This paper studies the implementation of Apriori Algorithm on cause for chronic diseases. The outcome of the study is that this algorithm can be efficiently used to discover hidden patterns and generate associated rules from datasets. Percentage of possibility for chronic disease is calculated from the each symptoms of all considered chronic disease. With the more number of symptoms, the accuracy of calculating the disease possibility will be higher. The review of the implementation of Apriori Algorithm gives the results that can be used by physicians and patients for effective decision making.

REFERENCES

- [1] PrasannaDesikan, Kuo-Wei Hsu and JaideepSrivastava, Data Mining for healthcare Management, SIAM International Conference on Data Mining, Hilton Phoenix East|Mesa, Arizona, USA.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

- [2] MsShweta and Dr. KanwalGarg, "Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithm", In: Proceeding of IJARCSSE, ISSN 2277-128X, Vol. 3, Issue 6, June 2013.
- [3] P.KasemthaweesabandW.Kurutach, "Association Analysis of Diabetes Mellitus (DM) With Complication States Based on Association Rules", 7th IEEE Conference on Industrial Electronics and Applications (ICIEA) 2012.
- [4] M. Ilayaraja and T. Meyyappan, "Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm", In: Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), 21-22 February.
- [5] B. M. Patil, R. C. Joshi, DurgaToshniwal, Association rule for classification of type -2 diabetic patients, Second International Conference on Machine Learning and Computing, 9-11 Feb. 2010.
- [6] SarojiniBalakrishnan,RamarajNarayanaswamy, Nickolas Savarimuthu and Rita Samikannu.(2008). SVM Ranking with Backward Search for Feature Selection in Type II Diabetes Databases.IEEE.0 (0), p2628-2633.
- [7] Dilly,Ruth.DataMining.2002http://www.pcc.qub.ac.uk/tec/courses/datam ining/stu_notes/dm_book_1.html.
- [8] ShusakuTsumotoand Shoji Hirano.(2008). Mining Trajectories of Laboratory Data using Multiscale Matching and Clustering.IEEE.0 (0), p626-631.
- [9] Hideno Abe, Hideto Yokoi, Miho Ohsaki and Takahira Yamaguchi. (2008). Developing an Integrated Time-Series Data Mining Environment for Medical Data Mining. IEEE. 0 (0), 127-131.
- [10] Jenn-Lung Su, Guo-Zhen Wu, I-Pin Chao. (2001). THE APPROACH OF DATA MINING METHODS FOR MEDICAL sDATABASE.IEEE.0 (0), p3824-3826.