

# Mining Medical Data for Identifying Frequently Occurring Diseases by using Apriori Algorithm

Rupali Hande  
Assistant Proferssor  
Department of Computer  
Engineering  
VESIT, University of Mumbai, India

Vishal Bulchandani  
Student  
Department of Computer  
Engineering  
VESIT, University of  
Mumbai, India

Hitesh Batreja  
Student  
Department of Computer  
Engineering  
VESIT, University of  
Mumbai, India

Karan Jaisinghani  
Student  
Department of Computer  
Engineering  
VESIT, University of  
Mumbai, India

Sagar Nagwan  
Student  
Department of Computer  
Engineering  
VESIT, University of  
Mumbai, India

## ABSTRACT

Data mining is a process of analyzing data from various perspectives and trims it into useful information. The data can be transformed into knowledge for future use and history patterns. Data mining has a vital role in the domain of information technology. There is a lot of complex data being generated by the health care industry. It includes the details of various patients, hospitals, diagnosis techniques, diseases, etc. The data mining methods prove to be useful for making decisions related to the curing of diseases. The information is hidden because the huge data gathered by the health care industry is not mined. Thus effective decisions cannot be made. The information gained after data mining can be used by doctors and health care administrators for improving the quality of service. In this paper, identification of frequent diseases in a specific location is done using Apriori algorithm. Association rules are applied to extract patterns that occur frequently within a data set. For extracting the results, WEKA tool is used.

## Keywords

Apriori algorithm, data mining, association rule

## 1. INTRODUCTION

The importance of Data Mining is the extraction of information from vast information. It is likewise named as knowledge mining from huge amount of data. There are such a large number of different terms which give comparative importance of data mining; they are pattern/data analysis, knowledge extraction, etc. The other broadly utilized terms are knowledge discovery from data. Decision making can be obtained by transforming data mining into knowledge and this mechanism is called knowledge discovery.

The iterative grouping present in knowledge discovery is 1. Data cleaning [all the inconsistent information and clamor are removed], 2. Data mix [the mix of various information sources are done], 3. Data determination [the applicable information is extricated to the examination undertaking from data base], 4. Data change [the data is changed into significant different structures], 5. Data mining [the data patterns are discovered by smart methods], 6. Pattern assessment [depending upon a few measures, we will distinguish the intriguing examples in knowledge]. Data mining is a step in KDD that comprises of applying data analysis and various

discovery methods that produce a specific list of patterns over the data. Data mining incorporates numerous different functions like association, classification, grouping, and prediction. The important properties of data mining are-discovery of patters, prediction of likely results, formation of noteworthy data and concentrate on substantial data sets and databases. Mining association rules is considered one of the vital data mining applications. In medical field it is utilized to find the frequently occurring diseases in diverse land areas at given time period. Thus the medicinal information is analyzed in this experimental work.

## 2. LITERATURE REVIEW

Jyothi soni. [1] Provided a review of most recent methods in foreseeing heart sicknesses utilizing information mining strategies of information revelation. Such a large number of investigations are led to contrast the exhibitions and with decide the results. The study uncovers that in precision savvy Bayesian grouping is having comparable results as of decision tree. At the point when these are contrasted with different techniques, as Neural Networks, Classification in light of grouping they are performing well. Decision tree calculation and Bayesian arrangement are enhanced by applying Genetic calculation ideal information sets are gotten by diminishing the genuine information size which is helpful in foreseeing Heart infections.

Carlos Ordenez [2] concentrated how to confine the affiliation rules with a specific end goal to foresee the heart infections. He proposed three things to diminish the number of examples.

Firstly, the required things required such that characteristics ought to show on one and only side of the rule. Secondly, separate the attributes into groups. Thirdly, to decrease the quantity of guidelines connected.

Maria-Luiza Antonie [3] examined distinctive data mining systems like association rule mining and neural systems in identification of tumor in computerized mammography.

## 3. APRIORI ALGORITHM

Apriori algorithm is one of the most classical and important algorithms for mining the frequent data itemsets. Apriori operates on databases containing transactions. For example, collection of items purchased by a customer. The algorithm finds subsets which are common to at least a minimum number C (cutoff) of the data itemsets. Apriori algorithm uses

a bottom up approach, where the frequent subsets are extended one at a time. This step is called candidate generation. Apriori algorithm uses the breadth-first search technique and a hash tree structure for counting the candidate item sets. Based on the principle of Apriori, any subset of a frequent itemset is also a frequent itemset. For example, if {AB} is a frequent itemset, then both {A} and {B} are also frequent itemsets. This is how apriori works-

1. Find all the frequent itemsets:
  - a. Get frequent items: Items whose instances in database is greater or equal to minimum support.
  - b. Get frequent itemsets: Generate candidates from frequent items and reduce the results to find the frequent itemsets.
2. Generate association rules from frequent itemsets: Rules which satisfy the minimum support and minimum confidence threshold.

## 4. PROPOSED WORK

### 4.1 Proposed System

Data mining tools are developed for efficient analysis of medical data for helping the clinician in improving the diagnosis. In this exploration work, the specialist can gather information from Hospital Information System (HIS) which has the adequate subtle elements of patient including his name, age, ailment, area, region, and date from labs which continues developing a seemingly endless amount of time. Having gathered the information from hospital facility data framework, this exploration can discover the frequently occurring diseases with the assistance of association techniques. This examination work mines the data about the frequently occurring diseases with the assistance of weka tool connected over preparing data set.

In the medicinal field, the hospital facility data framework is utilized to get various types of medicinal records of infections and its patients who hail from diverse territories. It is a huge task to distinguish the regular illnesses and its reasons from the substantial information set.

Patients from distinctive areas approach diverse clinics. They don't meet in a same spot. Their records are kept up by the healing centers where they get treated. Gathering data about the frequently happening illnesses is not a simple task. The information accumulation in regards to these sorts of diseases should be possible through the association rules. Insights with respect to the event of these diseases in a specific time period can likewise be mined with the help of apriori algorithm.

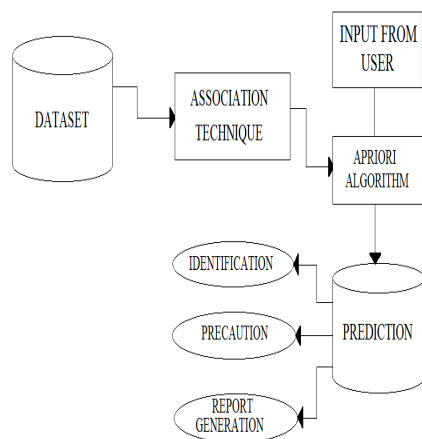


Figure 1: Proposed System

## 4.2 Algorithm

$C_n$ : Medical data itemset of size  $n$

$L_n$ : frequent itemset of size  $n$

$L = \{\text{frequent items}\}$ ;

**for** ( $n = 1; L_n \neq \emptyset; n++$ ) **do begin**

$C_{n+1}$  = Medical data derived from  $L_n$ ;

**each** transaction  $t$  in database **do**

Increment the count of all the medical data in  $C_{n+1}$  that are present in  $t$

$L_{n+1} = \text{min\_support medical data in } L_{n+1}$

**end**

**return**  $\cup_k L_n$ ;

## 4.3 Result of Research

The proposed strategy is valuable to distinguish the regular diseases in an expansive medicinal dataset. The result of this exploration will help the experts in making medicinal choices for often happening illnesses. The analysis is made on information from different topographical areas amid diverse time periods.

## 5. CONCLUSION

This exploration work proposes Apriori data mining algorithm based on association rule and produces the frequency of diseases and the number of patients influenced by these diseases. Based on different geographical areas and at different time periods the study is made. Existing medical details gained from various hospitals are used as training data set for data analysis. The outcome of the study is that this algorithm can be efficiently used to discover hidden patterns and generate associated rules from datasets. 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.

## 6. REFERENCES

- [1] Jyothi Soni, et al., "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction"
- [2] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules"
- [3] Maria-Luiza Antonie et al., "Application of Data Mining Techniques for Medical Image Classification"
- [4] M. Ilayaraja, T. Meyyappan, "Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm"
- [5] Arun K Pujari "Data Mining Techniques", Edition 2001.
- [6] Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200, 2006.
- [7] Divya Jain and Sumanlata Gautam, 2014, "APRIORI ALGORITHM FOR MINING FREQUENT ITEMSETS," International Journal of Computer Science and Communication Engineering, Volume 2 issue 4(November 2013 issue).
- [8] Priyanka, Er. Vinod Kumar Sharma, 2013, "Implementation of Apriori Algorithm in Health Care Sector," International Journal of Computer Science and Communication Engineering, Volume 3-Issue 3, July 2014. Pp. 232-236.

- [9] Gitanjali J, C.Ranichandra ,M.Pounambal, 2014, “APRIORI algorithm based medical data mining for frequent disease identification” IPASJ International Journal of Information Technology (IJIT), Volume 2, Issue 4, April 2014.
- [10] Paresh Tanna, Dr. Yogesh Ghodasara, 2014, “Using Apriori with WEKA for frequent pattern mining” International Journal of Engineering Trends and Technology (IJETT), Volume 12, Number 3, June 2014.