

# Data Mining Technique: An Implementation of Association Rule Mining in Healthcare

Ashwini Rajendra Kulkarni<sup>1</sup>, Dr. Shivaji D. Mundhe<sup>2</sup>

Assistant Prof., Sinhgad Institute of Management and Computer Application (SIMCA), Narhe, Pune, India <sup>1</sup>

Director-MCA, Sinhgad Institute of Management and Computer Application (SIMCA), Narhe, Pune, India <sup>2</sup>

**Abstract:** Data mining is used to find out the meaningful information from the large dataset. Data mining plays vital role in various domains including healthcare, banking, retail, education, telecommunication, information security, finance etc. Association rule mining is one of the data mining techniques which plays vital role for analysis and decision making and it is used to find out the various patterns or association in the item set. The present study focuses on the theoretical aspects of association rule mining as well as implementation aspects also. The paper describes about the Association rule mining and an Apriori Algorithm. Also the paper discuss about the reviews of research work done in this filed by diverse researchers, scholars, organizations etc. This paper is intended towards an association rule generation using in healthcare especially for the viral infective diseases.

**Keywords:** Data Mining, Text Mining, Association Rule, Apriori Algorithm, Diseases.

## I. INTRODUCTION

An association rule mining is nothing but finding the frequent patterns, correlations, associations, or casual structures among the sets of items or objects in transaction database, relational database, and other information repositories. In data mining, Association Rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness.

Based on the concept of strong rules, we can introduce association rules for discovering the symptoms of particular diseases for healthcare as {Fever, Cough, Chill, Bodyache} → {Influenza-Flu} this rule indicates that if the symptoms are Fever, Cough, Chill, Bodyache then a person is having Influenza-Flu disease; this kind of information is useful for healthcare, medical experts for the better decision making and proper treatment. The association rule plays vital role in many applications such as healthcare, baking, education, telecommunication, retail industry, intrusion detection, web mining, bioinformatics etc.

## II. ASSOCIATION RULE MINING

Association rule is having two main important properties namely Support and Confidence as

Support (AB) = P (AUB)

Confidence (AB) = P (B|A)

If we correlate support and confidence then

Confidence (AB) = P (B|A) = Support\_count (AUB)/Support\_count (A)

Support\_count (AUB): number of transaction containing the item sets AU B,

Support\_count (A): number of transactions containing the item set A. [2]

Association rule two steps , one is **Find all frequent item sets** : By definition, each of these item sets will occur at least as frequently as a predetermined minimum support count; and another is Generate strong association rules from the **frequent item sets**: By definition, these rules must satisfy minimum support and minimum support and minimum confidence.

### Apriori Algorithm: (Finding Frequent Item sets Using Candidate Generation)

Apriori is an influential algorithm for mining frequent item sets. The name of the algorithms is based on the fact that the algorithm uses prior knowledge of frequent item sets properties. Apriori employs an iterative approach known as a level-wise search.

```

Input:
    database D
    Mini Support  $\epsilon$ 
    Mini Confidence  $\xi$ 
Output:
     $R_t$  All association rules
Method:
01      $L_1 =$  large 1-itemsets;
02     for( $k=2$ ;  $L_{k-1} \neq \emptyset$ ;  $k++$ ) do begin
03          $C_k =$ apriori-gen( $L_{k-1}$ ); //generate new candidates from  $L_{k-1}$ 
04         for all transactions  $T \in D$  do begin
05              $C_t =$ subset( $C_k, T$ ); //candidates contained in T.
06             for all candidates  $C \in C_t$  do
07                 Count( $C$ )=Count( $C$ )+1; // increase support count of C by 1
08             end
09              $L_k = \{C \in C_t \mid \text{Count}(C) \geq \epsilon \times |D|\}$ 
10         end
11          $L_f = \bigcup_k L_k$ ;
12          $R_t =$ GenerateRules( $L_f, \xi$ )

```

Diagram 1: Apriori Algorithm [4]

### III.LITERATURE REVIEW

The researcher Doddi S, et.al had worked on association rule generation of medical data containing the patient's records and reported diagnosis. An Apriori algorithm is used for association rule generation of medical data. [1]

The research paper entitled "Association Rules Mining: A Recent Overview", briefed about the various applications of association rule mining. The researcher also focused on theoretical aspect of Apriori algorithm; also the merits and demerits along with the negative association rules are briefed. [3, 4]

The author Jagmeet Kaur and Neena Madan entitled a research paper "Association Rule Mining: A Survey", explained about the association rule theoretical views along with an Apriori algorithm, FP growth algorithm and genetic algorithm. The authors also discussed about the application of association rule such as Market Basket Analysis, CRM of the Credit Card Business, Medical Diagnosis, Census Data, Protein Sequences, etc. [5]

Rashid M., et.al entitled "Association Rules Mining Based Clinical Observations", aimed towards the development of an application for the patient's diseases prediction and also worked on occurrences of one disease relates to the other diseases through the system using association rule mining through Apriori algorithm. [6]

A research paper entitled "ASSOCIATION RULE MINING ON MEDICAL DATA TO PREDICT HEART DISEASE" by Said I., et.al illustrated association rule generation for medical data for prediction of heart diseases using Apriori algorithm. The researcher had used medical data through which the sick and healthy patient's details are categorized for the prediction of heart diseases. [7]

The authors Amin A., Talib R. and Raza S. (2014) in their research paper entitled "Extract association rules to Minimize the effects of dengue by using a text mining technique "specified on the use of XML for information retrieval and suggests architectural model of text mining system for text processing, mining association rule and visualization. The researcher had designed GRAW Algorithm (generating association rules from based on weighting scheme) which is focused to only XML files to extract the association rules for minimizing the effects of Dengue. [10]

Report of Oracle® Data Mining Concepts 11g Release 1 (11.1), 2008 focus on unstructured data and various text mining algorithms for Such as Naive Bayes, Generalized Linear Models, Support Vector Machine, k-Means, Non-Negative Matrix Factorization, Apriori, Minimum Descriptor Length etc. Report also concludes that Apriori algorithm is used for association mining function which generates various association rules for information retrieval. [11]

The research paper entitled APRIORI algorithm based medical data mining for frequent disease identification by the researcher Gitanjali J, et al. discussed on the large amount of healthcare data regarding the various aspect like diseases, patients, diagnosis, hospital resources, etc. Applying data mining in healthcare the discovered knowledge will be useful for further decision making. The paper aimed towards the identification of frequency of diseases for particular geographical location using Apriori an association rule mining technique. [12]

An author Kaur G. entitled Association Rule mining: a survey explained about the association rule mining techniques namely AIS, Apriori, and FP-Growth with its advantages and pitfalls. Author also stated about the applications of association rule mining in diverse areas like medical diagnosis, inventory control, telecommunication networks, risk and market management and many more. [16]

## IV. ASSOCIATION RULE GENERATION FOR VIRAL INFECTIVE DISEASES

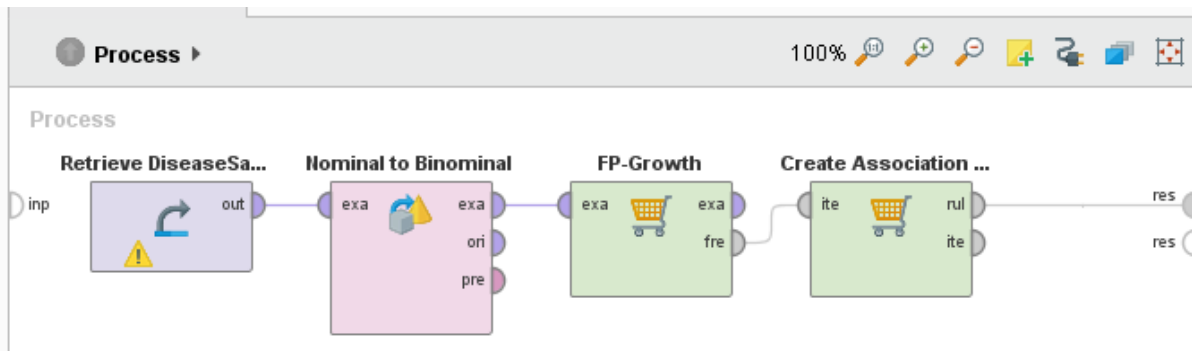


Diagram 2: Dataset for Association Mining in Rapid Miner

This research used an Apriori algorithm and FP-Growth an association rule generation algorithm for various viral infective diseases and their symptoms. The analysis of the diseases dataset is done using Rapid miner text mining tool. The Diagram.1 represents an association rule generation for viral infective diseases. The data source is retrieved through excel which contains diseases and their symptoms then he data is converted from nominal to binomial. The FP-Growth technique is used to set the min support and confidence and then association rules are generated. The association rules generated through the Rapid miner are represented in Diagram 3.

For example first rule indicates that

{swelling, fever, headache, feeling-sick, dry-mouth, joint-ache} → {Mumps}

{Mumps} → {swelling, fever, headache, feeling-sick, dry-mouth, joint-ache}

It will be helpful for various hidden pattern generation and interpretation for further decision making and making healthcare sector helpful.

**Result in Rapid Miner:**

## Association Rules

[Symptoms = swelling, fever, headache, feeling-sick, dry-mouth, joint-ache] --> [Disease = Mumps] (confidence: 1.000)

[Disease = Mumps] --> [Symptoms = swelling, fever, headache, feeling-sick, dry-mouth, joint-ache] (confidence: 1.000)

[Symptoms = fever, chills, cough, soar-throat, runny-nose, bodyache, headache, fatigue, vomitting] --> [Disease = InfluenzaFlu] (confidence: 1.000)

[Disease = InfluenzaFlu] --> [Symptoms = fever, chills, cough, soar-throat, runny-nose, bodyache, headache, fatigue, vomitting] (confidence: 1.000)

[Symptoms = fever ,jointpain, headache, musclepain, joint-swelling, rash] --> [Disease = Chikungunya] (confidence: 1.000)

[Disease = Chikungunya] --> [Symptoms = fever ,jointpain, headache, musclepain, joint-swelling, rash] (confidence: 1.000)

[Symptoms = cough,coryza,runny-nose, fever, sneezing, rash, boadyache, watery-eyes, Soar-Throat] --> [Disease = Measles] (confidence: 1.000)

[Disease = Measles] --> [Symptoms = cough,coryza,runny-nose, fever, sneezing, rash, boadyache, watery-eyes, Soar-Throat] (confidence: 1.000)

[Symptoms = chills,fever,coughing,sore-throat,runny-stuffy-nose,bodyaches,fatigue, nausea,vomiting] --> [Disease = Swine Flu] (confidence: 1.000)

[Disease = Swine Flu] --> [Symptoms = chills,fever,coughing,sore-throat,runny-stuffy-nose,bodyaches,fatigue, nausea,vomiting] (confidence: 1.000)

[Symptoms = abdominal-pain, nausea, vomiting] --> [Disease = Diarrhoea] (confidence: 1.000)

[Disease = Diarrhoea] --> [Symptoms = abdominal-pain, nausea, vomiting] (confidence: 1.000)

[Symptoms = Rash, Red-spot, Soar-throat, high-fever, vomitting, headache, tiredness, Painfull-Blisters] --> [Disease = Chickenpox] (confidence: 1.000)

[Disease = Chickenpox] --> [Symptoms = Rash, Red-spot, Soar-throat, high-fever, vomitting, headache, tiredness, Painfull-Blisters] (confidence: 1.000)

[Symptoms = Nausea, vomiting, Dehydration, Muscle-cramps, abdominal-pain] --> [Disease = Cholera] (confidence: 1.000)

[Disease = Cholera] --> [Symptoms = Nausea, vomiting, Dehydration, Muscle-cramps, abdominal-pain] (confidence: 1.000)

[Symptoms = High-Fever, Abdomen-Pain, vomiting, cold, headache, rash] --> [Disease = Dengue] (confidence: 1.000)

[Disease = Dengue] --> [Symptoms = High-Fever, Abdomen-Pain, vomiting, cold, headache, rash] (confidence: 1.000)

[Symptoms = Fever, Headache, Nausea, Vomiting, Anxiety, Hyperactivity, Confusion, Agitation] --> [Disease = Rabies] (confidence: 1.000)

[Disease = Rabies] --> [Symptoms = Fever, Headache, Nausea, Vomiting, Anxiety, Hyperactivity, Confusion, Agitation] (confidence: 1.000)

### Diagram 3: Association Rules Generated in Rapid Miner

## V. CONCLUSION

The data mining works with the structured data that is the data which is in row and column format. The above results are drawn for the diseases information with the disease name and symptoms of diseases which are in structured form. According to the report of Oracle Data Mining Concepts Now a day's, approximate 90% of data is in the unstructured format that is text, images, audio, video, graph, email, blog, etc and only 10 % of data is in the structured form. There is need to focus on unstructured data and getting the better results from it. The future scope of the present study is to work on text mining of healthcare data and generating the various patterns from it. This research used an Apriori algorithm an association rule generation algorithm for various viral infective diseases and their symptoms

## REFERENCES

- [1]. Doddi S. , Marathe A. S. S. Ravi, David C. Torney, Discovery of Association Rules in Medical Data
- [2]. M. Inbava Ili(2015).Efficient Analysis of Frequent item set Association Rule Mining Methods. International Journal of Scientific & Engineering Research, 6(4), 1-23, <http://www.ijser.org/researchpaper%5CEfficient-Analysis-of-Frequent-itemset-Association-Rule-Mining-Methods.pdf>
- [3]. Kotsiantis S., Kanellopoulos D.(2006).Association Rules Mining: A Recent Overview, GESTS International Transactions on Computer Science and Engineering, 32 (1), 71-82
- [4]. Zhao Q. ,Bhowmick S. (2003).Association Rule Mining: A Survey. Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116 ,1-20
- [5]. Kaur J. , Madan N. (2015). Association Rule Mining: A Survey. International Journal of Hybrid Information Technology ,8(7),pp.239-242 <http://dx.doi.org/10.14257/ijhit.2015.8.7.22>
- [6]. Rashid M. , Hoque T., Sattar A. Association Rules Mining Based Clinical Observations. Griffith University Nathan, QLD, Australia
- [7]. Said I. , Haruna A. , Garko A. (2015). ASSOCIATION RULE MINING ON MEDICAL DATA TO PREDICT HEART DISEASE International Journal of Science Technology and Management, 4(8), 26-35
- [8]. M. Devi, A. Sarojini (2012). APPLICATIONS OF ASSOCIATION RULE MINING IN DIFFERENT DATABASES, Journal of Global Research in Computer Science, 3(8), 30-34, <http://www.jgrcs.info/index.php/jgrcs>
- [9]. Ordonez C., Santana C., Braal L. Discovering Interesting Association Rule in Medical Data
- [10]. Amin A., Talib R., Raza S. (2014). Extract association rules to minimize the effects of dengue by using a text mining technique. International Journal of Computer Science and Mobile Computing, 3(4), 394-400 , [www.ijcsmc.com](http://www.ijcsmc.com)
- [11]. Report of Oracle® Data Mining Concepts 11g Release 1 (11.1) B28129-04. (2008) Retrieved from : [http://docs.oracle.com/cd/B28359\\_01/datamine.111/b28129/text.htm#BCEDHEDD](http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/text.htm#BCEDHEDD)
- [12]. Gitanjali J, C. Ranichandra, M. Pounambal. (2014). APRIORI algorithm based medical data mining for frequent disease identification, IPASJ International Journal of Information Technology (IJIT), 2(4),1-5, <http://www.ipasj.org/IJIT/IJIT.htm>
- [13]. Kaur G. (2014). Association Rule Mining: A Survey. (IJCSIT) International Journal of Computer Science and Information Technologies, 5(2) , 2320-2324 , [www.ijcsit.com](http://www.ijcsit.com)
- [14]. Rajput V., Manjhvar A. (2017). A Review Paper on Dengue Disease Forecasting Using Data Mining Techniques. IJSART, 3(4), 240-249, [www.ijcsart.com](http://www.ijcsart.com)
- [15]. Kumar K., S. Arumugaperumal (2013), Association Rule Mining and Medical Application: A Detailed Survey, International Journal of Computer Applications (0975 – 8887, 80(17), 10-19
- [16]. Kaur G. (2014). Association Rule mining: a survey. (IJCSIT) International Journal of Computer Science and Information Technologies, 5 (2) , 2320-2324