# An Application of Apriori Algorithms for Finding Frequent Itemsets in Pharmacy Database to Find Frequent Medicine Itemsets

**V. Pushpalatha[1] Mr. J.S. Ananda Kumar[2]**
[1]Student [2]Assistant Professor
[1,2]Department of Computer Applications
[1,2]KMM Institute of PG Studies, Tirupati, India

*Abstract*— Apriori algorithm has been successfully used for finding the persistent item sets in huge databases. In this paper study about finding associations between item sets contains data about diagnosis and treatment. It has shown that the algorithm is equally beneficent for finding the large item sets and thus generating the association rules in transactional databases. Healthcare is a data rich domain so we have applied Apriori algorithm for medical practices, insurance companies and other health related organizations that have collected huge volumes of data, thus attracting data mining researchers to explore it and find something beneficent from it.

***Key words:*** Association Rules, Persistent Pattern, Data Mining

## I. INTRODUCTION

Association Rule Mining (ARM) has become a mature field of research. So many research papers, articles are surveyed in the field of ARM. This paper details some fundamental about persistent item sets generations which helps to develop new algorithm for that process. The field of ARM is divided into the following areas: positive rule mining, Negative rule mining and interesting measures. Major area of work in ARM is coming under these three categories. The classical rules are called positive rules which are showed in the paper. The positive rules are mined from a set of persistent item sets. Due to the deficiency of persistent item set mining, the persistent item sets are extended to various formats like closed, maximum, sequential, complex persistent item set. The persistent item set mining is detailed in the paper. The above types of persistent item set are supported to constraints based rule mining. The constraint based rule mining is described in the section. The negative mining process using in persistent item set. The rules are mined from these kinds of in persistent item set relationship between item set are mined by rule that is called negative rule mining which is explained in section. The interestingness measures play an essential role in the field of ARM similar to the data mining process. Apriori we have to give transactional database and minimum support value as input and persistent item set as output.

## II. RELATIVE STUDY

Paula R.C. Silva et.al, 2015 expresses a novel approach to discover professional profile patterns from LinkedIn by using association rule mining to extract relevant pattern from the data warehouse, evaluate their approach academic activities and curricula in educational instructions [6]. P. Nancy et.al, 2013 discussed the Facebook 100 universities data set in United States from which association rules are mined. Knowledge pattern regarding the association between the major (course) and gender were identify [7]. Trand et.al, 2010 examined the community structures of facebook networks whose links represent “friendship” between user pages within each of five American universities[8]. Ahmet Seiman Bozkir et.al 2009 investigated demographic characteristics of facebook users and their frequency, time spent on facebook and membership in facebook group using association rules[9]. Xiao Cui et.al,2014 explored the relationship among different profile attributes in Sina Weibo using association rule mining to identified the dependency among the attributes [10].Balaji Mahesh, VRK Rao G Subrahmanya [11].

## III. PROPOSED ALGORITHM

First, we construct the matrix the above example in table (1). Then the matrix may be following.

$$\begin{array}{c c} & \begin{array}{ccccc} I1 & I2 & I3 & I4 & I5 \end{array} \\ \begin{array}{c} I1 \\ I2 \\ I3 \\ I4 \\ I5 \end{array} & \begin{pmatrix} 6 & 4 & 4 & 1 & 2 \\ 4 & 7 & 4 & 2 & 2 \\ 4 & 4 & 6 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 2 & 1 & 0 & 2 \end{pmatrix} \end{array}$$

Fig. 1: frequency matrix

Step 1

In support count=2, extract all possible frequent itemsets by matching rows and columns satisfying min-sup=2.

Then{I1,I2},{I1,I3},{I1,I5},{I2,I3},{I2,I4},{I2,I5}

Step 2

Remove extracted frequent itemset which included infrequent subsets

Remove extracted frequent item set which is not satisfied m*in_sup*

Then we get {I1, I2, I3},{I1,I2,I5},{I2,I4}

The result of frequent itemset may be{I1,I2,I3},{I1,I2,I5},{I2,I4}

The algorithm uses only 2 step for frequent itemset. Thus our proposed algorithm is less computational complexity than the Apriori algorithm.

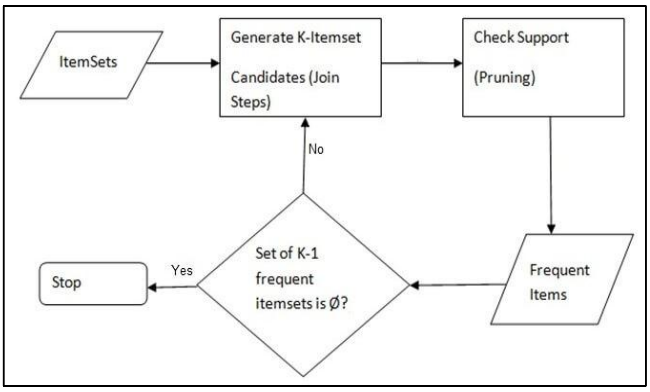
## IV. APPLICATION ARCHITECTURE

Step1: Apply minimum support to find all the frequent sets with k items in a database.

Step2: Use the self-join rule to find the frequent sets with k+1 items with the help of frequent k- itemsets. Repeat this process from k=1 to the point when we are unable to apply the self-join rule.

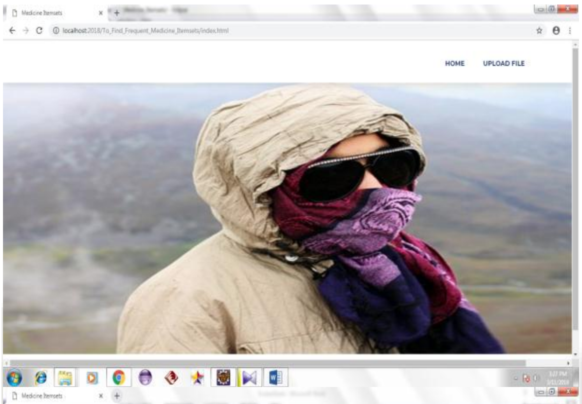
This approach of extending a frequent itemsets one at a time is called the “bottom up” approach.

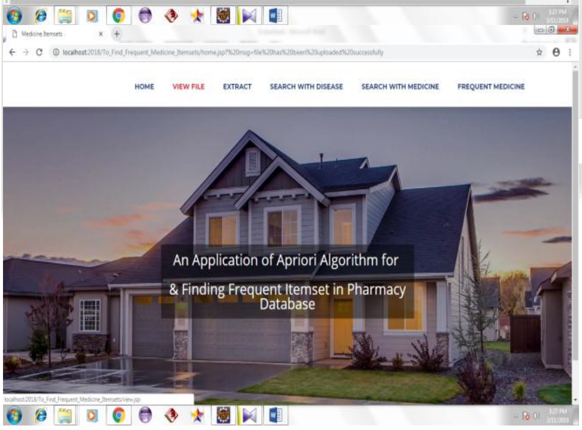
 




## V. APRIORI ALGORITHM ANALYSIS AND RESULTS

### A. *Home*



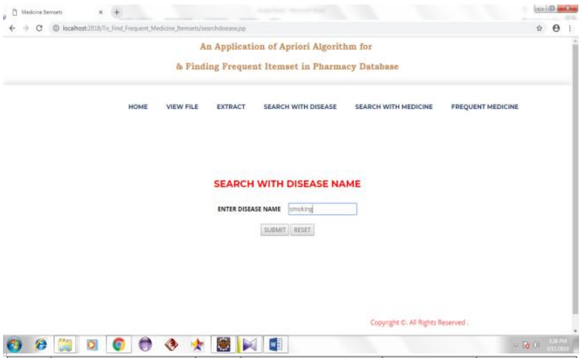

### B. *View files*

| S. No | Name | Age | Disease | Months | Medicine |
|---|---|---|---|---|---|
| 1 | Cheekati sai manoj | 90 | Lupus | August | Azithromycin |
| 2 | Kethavath Srinu Naik | 46 | thrush | August | Clindamycin |
| 3 | Immadi Divya Sai | 32 | Lupus | February | Citalopram |
| 4 | Vajjha Sri Valli | 32 | HIV | March | Hydrochlorothiazide |
| 5 | Akella V S Vinay karthik | 94 | HIV | July | Lisinopril |
| 6 | Bukya mamatha | 48 | Allergies | April | Acetaminophen |
| 7 | Pannem Gautami | 48 | Pertussis | March | Ciprofloxacin |
| 8 | Yelleni mounika | 64 | Smoking | June | Lbuprofen |
| 9 | M Pranav pratheek | 59 | Asthma | March | Lexapro |
| 10 | Voruganti Abhinava | 94 | Impotence | February | Doxycycline |
| 11 | P Harish kumar | 27 | Insomnia | March | Lexapro |
| 12 | Maddela deepak | 94 | Lenukemia | July | Loratadine |
| 13 | Gurram pavan | 32 | Pregnancy | June | Lbuprofen |
| 14 | Kancherla Sandesh | 59 | Thrush | October | Acetaminophen |

### C. *Extract*

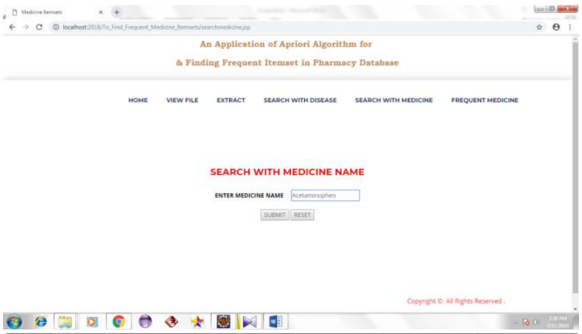
| S. No | Name | Age | Disease | Months | Medicine |
|---|---|---|---|---|---|
| 1 | Cheekati sai manoj | 90 | Lupus | August | Azithromycin |
| 2 | Kethavath Srinu Naik | 46 | thrush | August | Clindamycin |
| 3 | Immadi Divya Sai | 32 | Lupus | February | Citalopram |
| 4 | Vajjha Sri Valli | 32 | HIV | March | Hydrochlorothiazide |
| 5 | Akella V S Vinay karthik | 94 | HIV | July | Lisinopril |
| 6 | Bukya mamatha | 48 | Allergies | April | Acetaminophen |
| 7 | Pannem Gautami | 48 | Pertussis | March | Ciprofloxacin |
| 8 | Yelleni mounika | 64 | Smoking | June | Lbuprofen |
| 9 | M Pranav pratheek | 59 | Asthma | March | Lexapro |
| 10 | Voruganti Abhinava | 94 | Impotence | February | Doxycycline |
| 11 | P Harish kumar | 27 | Insomnia | March | Lexapro |
| 12 | Maddela deepak | 94 | Lenukemia | July | Loratadine |
| 13 | Gurram pavan | 32 | Pregnancy | June | Lbuprofen |
| 14 | Kancherla Sandesh | 59 | Thrush | October | Acetaminophen |
| 15 | M Pranav Pratheek | 78 | Over Weight | September | Azithromycin |
| 16 | Vijay mohan reddy | 50 | Eye disease | May | Clindamycin |
| 17 | Kodali Yamini | 48 | AIDS | November | Lexapro |

### D. *Search with Disease Name*




| S.No | Name | Age | Disease | Months | Medicine |
|---|---|---|---|---|---|
| 1 | Yelleni mounika | 64 | Smoking | June | Lbuprofen |
| 2 | Bangari gnaneshwar | 27 | Smoking | August | Lisinopril |
| 3 | Meruga akhil | 52 | Smoking | August | Azithromycin |
| 4 | Ranga gayathri | 50 | Smoking | September | Gabapentin |
| 5 | Sai sneha mulpuru | 48 | Smoking | February | Azithromycin |
| 6 | M Pranav pratheek | 64 | Smoking | April | Ciprofloxacin |
| 7 | Annamaraja sripada | 27 | Smoking | February | Ciprofloxacin |
| 8 | Kunta anil | 64 | Smoking | May | Lyrica |
| 9 | Vangaveti sai | 27 | Smoking | August | Doxycycline |
| 10 | Nagoor rakesh | 52 | Smoking | June | Acetaminophen |
| 11 | Annamaraja sripada | 78 | Smoking | April | Doxycycline |
| 12 | Mohammed | 78 | Smoking | December | Gabapentin |
| 13 | Mamidi sumikendar | 24 | Smoking | July | Lbuprofen |

### E. *Search with Medicine*




| S.No | Name | Age | Disease | Months | Medicine |
|---|---|---|---|---|---|
| 1 | Bukya Mamatha | 48 | Allergies | April | Acetaminophen |
| 2 | Kancherla sandesh | 59 | Thrush | October | Acetaminophen |
| 3 | K srinivas | 24 | Lung cancer | July | Acetaminophen |
| 4 | Chekuri bhavani | 69 | Insomnia | February | Acetaminophen |
| 5 | Ponnala jeevan | 27 | HIV | September | Acetaminophen |
| 6 | Konda kavya | 24 | Eye disease | July | Acetaminophen |
| 7 | R. Nehru | 52 | Pertussis | November | Acetaminophen |
| 8 | R. Nehru | 64 | Lenukemia | June | Acetaminophen |
| 9 | Yachamaneni preethika | 46 | Thrush | August | Acetaminophen |
| 10 | P Harish kumar | 52 | AIDS | August | Acetaminophen |
| 11 | Kancheria sandesh | 20 | Kidney disease | November | Acetaminophen |
| 12 | Kasarla mani kumar | 20 | Impotence | April | Acetaminophen |
| 13 | Nareddy chandra | 64 | Jaundice | February | Acetaminophen |

### *F. Frequent Medicine*

| Frequent medicine itemsets & disease | Count |
|---|---|
| Azithromycin , Lupus | 2855 |
| Lisinopril, Lupus | 4295 |
| Smoking , Thrush | 4585 |
| Clindamycin, thrush | 4240 |
| Citalopram, Lupus | 3350 |
| Hydrochlorothiazide, HIV | 4480 |
| Lisinopril, HIV | 1505 |
| Acetaminophen, Allergies | 3510 |
| Ciprofloxacin , Pertussis | 1725 |
| Lbuprofen , smoking | 4475 |
| Lexapro, Insomnia | 1685 |
| Loratadine , Lenukemia | 3435 |
| Ibuprofen, pregnancy | 1470 |
| Acetaminophen, Thrush | 4225 |

## VI. ALGORITHM ANALYSIS

The pseudo code for the algorithm is given below for a transaction database T, and a support threshold of E. Usual set theoretic notation is employed, though note that T is a multiset. $C_k$ is the candidate set for level K. At each step ,the algorithm is assumed to generate the candidate sets from the large item sets of the preceding level, heeding the downward closure lemma. Count[c] accesses a field of the data structure that represents candidate set C, which is initially assumed to be zero. Many details are omitted below, usually the most important part of the implementation is the data structure used for storing the candidate sets, and counting their frequencies.

$\mathrm{Apriori}(T, \epsilon)$
- $L_1 \leftarrow \{\text{large } 1-\text{itemsets}\}$
- $k \leftarrow 2$
- **while** $L_{k-1} \neq \emptyset$
  - $C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k-1\} \nsubseteq L_{k-1}\}$
  - **for** transactions $t \in T$
    - $D_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$
    - **for** candidates $c \in D_t$
      - $count[c] \leftarrow count[c] + 1$
  - $L_k \leftarrow \{c \mid c \in C_k \wedge \ count[c] \geq \epsilon\}$
  - $k \leftarrow k+1$
- **return** $\bigcup_k L_k$

## VII. CONCLUSION

The use of data mining is very wide it helps to growing in business applications, so, the research scope exists for new algorithms applied to the large amount of data stored in enterprise**'s** databases. The main objective of this paper is to observe the various Apriori algorithms and find the limitations of existing Apriori algorithms. From the experimental results, it is observed that the evaluation of support and confidence for traditional Apriori algorithm consumes more time, space and also spawns the number of candidate sets.

In future, there is a scope for development of improved or modified version of existing Apriori algorithm based on weighted value Apriori, hash tree Apriori, matrix, interest item sets, transaction compression. One can develop the hybrid algorithm using different existing Apriori algorithm for identifying frequent with focus on reducing the computational time and memory space using a retail data set.

## REFERENCES

[1] Changxin Song, "Research of Association Rule Algorithm Based On Data Mining," IEEE International Conference of Big Data Analytics(ICBDA), Pp.1- 4, 12-14 March 2016.

[2] O. Jamsheela and Raju.G, "Persistent Item set Mining Algorithms: A Literature Survey," IEEE International Advance ComputingConference(IACC), Pp. 1099-1104, 12-13 June 2015

[3] L. Fang and Q. Qizhi, "The Study On the Application of Data Mining Based On Association Rules," IEEE International Conference OnCommunication Systems and Network Technologies(CSNT), Pp. 477-480, 11-13 May 2012.

[4] Archana Singh and Jyoti Agarwal, "Proposed Algorithm for Persistent Item Set Generation," IEEE International Conference On Contemporary Computing(IC3), Pp.160-165, 7-9 August 2014.

[5] S. D. Patil and Dr R. R. Deshmukh, "Review and Analysis ofApriori Algorithm for AssociationRule Mining," IEEE International Journal of Latest Trends in Engineering and Technologies (IJLTET), Volume 6, Issue 4, March 2016.

[6] Gayana Femanda & Md Gapar MdJohar, Framework for social network data mining, International journal of computer application, April,2015.

[7] Nergis Y1, limazans Gulfem1 and KlarAlptekin, The Effect of Clustering in the Apriori Data Mining algorithm: A case study, Proceedings of the World Congress on Engineering , July,2013.

[8] S,Sphulari, P.U.Bhulchadra, Dr.S.D.Khamitkar&S.N.Lokhande ,Understanding rule behavior through apriori algorithm over social network data , Globle journal of computer science and technology software &data engineering,2012.

[9] Xiao Cui, Hao Shi, Xun Yi, Application of association rule mining theory in SinaWeibo, Journal of computer and communication,2014.

[10] P. Nancy, R. GeethaRamani&ShomonaGracia Jacob, Mining of Association Patterns in Social Network Data (Face Book 100 Universities) through Data Mining Techniques and Methods,Springer,2013.