



## Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms

Ms Shweta<sup>1</sup>

<sup>1</sup>M Tech. Scholar,

Dept. of Computer Science and Applications,  
Kurukshetra University,  
Kurukshetra, India

Dr. Kanwal Garg<sup>2</sup>

<sup>2</sup>Assistant Professor,

Dept. of Computer Science and Applications,  
Kurukshetra University,  
Kurukshetra, India

**Abstract**— In frequent pattern mining, there are several algorithms. Apriori is the classical and most famous algorithm. Objective of using Apriori algorithm is to find frequent itemsets and association between different itemsets i.e. association rule. In this paper, author considers data (bank data) and tries to obtain the result using Weka a data mining tool. Association rule algorithms are used to find out the best combination of different attributes in any data. In this paper author uses Apriori to find association rule. Here author consider three association rule algorithms: Apriori Association Rule, PredictiveApriori Association Rule and Tertius Association Rule. Author compares the result of these three algorithms and presents the result. According to the result obtained using data mining tool author find that Apriori Association algorithm performs better than the PredictiveApriori Association Rule and Tertius Association Rule algorithms.

**Keywords:** Weka, Apriori, Association rules, Frequent pattern mining, CLI (Command Line Interface).

### 1. INTRODUCTION

KDD(knowledge discovery in database) and data mining are used as synonyms to each other. But in real, Data mining is the core process of KDD. Data mining is used to extract the information from any system by analyzing the present in the form of data [1]. In this paper author focuses on the problem of frequent pattern mining. Frequent patterns are the patterns that that occur in database at least user given number of times. Problem of frequent pattern mining can be defined as: given a large database of transactions, each consists of set of items. Aim of this problem is to find all the frequent itemsets i.e. a set of items Y is frequent if greater than min\_supp % of all transaction in database contains Y and finding association rules from these frequent itemsets [2]. Association rules was first introduced by Agarwal [1]. Association rules are helpful for analyzing customer behavior in retail trade, banking system etc. Association rule can be defined as  $\{X, Y\} \Rightarrow \{Z\}$ . It means in retail stores if customer buys X, Y he is likely to by Z. this concept of association rule today used in many application areas like intrusion detection, biometrics, production planning etc. In this paper, author uses Apriori algorithm to find out the frequent pattern and then from them association rules. Rest of the paper is organized as follow: section 2 gives the review of literature. And in section3 author will demonstrate the basic concepts. In section 4 author give the working of Apriori algorithm using example and a data mining tool and does the comparative analysis of three association rule algorithms. Finally, in section 5 author concludes

### 2. Literature Review

In this section author has discussed some research papers which had been previously undertaken in the field of association rule mining, frequent pattern mining. Agrawal et al. introduced the notion of association rules. Later, that becomes very useful in the field of retail trade. Nahar J. et al. used the concept of frequent pattern in cancer prevention, here objective of the author was to find the relevant prevention factor for different type of cancer; for this author used Apriori, Predictive Apriori, and Tertius algorithms. Jogi. Suresh and T. Ramanjaneyulu addressed the mining of frequent itemsets using apriori algorithm with an example. Sunita B. Aher and Lobo performed the comparative study of association rule algorithms for course recommender system in E-learning.

### 3. Basic Conception

Basic concepts related to frequent pattern mining are given below:

**3.1 Association Rules:** association rule are the statements that find the relationship between data in any database. Association rule has two parts 'Antecedent' and 'Consequent'. For example,  $\{\text{egg}\} \Rightarrow \{\text{milk}\}$ . Here egg is the antecedent and milk is the consequent. Antecedent is the item that found in database, and consequent is the item that found in combination with the first. Association rules are generated during searching for frequent patterns [12].

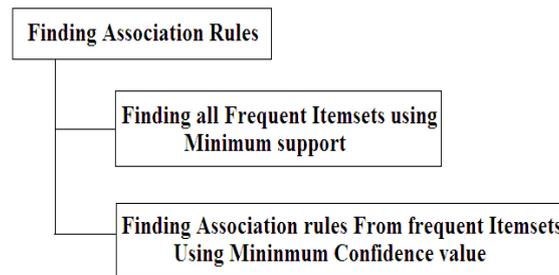


Figure 1 Generating Association rules

The problem of finding association rules is divided into two sub problems: first is to find frequent itemsets and second is to find association rules from these itemsets [12]. For important relationships association rule uses the criteria of ‘Support’ and ‘Confidence’ that are explained below:

→Support (s): it is an indication of item how frequently it occurs in database. For a rule  $A \Rightarrow B$ , its support is the percentage of transaction in database that contain AUB (means both A and B) [6].

→Confidence (c): it indicates the no of times the statements found to be true. Confidence of the rule given above is the percentage of transaction in database containing A that also contain B [6].

→Lift: the lift of rule is defined as:

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Supp}(A \cup B)}{\text{Supp}(B) * \text{Supp}(A)} \quad [6]$$

→ Conviction: conviction is similar to lift, but it measures the effect of the right-hand-side not being true. It also inverts the ratio. So, convictions is measured as:

$$\text{Conviction} = \frac{P(L).P(\text{not } R)}{P(L,R)} \quad [6]$$

#### 4. Apriori Algorithm

Apriori is the Latin word and its meaning is ‘from what comes before’. Apriori uses bottom up strategy. It is the most famous and classical algorithm for mining frequent patterns. This algorithm works on categorical attributes. Apriori uses breadth first search [5].

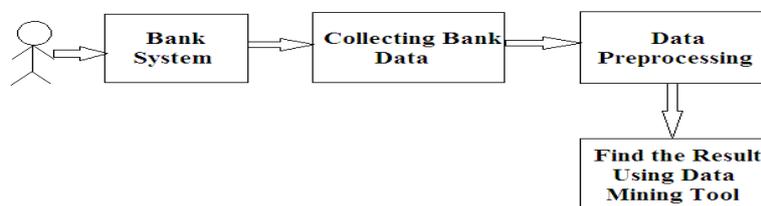


Figure 2 Framework to find the Association Rules

##### 4.1 Important Terms Used in Apriori

→Min\_supp: it is minimum support used for searching frequent patterns that satisfy this constraint.

→Min\_conf: it is Minimum confidence used for finding the strong association rule that satisfy this threshold [9].

→Frequent Itemset ( $L_i$ ): denoted by  $L_i$ , where  $I$  means  $i^{\text{th}}$  item, these are the item sets that satisfy the minimum support (min\_supp) threshold [9].

→Join Operation: for  $L_k$ , a set of candidate  $k$ - itemsets ( $C_k$ ) is generated by joining  $L_{k-1}$  with  $L_{k-1} (L_{k-1} \times L_{k-1})$  [9].

→Apriori Property: this property is very useful for trimming irrelevant data. It states that any subset of frequent itemset must be frequent.

→Prune step: used for finding frequent itemsets, for any  $(k-1)$ -itemsets that is not frequent cannot become subset of a frequent  $k$ -itemset [9].

Definitions:  $L_k$  – set of frequent itemsets of ‘ $k$ ’ size found using min support.

$C_k$  – set of candidate itemsets of ‘ $k$ ’ size

##### 4.2 Pseudo Code for Apriori Algorithm:

→ Scan the transactions to find  $L_1$  (i.e. the set of all frequent 1-itemsets) with their counts;

→For ( $k=2$ ;  $L_{k-1} \neq \emptyset$ ;  $k++$ ) {

Generate the set of candidate  $k$ -itemsets,  $C_k$ , from  $L_{k-1}$ , the set of frequent  $(k-1)$  itemsets by applying  $L_{k-1} \times L_{k-1}$  (i.e. Cartesian product).

Scan the transaction to count the occurrences of itemsets in  $C_k$ ;

A subset of  $C_k$ ,  $L_k$  is created with count more than minimum support value ;}

→ Return  $L_1 \cup L_2 \cup L_3 \cup \dots \cup L_k$ ; [4][10]

4.3 Generating Candidate Itemsets and Frequent Itemsets (Min\_supp=2 (20%))

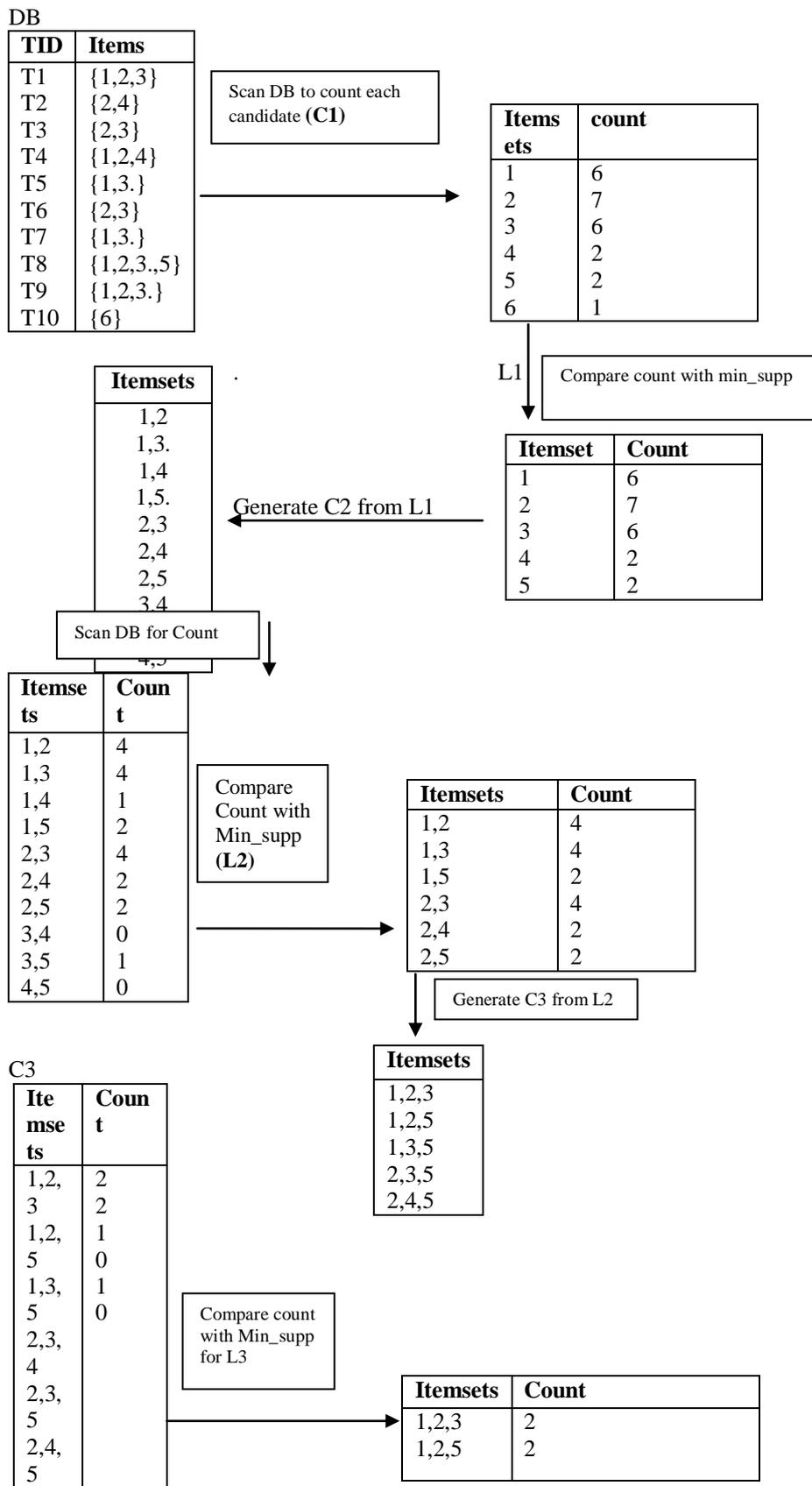


Figure 3 Generation of Candidate Itemsets and Frequent Itemsets (Min\_supp=2 (20%))

4.3.1 Advantages of apriori

- Easy implementation.
- ‘Apriori’, this word is originated from Latin. That means “from what comes before”.

- Initial Information- transaction database D and user-defined minimum support threshold Min\_supp.
- Algorithm uses information from previous steps to produce the frequent itemsets [8].

### 4.3.2 Limitations of Apriori

- It only explains the presence and absence of an item in transactional databases.
- In case of large dataset, this algorithm is not efficient [4].
- In Apriori, all items are treated equally by using the presence and absence of items.
- Apriori algorithm requires large no of scans of dataset [4].
- In this Algorithm, Minimum support threshold used is uniform. Whereas, other methods can address the problem of frequent pattern mining with non-uniform minimum support threshold [11].
- In case of large dataset, Apriori algorithm produce large number of candidate itemsets. Algorithm scan database repeatedly for searching frequent itemsets, so more time and resource are required in large number of scans so it is inefficient in large datasets [7].

### 4.3.3 Ways to Improve Apriori

- Transaction Reduction: transactions that do not consist of frequent itemsets are of no importance in the next scans for searching frequent itemsets [14].
- Hash based itemset counting: hashing table is used for counting the occurrences of itemsets.
- Partitioning: for any itemset i.e. frequent in database, then that itemset must be frequent in atleast one of the partition of database [6].
- By adding attribute Weight and Quantity: means how much quantity of item has been purchased.
- By adding attribute Profit: that can give the valuable information for business and customers.
- By reducing the number of scans.
- By removing the large candidates that cause high Input/output cost.

### 4.4 Working of Apriori Algorithm Using Data Mining Tool

bank contain huge data base, as the population increases day by day, database also increasing at alarming rate. Generally there are complex queries are that are difficult to answer in such large databases. So it becomes important to extract the useful information using an efficient mining algorithm to answer these questions. Following are the steps of working of Apriori algorithm using data mining tool Weka [15][3]:

Figure 4 shows how to import database in this tool. Figure 5 shows the preprocessing of database; in this figure at the left hand side list of attributes of database is given.

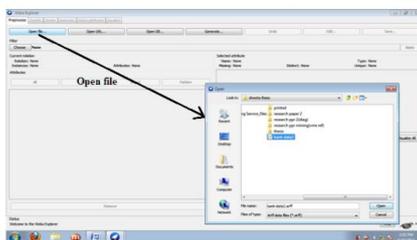


Figure 4 Importing Database

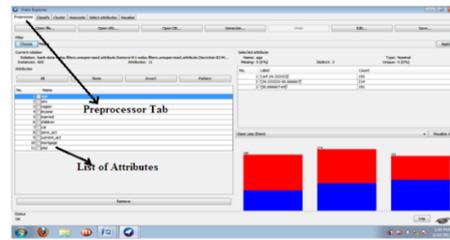


Figure 5 Preprocessing (Selection of Attributes)

In this example filtering and discretization are performed. In filtering irrelevant attributes are removed, in this example customer id is removed. Apriori only work on categorical attributes. In this example, age, income and children attributes are continuous. For converting them into categorical discretization is performed. Figure 6 elaborate the selection of different parameters for Apriori for mining frequent patterns. Different options available for Apriori are car, class index, delta, lower bound, min support, metric type, minimum metric, number of rules etc shown in figure 6. Figure 7 shows the ten best association rules using Apriori.

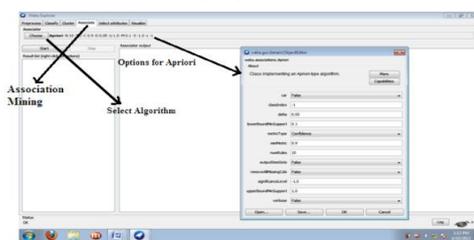


Figure 6 Selecting Parameters in Apriori

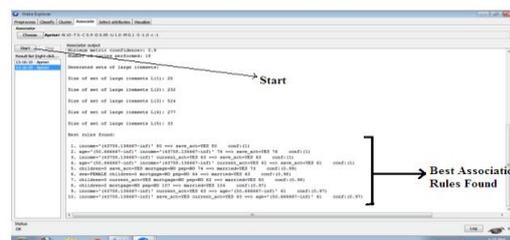


Figure 7 Association Rules Found in Apriori

GUI in WEKA only produces association rules and show number large itemsets L1, L2 etc. for frequent itemsets, Command Line Interface (CLI) is used shown in Figure 8. Type command in space given below. For example Apriori, type `java weka.associations.Apriori -N 10 -T 1 -C 1.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -I -t C:\Users\vishal\Desktop\bank-data2.arff`

Here -I is added for frequent itemsets. Figure 9 shows the association rules and frequent itemsets for Apriori using CLI [15].

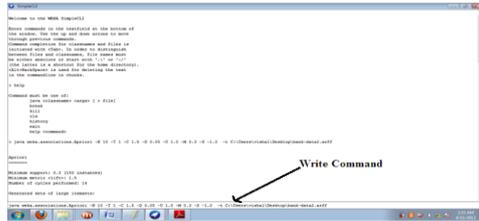


Figure 8 Command Line Interface



Figure 9 Association Rules in CLI

#### 4.5 Comparative Study of Association Rule Algorithms

here author considers three association rule algorithms i.e. PredictiveApriori association rule, Tertius association rule, Apriori Association rule. Author compares the results of these association rule algorithms with the help of data mining tool. Apriori Association Rule Algorithm is explained in the previous sections. Now brief description of other two algorithms i.e. PredictiveApriori Association Rule Algorithm and Tertius Association Rule Algorithm is given:

##### 4.5.1 PredictiveApriori Association Rule Algorithm

In predictive Apriori association rule algorithm, support & confidence is combined into a single measure called predictive 'Accuracy'.

{Support, Confidence}=> Accuracy

In this predictiveApriori association rule algorithm, this predictive accuracy is used to generate the Apriori association rule. In Weka, this algorithm generates "n" best association rule based on 'n' specified by the user

##### 4.5.2 Tertius Association Rule Algorithm

This algorithm finds the rule according to the confirmation measures. It uses first order logic representation. It includes various option or parameters like class Index, classification, confirmation Threshold, confirmation Values, frequency Threshold, horn Clauses, missing Values, negation, noise Threshold, number Literals, repeat Literals, roc Analysis, values Output etc [13].

##### 4.5.3 Experimental Result

Author finds the results using these three association rule algorithms. In this author uses bank data for comparison with 11 attributes and 600 records. And find association rules using Weka.

Table 1 represents the results using PredictiveApriori AssociationRule Algorithm. First column represents the list of relevant attributes after filtering the irrelevant attributes in this example customer id is removed.

Table 2 represents the results using Tertius Association Rule Algorithm.

Table 3 represents the results using Apriori Association Rule Algorithm (given at the end of the paper).

The ranges of income attribute for convenience can be denoted as: '(-i-24386]'→'(-inf-24386.173333]', '(24386-43758]'→ '(24386.173333-43758.136667]', '(43758-i)'→'(43758.136667-inf)'

And age attribute can be denoted as: '(-i-34]'→'(-inf-34.333333]', '(34-50]'→'(34.333333-50.666667]', '(50-i)'→'(50.666667-inf)'

Following figure 10 compares the time taken by three algorithms.

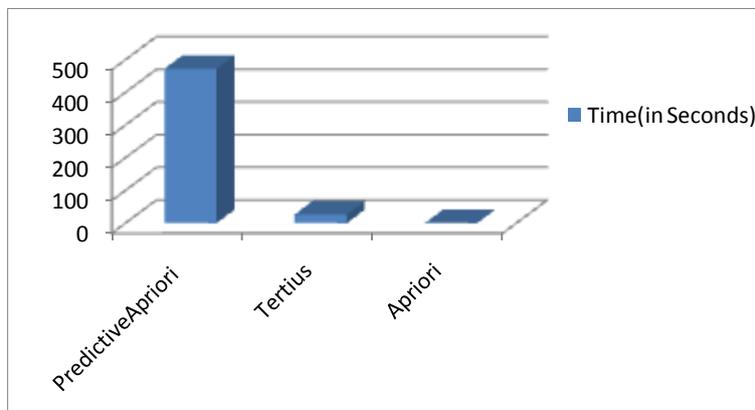


Figure 10 Comparison of PredictiveApriori, Tertius and Apriori on the Basis of Elapsed Time

#### 5. Conclusion

In this author tried to find the best association rules using data mining tool Weka. Based on the result author gives the advantages and limitations of Apriori association rule algorithm. And author discussed the ideas for improving the efficiency of Apriori association rule algorithm. And in second part author compared the association rule produced using

three association rule algorithms i.e. PredictiveApriori association rule algorithm, Tertius association rule algorithm and Apriori association rule algorithm. After comparing elapsed time by these three association rule algorithms, author finds that Apriori is faster than other two algorithms. **Future Scope:** Therefore these algorithms can be used in other domains to bring out interestingness among the data present in the repository. Association rules produced by these three algorithms can be combined for better results for any real life application. Algorithms can also be combined to for an efficient algorithm.

**REFERENCES**

- [1] Rakesh Agrawal, T. Imieliński, A. Swami, "Mining association rules between sets of items in large databases". In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93, 1993 pp. 207-216.
- [2] Sanjeev Rao, Prianka Gupta, "Implimenting Improved Algorithm Over APRIORI Data Mining Association Rule Algorithm", In: proceeding of IJCST, ISSN 0876-8491, VOL.3, Issue 1, Jan-March 2012.
- [3] Rayner Alferd , "Knowledge Discovery: Inhacing Data Mining and Dicision Support Integration", The University of York, Artificial Intelligence Group Department of Cpmputer Science, The University of York, United Kingdom, 2005.
- [4] Mamta Dhanda, Sonali Guglani, "Mining Efficient Association rules Through Apriori Algorithm Using Attributes", In: Proceeding of IJCST, ISSN 0876-8491, Vol. 2, Issue 3, September.
- [5] Jaishree Singh, Hari Ram, Dr. J.S. Sodhi, "Improving Efficiency of Apriori Algorithm Using Transaction Reduction", In: proceeding of International Journal of Scientific and Research Publication (IJSRP), ISSN 2250-3153, Volume 3, Issue 1, January 2013.
- [6] Jogi.Suresh, T.Ramanjaneyulu, "Mining Frequent Itemsets Using Apriori Algorithm", In: Proceeding of International Journal of Computer Trends and Technology, ISSN 2231-2803, Vol. 4, Issue 4, April 2013.
- [7] Suhani Nagpal, "Improved Apriori Algorithm Using Logrithmic Decoding and Pruning", In: Proceeding of International Journal of Engineering Research and Applications, ISSN 2248-9622, Vol. 2, Issue 3, pp. 2569-2572, May-June 2012.
- [8] K. Geetha, Sk. Mohiddin, "An Efficient Data Mining Technique for Generating Frequent Item Sets", In: Proceeding of IJARCSSE, ISSN 2277-128X, Vol. 3, Issue 4, April 2013.
- [9] Sunita B.Aher, Lobo L.M.R.J., "A Comparative Study of Association Rule Algorithms for course Recommender System in E-learning", In: Proceeding of IJCA, ISSN 0975-8887, Vol. 39-No.1, February 2012.
- [10] Partibha Parikh Dinesh Waghela, "Comparative Study of Association Rule Mining Algorithms", In: Proceeding of UNIASCIT, ISSN 2250-0987, Vol. 2, Issue 1, 2012.
- [11] Badri Patel, Vijay K Chaudhary, Rajesneesh K Karan, YK Rana, "Optimization of association Rule Mining Apriori Algorithm Using ACO", In: Proceeding of IJSCE, ISSN 2231-2307, Volume-1, Issue-1, March 2011.
- [12] Han, J., Pei, and Yin: "Mining frequent patterns without candidate generation". In: proceeding of the 2000 ACM SIGMOD International Conference on Management of Data, pp.1-12. ACM Press, New York, 2000.
- [13] S. Anupma Kumar, Dr. Vijaylakshmi M.N., "Discerning Learner's Education Using Data Mining Techniques ", In: proceeding of the International Journal On Integrating Technology in Education, Volume-2, No-1, March, 2013.
- [14] Rachna Somkunwar, "A Study on Various Data Mining Approaches of Association Rules", In: proceeding of International Journal of Advanced Research in Cmputer science and Software Engineering, ISSN 2277-128X, Volume-2, Issue-9, Page-141-144, September-2012.
- [15] Weka(2007).http://www.cs.waikato.ac.nz/ml/weka/ dated on May 10, 2013.

**Table 1** Results of PreidictiveApriori Association Rule Algorithms Using Data Mining Tool

<b>PreidictiveApriori Association Rule Algorithm</b>	
List of Attributes	Best Rules Found
1. age	1. income='(43758-i)' 80 ==> save_act=YES 80 acc:(0.99499)
2. sex	2. sex=FEMALE income='(43758-i)' 41 ==> age='(50-i)' save_act=YES 41
3. region	acc:(0.99492)
4. income	3. income='(43758-i)' car=NO 36 ==> age='(50-i)' save_act=YES 36 acc:(0.99489)
5. married	4. age='(34-50]' children=1 current_act=YES 35 ==> pep=YES 35 acc:(0.99488)
6. children	5. income='(43758-i)' married=YES pep=YES 31 ==> age='(50-i)' 31 acc:(0.99483)
7. car	6. income='(43758-i)' mortgage=YES 25 ==> age='(50-i)' 25 acc:(0.99468)
8. save_act	7. children=3 save_act=NO 22 ==> pep=NO 22 acc:(0.99453)
9. current_act	8. income='(43758-i)' married=YES pep=NO 22 ==> children=0 22 acc:(0.99453)
10. mortgage	9. age='(50-i)' married=YES children=1 car=YES 20 ==> save_act=YES 20
11. pep	acc:(0.99438)
	10. region=TOWN income='(43758-i)' 16 ==> age='(50-i)' save_act=YES 16
	acc:(0.99378)
Elapsed Time of PredictiveApriori Association Rule Algorithm - 472.478s	

**Table 2** Results of Tertius Association Rule Algorithms Using Data Mining Tool

<b>Tertius Association Rule Algorithm</b>	
List of Attributes	Best Rules Found
1. age	1. age = '(-i-34]' ==> income '(-i-34]'
2. sex	2. income = '(-i-24386]' ==> age = '(-i-34]'
3. region	3. age = '(-i-34]' ==> income = '(-i-24386]' or children = 3
4. income	4. age = '(-i-34]' ==> region = SUBURBAN or income = '(-i-24386]'
5. married	5. income = '(-i-24386]' ==> save_act = NO or age = '(-i-34]'
6. children	6. income = '(-i-24386]' ==> children = 3 or age = '(-i-34]'
7. car	7. age = '(-i-34]' ==> income = '(-i-24386]' or save_act = NO
8. save_act	8. income = '(-i-24386]' ==> region = SUBURBAN or age = '(-i-34]'
9. current_act	9. current_act = YES and age = '(-i-34]' ==> income = '(-i-24386]'
10. mortgage	10. age = '(-i-34]' ==> income = '(-i-24386]' or children = 1
11. pep	
Elapsed Time of Tertius Association Rule Algorithm- 26.941s	

**Table 3** Results of Apriori Association Rule Algorithms Using Data Mining Tool

<b>Apriori Association Rule Algorithm</b>	
List of Attributes	Best Rules Found
1. age	1. age='(-i-34]' 195 ==> income='(-i-24386]', current_acnt=YES 138, conf:(0.71) < lift:(1.97)> conv:(2.16)
2. sex	
3. region	2. income='(-i-24386]' current_acnt=YES 215 ==> age_='(-i-34]' 138 conf:(0.64) < lift:(1.97)> conv:(1.86)
4. income	
5. married	3. income='(-i-24386]' 285 ==> age='(-i-34]' car=NO 100 conf:(0.35) < lift:(1.97)> conv:(1.26)
6. children	
7. car	4. age='(-i-34]' car=NO 107 ==> income='(-i-24386]' 100 conf:(0.93) < lift:(1.97)> conv:(7.02)
8. save_act	
9. current_act	5. age='(-i-34]' 195 ==> income='(-i-24386]' pep=NO 111 conf:(0.57) < lift:(1.94)> conv:(1.62)
10. mortgage	
11. pep	6. income='(-i-24386]' pep=NO 176 ==> age='(-i-34]' 111 conf:(0.63) < lift:(1.94)> conv:(1.8)
	7. age='(-i-34]' 195 ==> income='(-i-24386]' save_acnt=YES 106 conf:(0.54) < lift:(1.91)> conv:(1.55)
	8. income='(-i-24386]' save_acnt=YES 171 ==> age='(-i-34]' 106 conf:(0.62) < lift:(1.91)> conv:(1.75)
	9. income='(-i-24386]' 285 ==> age='(-i-34]' mortgage=NO 113 conf:(0.4) < lift:(1.9)> conv:(1.3)
	10. age='(-i-34]' mortgage=NO 125 ==> income='(-i-24386]' 113 conf:(0.9) < lift:(1.9)> conv:(5.05)
Elapsed Time of Apriori Association rule Algorithm- 0.546s	