The Association Rule Optimization Technique Uses Genetic Algorithm To Improve The Quality Of Association Rule

NainaChoukade[1], Pooja Hardiya[2]
[1, 2] Dept of Computer Science and Engineering
[1, 2] Rajiv Gandhi Technical University
Bhopal (M.P.)

Abstract- *Data mining empower us for analyzing a large amount of data by minimizing the user efforts. These techniques utilize the computational algorithms that process the input information and produce the essential outcome patterns by which classification, prediction, and clustering are performed. In the comprehensive system of mining organization algorithms the mining process has its significance.This technique is used to find the relationship between the data attributes based on frequencies or occurrence of two or more attributes. This combination of attributes helps to make predictions and relationships building among two attributes. In this paper the organization managing mining strategies is the main learning area and learning two popular algorithms. In the study that is found the developed rules from both the algorithms are sometimes different andcan affect the prediction.The work therefore involves a genetic algorithm for compliance with a set of rules produced by both mining algorithms. During optimizing rules first common rules from both the algorithms are picked and then the remaining rules are combined as final outcome.The implementation of this program is done using JAVA technology. The performance shows the proposed technique optimizes the set of rules and having the good a quality of rules for different data mining and machine learning operations.*

Keywords- Data Mining, Association Rule, FP-Tee, Apriori, Frequent Pattern mining, Association Rule Mining, Genetic Algorithm

I. INTRODUCTION

In data mining, there are a number of techniques in data mining available that are used for modelingof data. These techniques basically used for the representation of using different forms such as using weights, trees, graphs, rules and others. Using this representation of the data model the data mining applications recover the fruitful patterns by which decision making, prediction, and other valuable tasks are performed. Among them, the Association rules for mining plays an important role in mining research there point is to discover fascinating relationships between arrangements of things in information bases. Basically, the association rule mining techniques are helpful for exploring and analyzing the transactional data such as web server access logs, any e-commerce store transactions, and other similar kinds of data. Using these data the store owner or the web server administrator makes effective decisions about the website's contents and store items that are purchased by the users.

In this context, the proposed work is focused on exploring and investigating the techniques of association rule mining. Therefore there are two popular algorithms namely the apriori algorithm and the FP-Tree algorithm are explored for addressing the issues in rule generation. In addition to that, a new method using the genetic algorithm is proposed that utilizes the generated rule sets by both the algorithms for optimizing the rules and obtaining the more fruitful rules using both the algorithms. The aim of the proposed optimization technique is to discover the potential rules from both the algorithms.

II. LITRETURE SURVEY

This section includes the different techniques and applications that are developed recently or implemented to improve the association rule mining techniques.

There is widespread access to big data steps and there is a need to convert this data into important data as information. This interest in knowledge disclosure has prompted the improvement of numerous algorithms to decide the affiliation rules. One of the most serious problems with these algorithms is the age of the candidate sets. But the FP-Tree algorithm provides organizational rules without producing candidate sets.But in the process generates many CP-trees. ***Vandit Agarwal et al. [14]***proposed FP tree algorithm with modified header table, backup table and MFI algorithm for organization mines.This algorithm generates normal object sets without using candidate sets and CP trees.

Extracting the desired information and data extraction is associated with a larger data size. For data

extraction there are various strategies. Detects hidden or pattern in big data. Among the existing strategies the FP growth algorithm is an effective algorithm for organizational rules. For analysis, the site is scanned. Producing a large number of conditional FP trees, this is a problem for FP growth. ***Meera Narvekaraetal***. ***[15]*** devised a strategy to excavate all non-permanent objects except for years of adjacent FP trees. Not at all like FP tree it filters the information base just once which decreases the time. It additionally discovers the recurrence of the thing sets to discover the ideal rules.

Collecting data that grows exponentially becomes a herculean task. In order to exploit common mining resources, organizational mining is used in data mining. Parallelizing the algorithm at different degrees of computation won't just accelerate the cycle yet additionally permit it to deal with scalable data. ***SheetalRathi et al. [16]*** proposes a model to parallelize the regular itemset mining measure without extra burden utilizing superior figuring. GPU has been utilized which offer better execution effortlessly and furthermore energy productive when contrasted with multi-center multiprocessors.

TamirTassa et al. [17]to propose a law of corporate protected mining laws on horizontally distributed information. This legal process, similar to theirs, is based on the Fast Distributed Mining (FDM) algorithm, which is a distributed unstable version of the Apriori algorithm. Basic protocol adjustments are algorithms for two secure multi-team algorithms — one comprising a union of sub-sets owned by each of the speaking actors, and another that evaluates the inclusion of one component in a small set in another. This protocol offers improved protection regarding the protocol. Likewise, it is less complex and is fundamentally more effective as far as correspondence adjusts, correspondence cost and computational expense.

The acquisition of organizational rules is a growing area of interest. A challenging point is to find the rules for numerical values with a numerical value system. Many methods of compound law cannot be applied to numerical data. Solving the problem of dealing with numerical data is an ongoing effort.Approaches suffer from problems are discussed by ***B. Minaei-Bidgoli et al. [18]*** and proposes a multi-purpose algorithm for mining organization rules. To determine efficient rules, several measures are there. Confidence, Interest, and Understanding are three methods that have been used. Finally, he found the best rule. This method uses rough numbers, based on the concept of rough patterns, defined by the upper and lower intervals to represent a set of values. Transformation and crossover operators offer powerful testing capabilities and allow you to find the best numerical intervals. The rules produced by this method as a result of the results are very appropriate - based on a few factors.

Among all the other mining algorithms based on the Apriori system, organization rules, mining resources and organizations of interest on the site, not only the organization's mining process is used. It turns out that there are two major issues in the traditional Apriori algorithm scanning regular data and a large number of candidate sets construct. Therefore, some related improvements are made: 1) using a database map to avoid scanning; 2) regular pruning of sets of items and sets of selected items to improve efficiency; 3) use a scattering strategy to calculate support to achieve maximum efficiency. The advanced Apriori algorithm is proposed by ***Xiuli Yuan et al. [19]*** compared with other improved algorithms improved performance.This method uses rough values, based on the notion of rough

III. PROPOSED WORK

The proposed work is intended to optimize the association rules for finding the best possible rules generated by different rule generation algorithms.This chapter provides a comprehensive understanding of the proposed strategy for performing the required task

A. System Overview

In data mining and machine learning, legal data modeling is used in a few applications. Rules can be used to identify common pattern, recognition, classification, prediction, and decision making. Law-based data modeling is therefore an important source of data mining.The rules are also extracted using the association mining techniques and can also be used for the same kinds of applications. The association based rules are developed based on the attributes frequency and the respective frequencies of different other attributes. Using these computed frequencies of the participating items the rules are developed that are used for prediction and decision making.

In this work, the apriori algorithm and FP-Growth algorithm is used for computing the association rules. Rule mining using the apriori algorithm is computationally expensive and the FP-Tree does not provides clearer rules due to multiple tree generation. Therefore both techniques are not much effective for acceptance. In this context, a technique is suggested in this work for the selection of the optimal rules from both the generated rule sets. Therefore the selected rules are used to map all the search or problem space on which the rule generation is performed. To make the selection of the

required rules, a genetic algorithm is used. The genetic algorithm is a search algorithm that helps to select the same operating rules for both sets of rules. Additionally, their Metaheuristic helps to choose the strongest quality of rules. In this section, an overview of the model is presented in the next section on how to select a rule based on the genetic algorithm is explained.

B. Methodology

The description of the proposed methodology is defined in figure 3.1. This diagram contains the different functional blocks that are helping to process the data for finding the required outcomes.

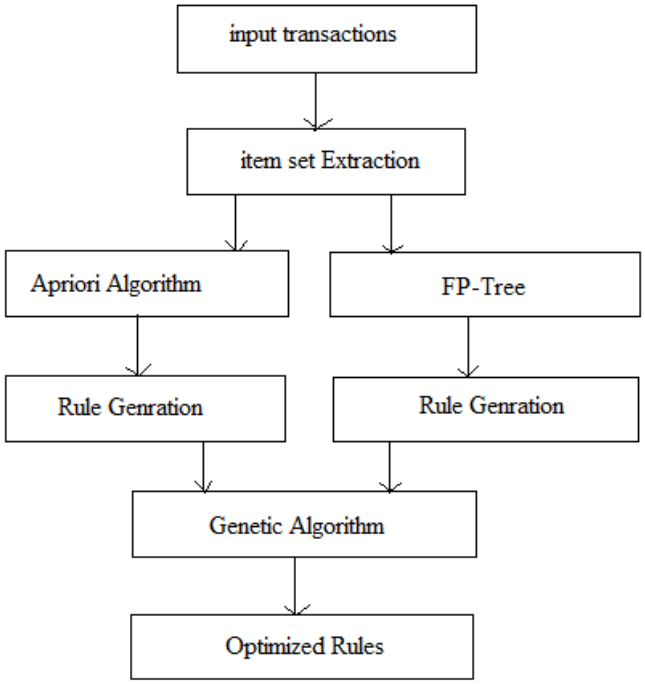

Figure 3.1 proposed Rule Generation Technique

Transaction Set: Machine learning and data mining algorithms are works with a special kind of data format by which the algorithm computes required patterns. In the context of mining law, data is obtained through transaction. These activities are grouped by a series of limited objects. In addition to using these limited resources all functions are improved.

Item set extraction: The inclusion of available data is a practice, therefore, all transactions are scanned in this category and the items found in these transactions are subtracted by their calculation using this number of items the calculated frequencies.

Apriori algorithm: The Apriori algorithm initially uses object sets to generate a potential candidate set. This candidate set is utilized with the confidence for computing the frequent patterns from the input transaction set.

FP-Tree algorithm: The FP-Tree algorithm produces standard object sets and based on these pre-set objects the tree structure is computerized. This tree starts with a null node and based on the size of the computer on which the tree grows. This tree is also used for law enforcement.

Rule Generation Apriori: The apriori algorithm provides standard object-based sets that are always set the rules are developed using the apriori algorithm effect.

Rule Generation FP-Tree: The FP-Tree algorithm generates tree structures to reflect organizational rules. These trees are transformed into this phase of finding rules through these tree structures. Thus all tree branches are transformed into organizational rules produced by FP-Tree.

Genetic Algorithm: Broadlythe genetic algorithm is a search algorithm. That algorithm accepts the query sequence as input and the other hands the database sequences for finding the best match sequences to the target sequences. In this proposed scenario both the generated rule sets are provided as input to the genetic algorithm. Among both of them, one set of rules are considered as the population set or solution set and the other set of rules are considered as the query for the search process.Therefore one rule is picked from the one rule set and the second rule set is treated as the population set for the genetic algorithm. The genetic algorithm compares the selected rules from all the given rules from other algorithm and based on their fitness value the common set of rules are generated.

Optimized Rules: That is the final outcome of the genetic algorithm, after processing the generated rules from both the algorithms the selected rules through the genetic algorithm is produced in this step as resultant of the system.

C. Proposed Algorithm

This section explains the proposed steps of an algorithm that summarizes the above listed methodology process. Table 3.1 contains the required algorithm steps:

Table 3.1 proposed algorithm

Input: input transactions T Output: optimized rules R
Process: 1. $D_n = ReadTransactionalDataset(T)$ 2. $for(i = 1; i \leq n; i++)$ a. $I_i = ExtractItemset(D_i)$ 3. $end\ for$ 4. $AR_n = Apriori.ComputeRules(I_n, D_n)$ 5. $FP_m = FPTree.GenrateRules(I_n, D_n)$ 6. $for(i = 1; i \leq n; i++)$ a. $P_i = SelectARule(AR_n)$ b. $R_i = GeneticAlgo.Optimize(R_i, FP_m)$ 7. End for 8. Return R

IV. RESULTS ANALYSIS

The performance of the proposed hybrid approach over implemented two data mining algorithm using genetic algorithm for generating classification rule are evaluated and analysis in this chapter. Consumed memory and required time for rules generation are performance factors that are evaluated and provided.

A. Memory Consumption

The measure of primary memory needed to perform information examination utilizing the calculation is named here as memory utilization or space multifaceted nature. The estimated memory consumption of the implemented algorithm i.e. genetic is reported using figure 4.1 and table 4.1. Following is the equation by which we can assess expended memory:

$$\text{Memory Consumption} = \text{Total Memory} - \text{Free Memory}$$

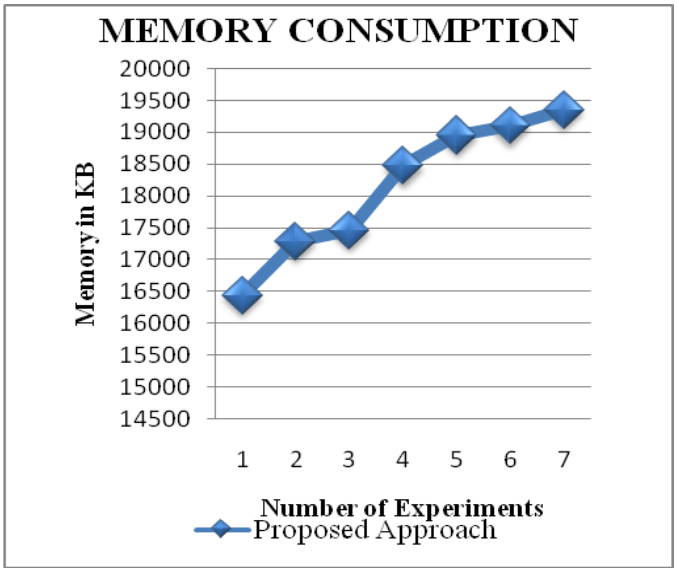

Figure 4.1 Memory Consumption

Table 4.1 Tabular form of Memory Consumption

Number of Experiments	Proposed Approach
1	16439
2	17284
3	17462
4	18472
5	18948
6	19094
7	19347

The use of memory or complexity of the mixed method used FP-Tree and Apriori algorithm generated by a viable solution using genetic algorithm is reported with the assistance of Figure 4.1 and Table 4.1. In this graph, the X-axis contains a variety of frame tests, and the Y-axis shows the amount of vaccine memory spent per KB (kilobytes). From the diagram provided, it clearly shows that the production of rules using a hybrid-based genetic algorithm consumes less memory than other standard data mining mining organization models.

B. Time Consumption

In order to process the FP tree the extended time scale, and the inclusion data is named here used in the Apriori-based organization rules such as the use of calculation time or the complexity of time. Use time to find under the given formula:

$$\text{Time Consumed} = \text{End Time} - \text{Start Time}$$

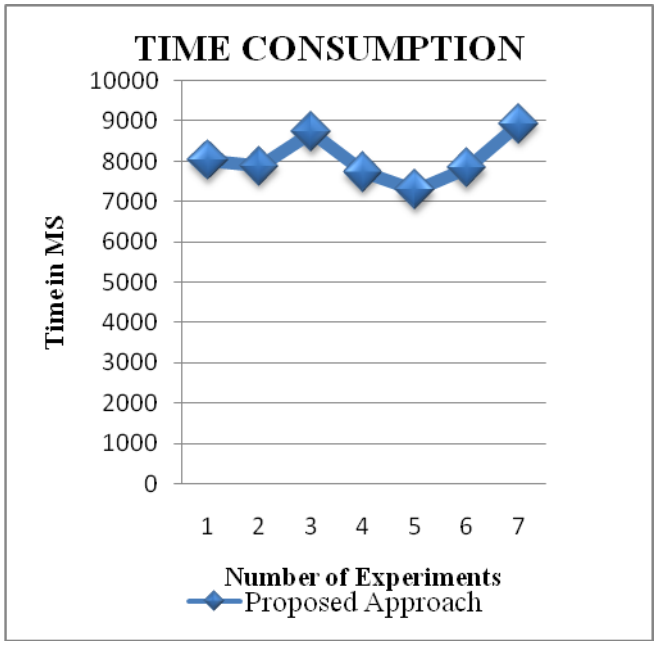

Figure 4.2 Time Consumption

In figure 4.2 and table 4.2 reported the use of the time method used. Different system-based tests are shown on the X-axis and the amount of time spent to produce the law of aggregation according to the milliseconds displayed on the Y-axis.According to the given performance, less amount of time is consumed by the implemented genetic algorithm of rule generation whereas apriori an FP-tree consumes more time for generating rules. But time management is increasing as the amount of data for the development of organizational rules grows.

Table 4.2 Tabular form of time consumption

Number of Experiments	Proposed Approach
1	8018
2	7882
3	8717
4	7726
5	7291
6	7829
7	8928

C. Number of Rules

Using minimum confidence and minimum support rules are generating in association rule mining that initially set for frequent mining. For finding the rule formula is given for calculating the rule for the implemented algorithm:

Table 4.3 Tabular form of Rules Generation

Number of Experiments	Optimized Genetic Rules	Total Rules
1	21	36
2	26	40
3	27	48
4	19	38
5	22	36
6	26	42
7	28	46

$$FuzzySimilarityOverlapping = \frac{1}{1 + D_m(a,b)}$$

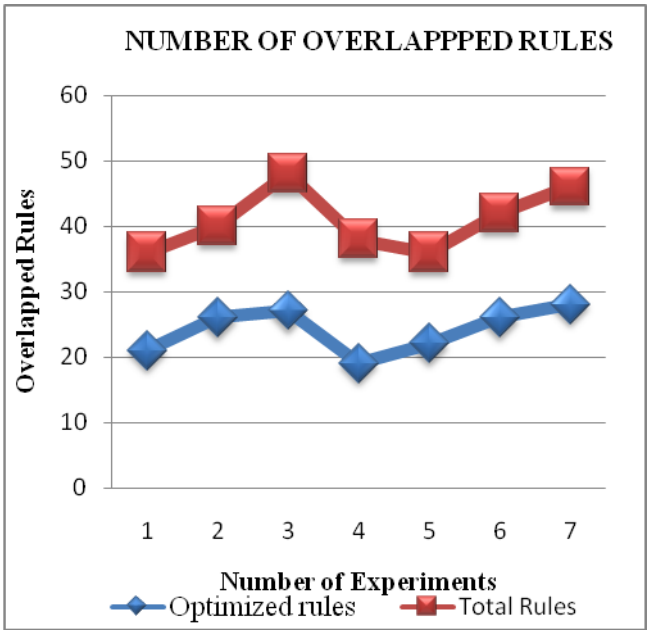

Figure 4.3 Rule Comparison

In figure 4.3 and table 4.3 the rule generation of the implemented algorithms of Apriori algorithmand the FP-tree along with the optimized rule that minimizes the rule is reported. The numbers of runs performed with the system are shown on X-axis and On the Y-axis rules generated using an algorithm are shown, at the same time the previous mining law of the organization is provided. Based on the function provided, the algorithms used by FP-Tree and apriori show rules that reflect the most appropriate law for matching.

D. Lift

The lift value of organizational law is part of the legal trust and the expected trust of the law. The expected confidence in the law is seen as a product of the supporting values of the legal entity and the head of law divided by the support of the legal entity.

$$Lift(X,Y) = \frac{frequency(X,Y)}{Support(X) * Support(Y)}$$

Table 4.4 Lift

Number of Experiments	Lift
1	4.72
2	6.28
3	5.47
4	4.58
5	5.44
6	4.29
7	4.94

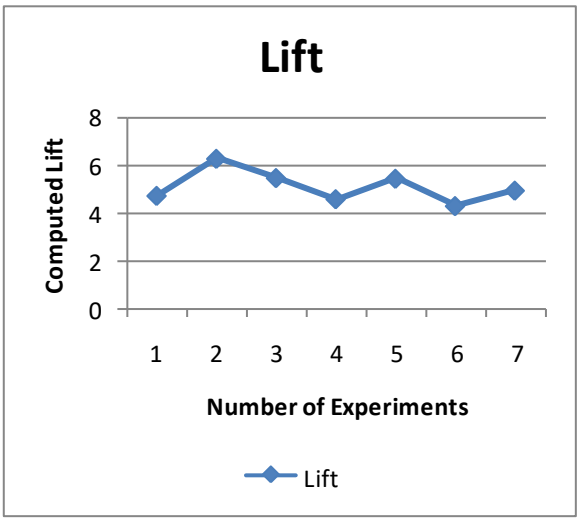

Figure 4.4 Lift calculation

The computed lift value for the optimized rule set is given using figure 4.4 and table 4.4In this figure, the X-axis shows the various tests performed on the frame and the Y-axis shows the height of the obtained rules. The proposed rules reflect the level of legislation. Given the high value of the acquired legislation, the proposed method has been found to be appropriate and effective in implementing the real-world law development approach.

V. CONCLUSION AND FUTURE WORK

The chapter provides a summary of the proposed conducted work for optimizing the generated rule set for association rule mining. Therefore the conclusion of the work based on experiments performed and future extension of the work is also listed in this work.

A. Conclusion

Data mining methods help to understand data and patterns hidden in data.To discover and explore an available set of information,data mining based computational algorithms are utilized. These algorithms compute weights, relationships, and other measurements for establishing the data models. These data models are used for various tasks such as segmentation, disposal, decision making, and more. Organizational mining is one of the most popular methods of data mining technology. The association rule mining techniques establish the relationship among the items by which predictions are possible based on the items combinations. For example in a store the possibility of purchasing two items in combination or predicting the items which can be purchased one after another.

In this presented work two popular data mining techniques namely apriori algorithm and FP-Growth algorithms are considered. Both the algorithms are frequently used for association rule mining is transactional databases. But both the models have their limitations such as ambiguity in rule generation, resource consumption, and others. Therefore which rules are most beneficial for the different applications are very complex to identify. In this context, the proposed work involves the ability of a genetic algorithm for the selection of strong rules from both the generated sets of rules. The technique accepts both the rules sets and finds the similar and common rules among them based onthe fitness values of the genetic algorithm.

JAVA Technology is used to implement the proposed strategy. Additionally, tests on different data sets were performed to detect variations in performance. Results obtained based on results are shown using Table 5.1.

Table 5.1 performance summary

S. No.	Parameters	Remark
1	Memory usages	The memory usage of the proposed algorithm is consistent and does not differ unequally in different tests
2	Time consumed	The time consumption of the proposed technique is increase according to the number ofrules produced for optimization
3	Number of generated rules	The number of rules are obtained in regular basis and only the selected rules are used for applications
4	Lift	The lift demonstrate the quality of rules according to the obtained lift values the system able to generate high-quality association rules

According to the obtained results, the proposed technique is suitable for selecting the optimal and strong rules from the classical association rule generation techniques.The proposed work is therefore welcome in the use of real-world applications.

B. Future work

The main aim of the proposed work is to combine the association rules generated by the apriori and FP-tree for finding the most effective rules among both the techniques. The implementation and performance analysis says the proposed technique works as required and help to pick up the

optimal rule among both of them.In the future, the following expansion of the project is proposed:

1. The proposed technique just used to search the common and strong rules among both of the algorithms generated rules. In near future, the work is extended for combining the non matched rules for optimizing them
2. Organizational rules can be re-created using genetically inspired algorithms so in the near future the work will be expanded to produce rules on the basis of metaheuristic algorithms.

REFERENCES

[1] Bhandari, Akshita, Ashutosh Gupta, and Debasis Das, "Improvised apriori algorithm using frequent pattern tree for real time applications in data mining", Procedia Computer Science 46 (2015): 644-651.

[2] N. Padhy, P. Mishra and R. Panigrahi, "The Survey of Data Mining Applications and Future Scope," International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), vol. 2, no. 3, June 2012.

[3] B. K. Baradwaj and S. Pal, "Mining Educational Data to Analyze Students' Performance," (IJACSA) International Journal of Advanced Computer Science and Applications, volume 2, Number 6, 2011.

[4] Han, Kamber (2001) Data mining concepts and techniques, San Francisco: Morgan Kauffmann Publishers.

[5] "Architecture of Data Mining", available online at: http://knowledgebounce.com/2016/03/14/architecture-data-mining/. [March 14, 2016]

[6] Dr. Poonam Chaudhary, "Data Mining System, Functionalities and Applications: A Radical Review", International Journal of Innovations in Engineering and Technology (IJIET), Volume 5 Issue 2 April 2015, pp. 449-455.

[7] Bhandari, S., Sharma, T., Singh, J., "A Review: Data Mining, its Issues, Functionalities and Applications", International Journal of Research (IJR), July 2014.

[8] Liao, Shu-Hsien, Pei-Hui Chu, and Pei-Yuan Hsiao, "Data mining techniques and applications–A decade review from 2000 to 2011", Expert systems with applications 39, no. 12 (2012), pp. 11303-11311.

[9] J. Usharani, Dr. K. Iyakutti, "Mining Association Rules for Web Crawling using Genetic Algorithm", International Journal Of Engineering And Computer Science, Volume 2 Issue 8 August, 2013 Page No. 2635-2640

[10] Agrawal, Rakesh, and RamakrishnanSrikant, "Fast algorithms for mining association rules." Proc. 20th international conference very large data bases, VLDB, Volume 1215, 1994.

[11] Agrawal R, Imielinski T, Swami A (1993) 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), Washington, DC, pp 207–216

[12] "Association Rules Mining", available online at: https://www.vskills.in/certification/tutorial/data-mining-and-warehousing/association-rules-mining/

[13] RanaIshita and RanaIshita, "Frequent Itemset Mining in Data Mining: A Survey", International Journal of Computer Applications (IJCA), Volume 139 – No.9, April 2016.

[14] Vandit Agarwal, MandhaniKushal and Preetham Kumar, "An Improvised Frequent Pattern Tree Based Association Rule Mining Technique with Mining Frequent Item Sets Algorithm and a Modified Header Table", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.2, March 2015

[15] Narvekar, Meera, and ShafaqueFatma Syed, "An Optimized Algorithm for Association Rule Mining Using FP Tree." Procedia Computer Science 45 (2015): 101-110.

[16] SheetalRathi and Dr.ChandrashekharDhote, "Parallelizing Frequent Itemset Mining Process using High Performance Computing", International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), March-May, 2014, pp. 58-63

[17] Tassa, Tamir. "Secure mining of association rules in horizontally distributed databases." IEEE Transactions on Knowledge and Data Engineering 26.4 (2014): 970-983.

[18] Minaei-Bidgoli, Behrouz, R. Barmaki, and Mahdi Nasiri, "Mining numerical association rules via multi-objective genetic algorithms." Information Sciences 233 (2013): 15-24.

[19] Yuan, Xiuli, "An improved Apriori algorithm for mining association rules", AIP Conference Proceedings, Vol. 1820, No. 1, AIP Publishing, 2017.