Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766155

FREQUENT ITEMSET GENERATION ON CUSTOMER DATASET USING ASSOCIATION RULE MINING

U. Latha
Assistant Professor, PG Department of Computer Applications, D.D.G.D Vaishnav College, Arumbakkam, Chennai, India. E-mail: dgvclatha@gmail.com

R. Velmurugan
Associate Professor, Department of Computer Science, Presidency College, Chepauk, Chennai, India. E-mail: vel_ram@yahoo.com

T. Velmurugan
Associate Professor, PG and Research Department of Computer Applications, D.D.G.D Vaishnav College, Arumbakkam, Chennai, India. E-mail: velmurugan_dgvc@gmail.com

Abstract: Data mining is an important tool in extracting interesting patterns from large datasets to represent knowledge. Association rule mining is one of the important concepts in Data Mining. Frequent itemset Generation is one the step in Association Rule Mining. In order to discover the relationships among the data items in large size of database, the most of the research activities focus on it. This research work is mainly implemented by focusing on the analysis of frequent itemset generation in customer dataset to find out customers buying behavior. The traditional algorithms Apriori and existing algorithm Cluster Based Bit Vector Association Rule Mining (CBVAR) and a proposed algorithms namely Improved Cluster Based Bit Vector Association Rule Mining (ICBVAR) are taken to find the efficiency of the algorithms in terms of its execution time and occupied space. A comparative analysis of the all the algorithms is carried out and the best algorithm is based on its performance that is suggested. From the experimental results, the proposed algorithm ICBVAR is faster and gives high recognition results.

Key words: Association Rule Mining, Apriori Algorithm, Cluster Based Bit Vector Association Rule Mining, ICBVAR, Frequent Itemset Generation.

I. INTRODUCTION

Data mining is an important tool in extracting interesting knowledge from large databases. It proposes many solutions for the extraction of significantly and potentially useful patterns from a large collection of data. Thus, mining process depends on the user requirements, who may be a business analyst or a marketing manager. The need for information differs from user to user. Data mining, also called knowledge discovery, it is the process of discovering interesting and useful patterns in large volumes of data, which have not previously been discovered, and relationships among them [1]. The different data mining techniques are suggested based on their requirements. One among them is association rule mining. Association rule mining is one of the most widely used data mining technique. Association rule mining finds all rules in the database that satisfies some minimum support and minimum confidence constraints.

This research work mainly focused on the analysis of frequent itemset generation in grocery itemset via finding customer behavior. The existing algorithms Apriori, Cluster Based Bit Vector Association Rule Mining (CBVAR) and proposed algorithms namely Improved Cluster Based Bit Vector Association Rule Mining (ICBVAR) is taken to find the efficiency of the algorithms in terms of its execution time and occupied space. A comparative analysis of the four algorithms is carried out and the best algorithm among these is based on its performance is suggested. Section II discusses about the literature survey used for this research work. The experimental results are illustrated in the section III. Finally, this research work concludes with its innovative information in section IV.

II. LITERATURE SURVEY

Association rule mining algorithms attempt to find out interesting relationships or hidden patterns among data. There are two steps involved in association rule mining: First is finding frequent itemset generation in transactional dataset, which is the major research field in data mining. Second is the discovery of mining frequent itemset that finds correlations among items in a large amount of transactional dataset. Applications of frequent itemset are in many domains including market basket analysis, financial flow analysis and in 2 several real-time based datasets. Frequent itemset mining enables the discovery of associations and correlations among items in large transactional or relational data sets. The enormous amounts of data continually being collected and stored, many organizations and industries are interested in mining such patterns from their databases. The discovery of these patterns and interesting correlation from the huge amount of business data helps in crucial decision making process such as cross-marking, customer buying behavior, and so on

Akshay palekar et al., [2] study on Data mining offers a better scope to extract of hidden information as it goes beyond the method of the traditional decision support systems which trusted mainly on analyzing of the past transactions of dataset. It plays vital role, very simple and cost effective to find frequent itemset. Aswathy Wilson et al., pointed out that the cause of death is heart disease is one of the reasons. There are many mining techniques that help to diagnosis of heart related diseases. In this comparison study of different mining algorithms such as K-Means Clustering with Decision Tree, Weighted Associative Classifier with Apriori algorithm and Naïve Bayes [3]. Baralis E, Cerquitelli T, Chiusano S et al., mention about P-Mine algorithm [4]. It is a parallel disk-based approach that proposes to mine frequent itemset. VLDBMine data structure is used to store data base. Multiple projections of the dataset have been loaded into different processor cores using a pre-fetching technique to mine frequent Itemset.

Bastide et al., revealed that the PASCAL algorithm is named after the French mathematician Blaise Pascal who invented an early computing device and it is an optimization of the Apriori algorithm. In this approach, the concepts of key patterns from frequent patterns are inferred without access to the database [5]. Chao et al., in their paper introduced an improved Apriori algorithm based on matrix [6]. In this algorithm, association rules are mined to form clusters and its length. The matrix construction for each cluster takes time and is not appreciable. Chavan et al., investigated FIM techniques that had applicability on the Map Reduce platform [7]. In this research work, two parallel algorithms, Dist-Eclat and Big-FIM algorithms are discussed. The results show that the 3-FIs together with even a basic Round-Robin allocation theme results in a good work distribution.

Dhanashree Shirke, Prof. Deepti Varshney observe it is a difficult to process to mining big data. Frequent itemset are obtaining from many different approaches. The traditional mining algorithm have some draw backs like lack of mechanisms of load balancing, data distribution and fault tolerance [8]. Erwin et al., proposed a CTU-PRO algorithm to find frequent itemset from both sparse and dense datasets. This is the main feature of it. Frequent pattern mining ascertains patterns in transaction databases centered on the relative frequency of occurrence of items sans considering their use [9].

Gayathri. G Observes that the frequent itemset play an important role in many data mining tasks that try to find the interesting patterns from databases such as association rules, correlations, sequence, classifier and clusters [10]. From the graph, it concludes FP-growth 21 algorithm is best to generate the frequent itemset for smaller datasets, whereas Apriori is best for larger datasets and éclat algorithm is used to generate the frequent itemset for smaller and larger datasets. Han et al., [11] proposed an algorithm in which the divide and conquer method is used to decompose the task into a set of smaller tasks for constructing and traversing of frequent pattern tree. Ilayaraja M, Meyyappan T describe the data mining is significant way of analyzing large amount of data on different perspectives and providing it into useful information on making decision. Finding frequent itemset generation plays vital role in mining [12].

Jadhav Kalyani B et al., [13] explains that there is an increasing in amount of data collected and ability to obtain the large amount of data increased importantly in modern era, because of modern software platform and up gradation of hardware. It reviews that parallel frequent pattern mining and analysis of it through the Big Data. Karthikeyan and Ravikumar performed theoretical survey on a number of existing algorithms. The ideas behind association rules are provided at the beginning followed by a summary of a few research works previously done related to this area. The advantages and disadvantages were discussed in detail and concluded with a suggestion [14]. Mall et al., proposed a Perturbed Frequent Itemset based Classification Technique (PERFICT) [15], a completely unique associative classification approach centered on perturbed frequent items. Experiments carried out on the UCI repository datasets demonstrate that PERFICT is highly economical with regard to accuracy in comparison with popular associative classification techniques.

III. EXPERIMENTAL RESULTS

In this research work, an improved Cluster based Bit Vectors for Association Rule mining (CBVAR) is proposed to discover frequent itemset efficiently. Finding frequent itemset and generating association rules are the two significant steps in the discovery of association rules. The similar items are grouped together to form a cluster which helps to make faster decisions and to explore data efficiently. The iterative process of creating the cluster is time consuming and requires more space. This issue is solved by the strategy of adding bit vector to each cluster. CBVAR which uses multiple tables for finding frequent itemset generation does not eliminate redundancy. To overcome the problems of CBVAR, ICBVAR is proposed which avoids redundancy. The traditional association rule mining algorithm Apriori and the existing algorithm CBVAR and the proposed algorithms ICBVAR are compared in terms of space and time complexity.

The data is collected from the web repository namely www. salemmarafi.com, which comprises groceries items to be used to find frequent itemset. The data sources are classified as item based. Some sample datasets are given as milk, beer, diapers, coke, bread, yolk, chicken, pip fruit, yogurt, cream cheese and meat spreads. The dataset contains 175 distinct items and 100K transactions. To benchmark the proposed algorithm, we need to create various sub datasets from this mother dataset based on parameters such as number of distinct items and number of transactions. Each sub dataset takes a 6 form as "#items #transactions.csv." For example, "3i10t.csv" means this sub dataset contains 3 distinct items and 10 transactions. Similarly, our largest dataset split is "175i100Kt.csv" meaning 175 distinct items and 100K transactions. Thus, the groceries dataset is split into sub-datasets of various sizes from 10Kt to 100Kt. Figure 1 shows the flow of methodology about this research work.

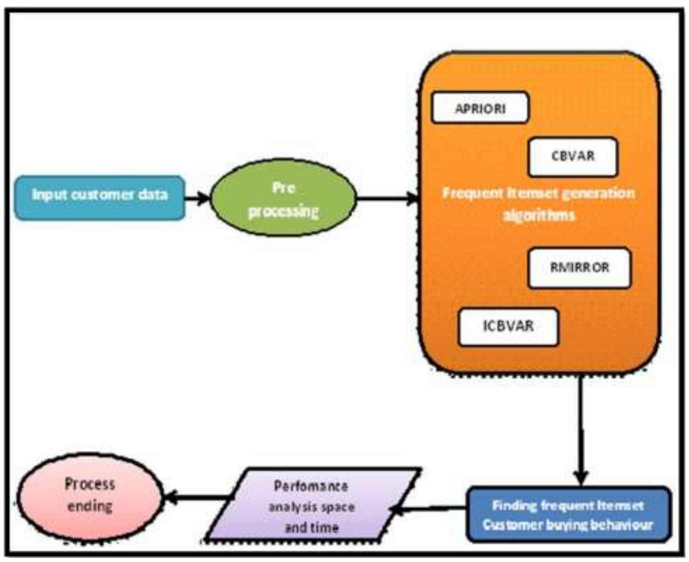

Figure 1. Flow of Methodology

The steps involved in the proposed method are as follows:

Step 1: Scan the customer data set. Before finding frequent item, the dataset that has to be preprocessed. That is, it has to be check whether it has any incomplete, noisy, inconsistent data, otherwise it may cause incorrect or even misleading result.

Step 2: Preprocess customer data using Median filter, Gaussian Filter, Weiner filter to remove the noise. The noisy data contains errors or outliers such as age="-28." It arises from data collection process, data entry, transmission. The filter methods are applied to preprocessing.

Step 3: Apply traditional algorithms Apriori, CBVAR and proposed algorithms ICBVAR to find frequent itemset generation. The Candidate set generation based algorithm Apriori and Bit Vector Based algorithm CBVAR are existing algorithms and the proposed algorithms ICBVAR are used for mining frequent itemset.

Step 4: Finding the customers buying behavior. Find the dataset which are appeared more times in the customer dataset. The experimental results are used to find frequent itemset. It helps to find Buying Behavior of customer.

Step 5: The results of all four frequent itemset generation algorithm findings are compared with respect to run time and memory space. Each algorithm consumes memory space for storing dataset and also takes time for processing to mine frequent itemset. These values will be taken for comparison of all algorithms.

Step 6: Find the performance of algorithms based on its accuracy. The three combination of algorithms Apriori, CBVAR, ICBVAR and CBVAR, ICBVAR are compared due to time and space for different size of dataset.

Experiments were done on both Mac OS X (Mountain Lion) & Windows 8 laptops running on Intel i5 Core processor with 4 GB RAM. Python, a high level programming language, which is used for implementing the proposed algorithm ICBVAR and RMIRROR algorithm and its parent algorithms - CBVAR and Apriori algorithm. The dataset used for comparing the performance of the proposed algorithms with CBVAR and Apriori algorithm is grocery dataset. This dataset contains about 100000 transactions of customers' buying behaviour.

Table 1 compares the time consumed of APRIORI, CBVAR and ICBVAR algorithm by applying 10 groups of transactions. The time difference between Apriority and the other is constantly increasing. For 1K, the time taken by APRIORI, CBVAR and ICBVAR are 17.80, 16.00, and 14.9 respectively. If it is 2K then the time consumed by the algorithm APRIORI is 18.16 sec. While CBVAR consumes 16.32 seconds for the same number of transactions, ICBVAR consumes 15.19 seconds. For 3K transactions, APRIORI takes as much as 19.06 seconds; CBVAR and ICBVAR take only 16.97 seconds and 15.805 seconds respectively.

If the number of transactions increases, the computing time too increases. Therefore, for 4K transactions, APRIORI, CBVAR, and ICBVAR take 20.78 seconds, 17.99 seconds and 16.75 seconds .Similarly APRIORI,CBVAR and ICBVAR take 23.48 seconds,19.43 seconds and 18.09 seconds to finish as many as 5K transactions. The algorithms finish 6K transactions in 27.47 seconds, 21.57 seconds, and 19.904 seconds. They operate 7K transactions in 33.52 seconds, 24.59seconds, and 22.29seconds. For 8K transactions, APRIORI 120 takes 42.57 seconds, CBVAR takes 29.01 seconds while ICBVAR consumes only 25.63 seconds respectively. When APRIORI runs 9K transactions, it uses up 55.76 seconds. For the same number of transactions, CBVAR and ICBVAR use up only 35.40 seconds and 30.25 seconds. To run 10K transactions, APRIORI, CBVAR, and ICBVAR consumes 75.80 seconds, 44.95 seconds and 36.906 seconds respectively.

Table 1: Time Complexity of APRIORI, CBVAR, and ICBVAR (1K to 10K)

Synod	Transactions	Computing Time (Time/Sec)		
		APRIORI	CBVAR	ICBVAR
1	1K	17.80	16.00	14.9
2	2K	18.16	16.32	15.198
3	3K	19.06	16.97	15.80592
4	4K	20.78	17.99	16.75428
5	5K	23.48	19.43	18.09462
6	6K	27.47	21.57	19.90408
7	7K	33.52	24.59	22.29257
8	8K	42.57	29.01	25.63645
9	9K	55.76	35.40	30.25102
10	10K	75.80	44.95	36.90624

When APRIORI runs 9K transactions, it uses up 55.76 seconds. For the same number of transactions, CBVAR and ICBVAR use up only 35.40 seconds and 30.25 seconds. To run 10K transactions, APRIORI, CBVAR, and ICBVAR consumes 75.80 seconds, 44.95 seconds and 36.906 seconds respectively.

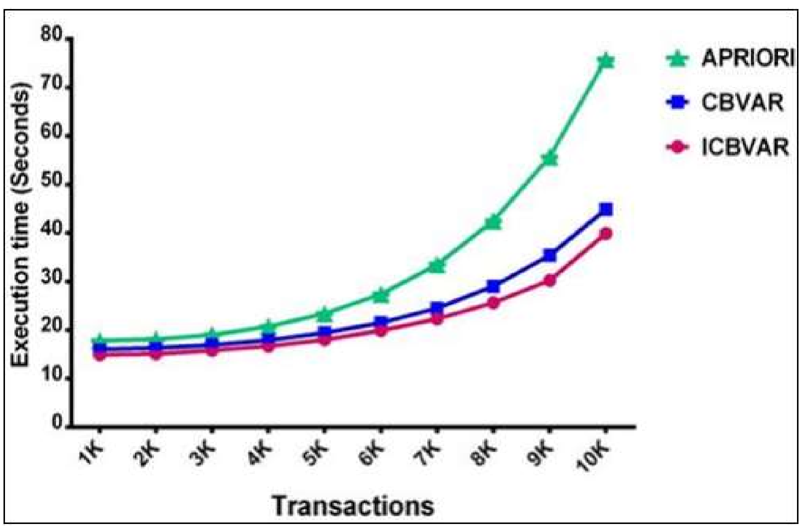

Figure 2: Time Complexity of APRIORI, CBVAR, and ICBVAR (1K to 10K)

Figure 2 gives clear picture about the time consumed by the algorithms such as APRIORI, CBVAR, and ICBVAR from 1K to 10K transactions. The amount of time taken by CBVAR and ICBVAR up to 4K transactions is more or less same. While APRIORI shows an upward spiral from the point of 5K transactions, the other two algorithms do not show much of difference in time consumption for 8K transactions. In the case of APRIORI, it takes less than 20 seconds or so for 4K transactions and from the point of 5K transactions, the algorithm sees far more time consumption than the other algorithms. It consumes as high as 80 seconds for its 10K transactions.

Table 2 shows the memory space being occupied by ten different transactions of the algorithms such as APRIORI, CBVAR, and ICBVAR. For 1K, the memory used by APRIORI, CBVAR and ICBVAR are 52.00KBs, 15.08KBs, and 9.05KBs. If it is 2K, then the memory space taken by the algorithm APRIORI is 59.00KBs while CBVAR consumes memory space of 17.68 KBs for the same number of transactions, and ICBVAR consumes 10.61 KBs respectively. For 3K transactions, APRIORI takes as much as 67.00 KBs of memory, while CBVAR and ICBVAR take only 19.00 KBs and 11.37KBs. If the number of transactions increase, the memory space taken by the algorithms also increases. So, for 4K transactions, APRIORI, CBVAR, and ICBVAR take 76.00 KBs, 22.00 KBs and 13.18 KBs of memory.

Similarly, APRIORI, CBVAR, and ICBVAR take a memory space of 90 KBs, 27.30 KBs, and 16.38 KBs for as many as 5K transactions. The said algorithms finish 6K transactions by occupying a memory space of 100 KBs, 34 KBs, and 20.42 KBs respectively. They operate 7K transactions by occupying a memory space of 111 KBs, 43 KBs, and 25.76 KBs. For 8K transactions, APRIORI takes a memory space of 123 KBs; CBVAR takes a memory space of 55 KBs and ICBVAR 33.86 KBs. When APRIORI runs 9K transactions it uses up a memory space of 136 KBs. For the same number of transactions, CBVAR and ICBVAR use up 65 KBs of memory and 39.12 KBs of memory. To run 10K transactions, APRIORI, CBVAR, and ICBVAR have taken a memory space of 149 KBs, 76.70 KBs, and 46.02 KBs respectively.

Table 2: *Space Complexity of APRIORI, CBVAR, and ICBVAR (1K to 10K)*

Synod	**Transactions**	**MEMORY(KB)**		
		APRIORI	**CBVAR**	**ICBVAR**
1	1K	52.00	15.08	11.6
2	2K	59.00	17.68	13.6
3	3K	67.00	19.00	14.4
4	4K	76.00	22.00	16.2
5	5K	90.00	27.30	21.0
6	6K	100.00	34.00	25.3
7	7K	111.00	43.00	32.1
8	8K	123.00	55.00	40.4
9	9K	136.00	65.00	49.6
10	10K	149.00	76.70	59.0

In Figure 2, the starting point of 1K transactions in APRIORI is from the 50KBs of memory space and the upward spiral of space consumption reaches as high as 150KBs or so for the algorithm in its 10K transactions. For 4K transactions, CBVAR consumes less than 20KBs of memory space while up to the point of 7K transactions in ICBVAR it is less than 25 KBs or so. Algorithm CBVAR takes the memory space of 20 KBs,30 KBs and 40 KBs for 5K, 6K and 7K transactions, while algorithm ICBVAR witnesses the space consumption of 25 KBs, 30 KBs and 40 KBs for 8K,9K and 10K transactions respectively. For 7K-10K transactions, CBVAR takes a memory space within the range of 40 KBs to 70 KBs respectively.

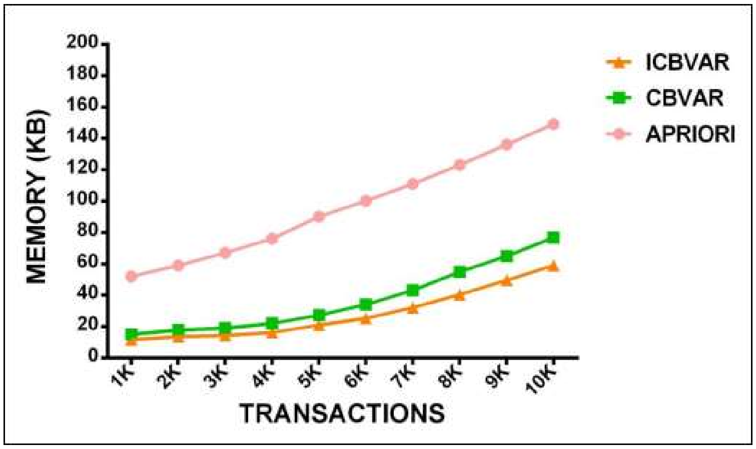

Figure 2: Space Complexity of APRIORI, CBVAR, and ICBVAR (1K to 10K)

This research work evaluates the performance of Apriori, CBVAR, ICBVAR and RMIRROR algorithms in terms of execution time and memory consumption. The algorithms are tested on customer dataset of different size (1K to 10k) to find frequent itemset. The experimental results show that the Apriori algorithm takes more memory space for storing data items but the algorithm CBVAR consumes less memory space since items in transactions are converted into bit vectors. It is shown that the CBVAR is better than Apriori. Based on the running time and memory consumption of the three algorithms Apriori, CBVAR, and ICBVAR, it is observed that the performance of ICBVAR is better than the other two algorithms for the chosen data set. The Performance analysis of the CBVAR, ICBVAR reveals that ICBVAR takes less execution time and memory consumption than Apriori and CBVAR since ICBVAR avoids redundancy and outperforms other algorithms.

IV. CONCLUSIONS

Association rule mining is one of the important concepts in Data Mining. Frequent Itemset Generation is one the step in Association Rule Mining. In order to discover the relationships among the data items in large size of database, the most of the research activities focus on it. The one of the process of Association Rule Mining is Frequent Itemset Generation. The primary goal of this research is to find a novel approach for finding frequent itemset from huge amount of the transactional dataset which outperforms in terms of execution time, memory consumption. the proposed algorithms ICBVAR and its parent algorithms CBVAR and Apriori algorithm are implemented and experimental results made on customer data. The performance of these 140 algorithms are analyzed based on processing time, memory used for storage. The existing traditional algorithm Apriori is very simple method to understand but it produces Candidate itemset every time then the Frequent itemset will be generated. The existing algorithm CBVAR is taken to analyze this research with implementation. It stores items of each transaction instead of storing as its representation, but the representation of each item in transaction will be zero or one. This representation helps in reducing memory storage. The CBVAR and ICBVAR algorithms take lesser memory space and time. Through the designed algorithms, 90 percent of accuracy could be achieved. The algorithms also designed and implemented in the study have proved to be good with respect to time and memory and is compatible with different datasets. The experimental results showed Frequent itemset can be mined efficiently, both in terms of time and space over the previous algorithms.

REFERENCES

[1] Agarwal R, Aggarwal C, Prasad V., “A tree projection algorithm for generation of frequent itemsets”, Journal of Parallel and Distributed Computing AIML Journal, Vol. 61 (3), pp.350-371, 2001.

[2] Akshay palekar, Komal narwekar, Pooja manjeshwar, Yugandhara Rane.,” Data mining using APriori Algorithm,” International journal of engineering trends and technology (IJETT), Vol. 28, pp.190 -192, 2015.

[3] Aswathy Wilson, Jismi Simon Liya Thomas, Soniya Joseph., “Data Mining Techniques for Heart Disease Prediction”, International Journal of Advances in Computer Science and Technology, Vol.3 (2), pp. 113-116,2014.

[4] Baralis E, Cerquitelli T, Chiusano S, Grand A., “P-Mine: parallel itemset mining on large datasets”, IEEE 29th international conference on data engineering workshops (ICDEW), Brisbane, 2013.

[5] Bastide Y, Taouil R, Pasquier N, Stumme G, Lakhal L., “Mining frequent patterns with counting inference”, ACM SIGKDD Explorations Newsletter, Vol.2 (2), pp.66-75, 2000.

[6] Chao L, Zhao-Ping Y., “Improved method of Apriori algorithm based on matrix[j]”, Computer Engineering of China, Vol.23, pp. 68-69, 2006

[7] Chavan K, Kulkarni P, Ghodekar P, Patil S.N., “Frequent itemset mining for Big Data”, IEEE, Conf. on Green Computing and Internet of Things (ICGCIoT), Noida, India, pp.1365-1368, 2015.

[8] Dhanashree Shirke, Deepti Varshney.,A Survey on Techniques of Parallel Mining of Frequent Itemsets Using MapReduce, International Journal of Advance Research in Computer Science and Management Studies, Vol.4 (11), pp.28-32, 2016.

[9] Erwin A. Gopalan R. P, Achuthan N. R., "CTU-Mine: An efficient high utility itemset mining algorithm using the pattern growth approach”, Conf. on Computer and Information Technology, pp. 71-76, 2007.

[10] Gayathri.G., Performance Comparison Of Apriori, Eclat And Fp-Growth Algorithm For Association Rule Learning”, International Journal of Computer Science and Mobile Computing, Vol.6 (2), pp.81-89, 2017.

[11] Han J, Cheng H, Xin D, Yan X., “Frequent pattern mining: current status and future directions”, Data Mining and Knowledge Discovery, Vol. 15(1), pp. 55- 86, 2007.

[12] Ilayaraja M, Meyyappan T., “Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm”, International Conference on Pattern Recognition, Informatics and

Mobile Engineering, pp.294-199, 2013.

[13] Jadhav Kalyani B, Tamhane Manisha S, Surwase Sonali U, Pallavi Patil, "A New Approach For Frequent Itemset Data Mining In Hadoop Environment", 7th International Conference on Recent Trends in Engineering, Science & Management, pp.971-976, 2017.

[14] Karthikeyan T, Ravikumar N., "A survey on association rule mining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, pp. 2278-1021, 2014. Li H. F, Shan M. K, Lee S. Y., "DSM-FI: an efficient algorithm for mining frequent itemsets in data streams", Knowledge and Information Systems, Vol. 17(1), pp.79-97, 2008.

[15] Mall R, Jain P, Pudi V., "PERFICT: Perturbed Frequent Itemset based Classification Technique. In Tools with Artificial Intelligence (ICTAI"), 22nd IEEE International Conference on IEEE, Aras, France, pp.79-86, 2010.

[16] Nandhini S, Mr. Yogesh M, Prabhu, Dr.S.Gunasekaran., Survey on Infrequent temset Mining Using Frequent Pattern Growth, International journal of Scientific and Engineering Research, Vol. 6 (4), pp.23-25, 2015.

[17] Nicolas Pasquier, Yves Bastide, Rafik Taouil, and Lotfi Lakhal.," Discovering Frequent Closed Itemsets for Association Rules", International Conference on Database Theory ICDT, pp.398-416,1999.

[18] Othman yahya, Osman hegazy, Ehab ezat.," An effective implementation of APRIORI algorithm based on hadoop –map reduce model", International journal of reviews in computing, Vol.12, pp.59 -64, 2009.

[19] Prajakta Kulkarni G, Khonde SR., "Performance improvement technique for frequent itemset mining using Map Reduce", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, pp. 11362-11368, 2017.

[20] Qu Z. G, Niu X. X, Deng J, McArdle C, Wang XJ, "Frequent itemset mining over stream data: Overview", IET International Conference on information and communication technologies, pp. 35-40, 2013.

[21] Raghunathan A, Murugesan K., Optimized Frequent Pattern Mining for Classified Data Sets, International Journal of Computer Applications (0975- 8887), Vol.1 (27), pp.20-29, 2010.

[22] Rupindee Kaur, Kamaljit kumar., "An improved web mining technique to fetch web data using APriori and decision tree", International journal of science and research,Vol.3 (6), pp.2094-2098, 2014.

[23] Saravanan uba, Dr. Christopher T.," An Efficient Frequent Pattern Mining Algorithm to Find the Existence of K-Selective Interesting Patterns in Large Dataset Using SIFPMM," International Journal of Applied Engineering Research, Vol.11 (7), pp.5038-5045, 2016.

[24] Suganya G, Paulraj Ananth K.J.," Analysis of Association Rule Mining Algorithms to Generate Frequent Itemset", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6(8), pp.15571-15577, 2017.

[25] Thabet Slimani, Amor Lazzez, "Efficient Analysis of Pattern and Association Rule Mining Approaches", International Journal of Information Technology and Computer Science, Vol.03, pp.70-81, 2014.

[26] Usha D, Rameshkumar K., “A complete survey on application of frequent pattern mining and association rule mining on crime pattern mining”, International Journal of Advances in Computer Science and Technology, Vol. 3, pp. 264 -275, 2014.

[27] Vaarandi R, Podins K., “Network ids alert classification with frequent itemset mining and data clustering”, In Network and Service Management (CNSM), International Conference on IEEE, pp.451- 456, 2010.

[28] Web H. Yu, J, Wang H, Jun L., “An improved Apriori algorithm based on the Boolean Matrix and Hadoop”, Advanced in Control Engineering and Information Science Procedia Engineering, Vol. 15, pp. 1827 -1831, 2011.

[29] Xun Y, Zhang J, Qin X, Zhao X., “FiDoop-DP: data partitioning in frequent itemset mining on Hadoop clusters”, IEEE Transactions on Parallel and Distributed Systems, Vol. 28, pp. 101-114, 2012.

[30] Yang GX., “The research of improved Apriori mining algorithm in bank customer segmentation”, 2nd international conference on optoelectronics engineering and material research, Zhengzhou, china, pp.2244-2249, 2013.

[31] Yubo Yuan, Tingzhu Huang., “Matrix Algorithm for Mining Association Rules”, Springer-Verlag Berlin Heidelberg, pp. 370-379, 2005.

[32] Zawaidah F.H.AL, Jbara Y.H, Marwan AL., “An improved algorithm for mining association rules in large data bases”, Vol.1 (7), pp. 311-316, 2011.

[33] Zhao Q, Bhowmick S. S., “Association rule mining: A survey”, Technical Report, CAIS, Nanyang Technological University, Singapore, pp.1-9, 2003.

[34] Zhi-Hong Deng, Sheng-Long LU., “Fast mining Frequent itemset using Node sets”, Expert Systems with application, Vol. 41, pp.4505-4512, 2014.