Prioritizing Effectiveness of Algorithms of Association Rule Mining

Nazish Ghafoor, Mansoor Ahmad

Abstract—From the last decade, clouds have become the most popular platform for data storage. In the current age, people and machines are engaged in transferring data on clouds. The devices like cameras, computers, mobile, and CCTV are being used to transmit data collectively on the clouds in each second. Such a huge collection of data on clouds is known as Big data. Data mining is a process of extracting useful information from a set of huge data. The different techniques like Association Rule Mining, Classification, and Clustering are some of the well-known techniques, which can be used for data mining purposes. Association Rule Mining is a process of mining associations and correlations among the items in a large data set. Some traditional algorithms like Apriori, FP-growth, and Eclat are being used for association rule mining, but no one provides an optimal solution. In our study, we ascertained the working algorithms, evaluated their performance, and finally ranked them based on their efficiency. We adopted a quantitative approach in our research. We framed queries, pinpointed pertinent work, gauged quality, summarized the evidence, and finally interpreted our findings.

Index Terms— Association Rule Mining, ARM algorithms, Big data, Data Mining, Systematic Literature Review.

1 Introduction

Association Rule Mining is a technique of data mining that pursues some interesting relationships among data items in a database. The information extracting technique that is mostly used is ARM (Association Rule Mining). It draws out the secreted pattern that links different items. It helps one to draws out data items that have a strong association. Considering basket analysis as an example, how the purchasing behavior of the customers one can estimate? One wants to keep all items together that a customer purchase mostly. One wants to group the items that are purchased together because items have somehow an association with one another [1].

Association rule mining is a method of taking out correlations and associations between items in a huge data set [2]. In association rule mining, one selects the item sets that have extra or identical support in comparison with the specified least support. Considering P also Y are items in any database, we need to know the probabilities of Q with the occurrence of P. One might state it as *P -> Q*. It satisfies two properties Confidence and Support. Support is the total count of occurrences of a particular item in total transactions. Confidence is the transaction percentage, comprising the second item; given that, it has already the first item. It is considered an example of 100 transactions, in which P arises 85 times, P with Q occurs 35 times, support is (30/100) 35% and confidence is (35/85), which is 41%.

Rule P-> Q
Support = frequency (P, Q) / N
Confidence = frequency (P, Q) / frequency (P)
N is the number of transactions in total.

Electronic Computers, mobiles, cameras, and various additional devices are combined forwarding data each second to the internet is called big data. The data available on the internet is in both ways, structured and unstructured. This data is originating from various sources from machines as well as from humans. The volume of data on the internet is being enlarged after every second is passed, from electronic devices that are working standalone and without human effort and are forwarding data to the internet. This type of data with many structures could not be organized through conventional database management systems. This is called Big Data [3].

Numerous specialists have utilized Affiliation Rule Mining in their corrections and studies yet this zone has still a hole to cover and work on it. Also, the overall explanation of this hole is the combined measure of information and its inclination [4]. This study covered the systematic review of all existing studies related to rule mining, identified the most relevant data, evaluated their performance, and ranked them according to their efficiency.

The purpose of this research was to review the working of the existing algorithm on data mining systematically. For this purpose, specific keywords were formatted from the RQ1 and RQ2 with the help of binary operators. With the support of these keywords, relevant research publications were explored from scientific databases, e.i. IEEE, ACM, Science Direct, Springer, and Google Scholar. After getting the relevant study, the quality of the study was assessed.

2 Basic Concepts

In an example of a supermarket having a large set of items for sale, various questions arise [5]. How to handle items and how to aware of the items that must be placed in that superstore? How to increase sales, and how to increase profit, etc.? Past sales data is used for analysis purposes to answer the above-given questions. Some organization used to store such data in huge amount and converted into databases. After maintaining such a huge database, the organization uses that data for extracting patterns or some interesting information. Discovering such patterns enables organizations to make important decisions for the business. Such decisions are applicable in Cross-Marketing, Attached-Mailing, Catalog-Design, Add-on Sales, Store-Layout, etc.

For an organization, this is crucial to discover useful patterns from the database. For extracting interesting patterns from the

• *Nazish Ghafoor is working as a Visiting Lecturer in the Department of Computer Science and IT, University of Sargodha, Pakistan. E-mail: nazshghafoor58@gmail.com.*
• *Mansoor Ahmad is working as a Lecturer in the LIMS (Lahore Institute of Modern Sciences) College, Sargodha Campus, Pakistan. E-mail:* mansoorahmaducp@gmail.com

database, association rule mining plays a vital role [6]. For working in this area in the future [7], it is more interesting to work with networks or with websites. The motives for researching association rule mining are due to the future necessities and importance of data mining [8]. One can use other algorithms along with the Apriori algorithm to syndicate the characteristics having similarities. One can disclose some solutions to improve the current techniques of association rule mining [9].

Association Rule Mining research has some gap for the researcher to work on it. The author discussed some research gaps for further study in this area. How to generate a large number of rules for knowledge discovery? How to optimize the current algorithms working for association rule mining [10]?

3 Research Questions and Review Process

The association rule mining is a process to find out the support and resistance of items and extract the items that are more relevant to each other [11]. Rule mining is a hot topic today and has different tools and techniques for data mining. All these techniques have some merits and some demerits concerning the performance and efficiency of these methods. In this research, the researcher conducted an SLR(Systematic Literature Review) by applying a quantitative approach. The main purpose of conducting the research was to evaluate existed study of association rule mining systematically and logically. To conduct the review process, we composed two research questions in the following section. In the proceeding sections, we defined the systematic literature review.

3.1 Explore Questions

The study will answer the below questions:

1. What are the existing well-known techniques that are currently working for Association Rule Mining?
2. Which technique or algorithm is more efficient as compared to the selected techniques for association rule mining based on the input size, speed, and accuracy?

3.2 Data Sources and Search Strategy

If it was needed to conduct a basic & unbiased investigation, it was required to investigate a huge range of digital libraries. There was an expansive variety of advanced digital libraries accessible on the web to examine the research. The digital library enlisted here encompasses a very large variety of computer and information technology domains [12].

Digital libraries:
Science-Direct [1]
Springer [2]
ACM-Digital-library [3]
IEEE-Xplore [4]

Search Engines:
Google-Scholar [5]
For RQ1: association rule mining, existing, algorithm
For RQ2: association rule mining, technique, beneficial

We also used synonyms of the selected terms in the search strings. The search strings have been composed using the logical expressions between the search terms. Each search string has been carefully formulated according to the constraints of specific databases.

The list of instructions considered when examining the primary data for selection are given as under:

1. For getting desired results, the primary data was selected from the rule mining domain.
2. The study selected from a paper published in a journal, conference or book, was suitable as an evidence-based study.
3. Selected English language for writing the thesis, as most of the people in the world, could understand English.
4. To consider the current trend, only selected the primary study published after or in 2010.
5. Selected only those articles that are free or have open access with the help of HEC.
6. In ordered to access the required articles or studies, research strings were composed by using standards given in the research strategy. These search strings were executed using digital databases.

3.3 Data Selection Process

It was a very simple process to select data from the digital databases. The research papers were downloaded from the digital library. After downloading, the title of the research paper was read by the researcher and selected or rejected based on relevancy. The abstract of research papers was also reviewed after selecting based on the title. The biasness and relevancy of the paper with research questions were verified and approved. Sometimes researchers failed to verify the biasness and relevancy with the problem. In such a case, the complete paper was reviewed thoroughly for selection or rejection. The duplicate paper was also rejected. Our study is an evidence-based study.

3.4 The procedure for Quality Assurance

During the inclusion & exclusion measures of the primary study, it was hard to judge the quality of the selected researches [13]. The main purpose of our study was to point out all differences between the papers selected in the primary study.

Outcomes were tested in addition to the quality selection process to reduce the chances of biasness. Internal and external validity was used to authenticate the data. The quality evaluation, by conduction a systematic literature review, is displayed in Appendix A.

3.5 Plotting Extracted Data

In the following text, we are going to discuss that how data was collected from primary researches. The general information about research question 1 and research question 2 is given in Appendix B. The primary study was evaluated by the criteria of quality assessment. The information that existed in Appendix B has listed after the evaluation of the primary study and then this study was ranked based on this evaluation. Appendix B contains data about the primary study like keywords, the title of study, author's name, conference/journal, publication year, organization type, and the review date.

Appendix C is containing data collected for research question 1 and research question 2. This data has been used in this study. Appendix C contains data related to these studies e.g. database, the country in which the study was done, the year in which the study was conducted, and even research methodology for the mentioned study in the appendix.

All the primary studies extracted in response to research ques-

1 http://www.sciencedirect.com/
2 http://www.springer.com/
3 http://dl.acm.org/
4 http://ieeexplore.ieee.org/
5 https://scholar.google.pk

tions 1 & 2 are listed in Appendix D. Data, extracted from primary studies, was examined and an assessment was accomplished to prioritize the performance of different techniques or algorithms.

3.6 Data Composition Plan

This part of the review gives how the results were custom fitted and from the past portions. As insights were mined from various looks into, each of them has its method so the outcomes were shown in table format.

4 Results and Discussion

4.1 Primary Studies Selection

The search string was executed in the search engine for collecting data from the digital databases. The searched research papers were relevant to the search string that was defined in the protocol. This task of searching the relevant data was completed in a couple of weeks. Fig 1. given below shows the sequence of steps to be followed for data collection. It produced 90 papers for RQ1 and 110 articles for RQ2 from different digital databases.

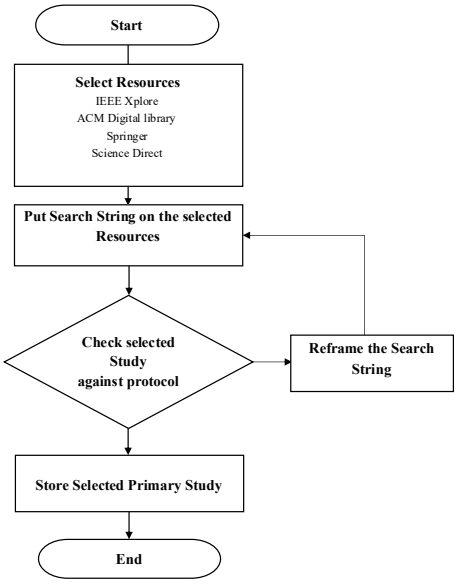

Fig 1. Steps in Search Strategy

In section 3.5, we mentioned a mechanism for selecting data, these criteria were applied to select articles from databases. The title and abstract of each retrieved paper were reviewed very carefully. After examining the papers, 110 research papers were selected as a primary study (Filter 1).

After filtering the papers in Filter 1, Filter 2 was executed. In this step, the full-length article was studied to find out the relevance. In this phase, a large number of studies were rejected due to irrelevancy. Finally, after Filter 2, 62 papers were selected. Details about the quantity of retrieved paper that is finally selected have been given in the following table.

TABLE 1
NUMBER OF RETRIEVED PAPERS FROM THE SELECTED SOURCES

Sr. No.	Data Source	Initially Scanned		After Filter 1		After Filter 2		Final Selection	
		RQ	RQ	RQ	RQ	RQ	RQ	RQ	RQ
		1	2	1	2	1	2	1	2
1	IEEE Xplore	16	27	17	21	7	11	5	5
2	Google Scholar	33	37	19	26	15	23	5	8
3	ACM Digital Lib.	21	27	0	7	0	0	0	0
4	Science Direct	17	15	4	9	3	3	1	1
5	Springer	3	4	3	4	0	0	0	0
	Total	**90**	**110**	**43**	**67**	**25**	**37**	**11**	**14**

Using search string 1, eleven research papers were selected finally. And for search string 2, there were fourteen papers, selected for further study. This data has been given in Table 1. The list of papers selected for Research Question 1 has been given in Table 2 and papers related to Research Question 2 have been given in Table 3. These studies were further filtered concerning the Quality Assessment. Each research paper was assigned Quality Assessment Score. The scores of the Quality Assessment have been summarized in ***Error! Reference source not found.***.

TABLE 2
IN RESPONSE TO RQ2 FOR PRIMARY STUDY SELECTED PAPERS

ID	Research Question Results	Reference Research Paper
[1]	RQ1	Wu et al., (2008)
[2]	RQ1	Kotsiantis et al., (2006)
[3]	RQ1	Lucchese et al., (2006)
[4]	RQ1	Lin et al., (2010)
[5]	RQ1	Hall et al., (2009)
[6]	RQ1	Chaure et al., (2016)
[7]	RQ1	Anaissi et al., (2015)
[8]	RQ1	Bakar et al., (2015)
[9]	RQ1	Cong et al., (2002)
[10]	RQ1	Ghani et al., (2002)
[11]	RQ1	Hipp et al., (2000)

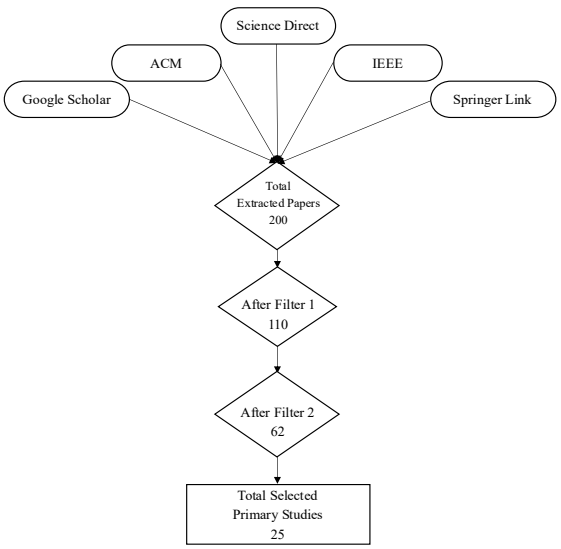

Fig 2. Selection of Primary Studies

TABLE 3
IN RESPONSE TO RQ2 FOR PRIMARY STUDY SELECTED PAPERS

ID	Research Question Results	Reference Research Paper
[1]	RQ2	Ayubi et al., (2009)
[2]	RQ2	Mazid et al., (2009)
[3]	RQ2	Kuo et al., (2011)
[4]	RQ2	Chung & Mangamur (2004)
[5]	RQ2	Singh & Sethi (2015, December)
[6]	RQ2	Li et al., (2005, August)
[7]	RQ2	Chadokar et al., (2013, July)
[8]	RQ2	Comito (2015, October)
[9]	RQ2	Min et al., (2010, December)

[10]	RQ2	Peng et al., (2009, March)
[11]	RQ2	Aher et al., (2012)
[12]	RQ2	Wei et al., (2014, May)
[13]	RQ2	Gu et al., (2015, December)
[14]	RQ2	Bharathi et al., (2016, January)

The above-given tables represent the list of primary term papers that were selected for accomplishing this study. In the mentioned tables, the studies have been distributed based on Research Questions.

4.2 Extracted Data Presentation

Data extracted from primary studies have been arranged in Appendix D. The general information about papers related to the RQ1 has been stored in Table 10 of Appendix B. The information extracted from research papers that were related to the RQ2 has been organized in Table 11 of Appendix B. Whereas details about data mining extracted from the studies have been given in ***Error! Reference source not found.***. Appendix C constitutes the information about data mining extracted from the primary studies related to the RQ1 (Table 12) while the data related to the RQ2 has been organized in Table 13 that were collected from the primary studies.

4.3 Latest Algorithms for Association Rule Mining

The study conducted by Bharathi & Krishnakumari [14], provides the analysis of currently available algorithms for association rule mining. This research provides a list of algorithms that are currently working for association rule mining. The following is a list of algorithms discussed in this study:

TABLE 4
LIST OF LATEST ARM ALGORITHMS

Sr.	Name of Algorithm
[1]	AIS
[2]	SETM
[3]	Apriori
[4]	AprioriTID
[5]	Apriori Hybrid
[6]	Eclat
[7]	Recursive Elimination
[8]	TERTIUS
[9]	FP-Growth
[10]	Dyn- FP Growth
[11]	PSO

4.4 Comparative Analysis of ARM Algorithms:

The table given below provides the scores of each variable related to the algorithm for association rule mining. The details of these variables have been given below with their meanings [15]. Each variable of the algorithm has been divided into four scores ranging from 1 to 4, 1 is for least whereas 4 is for the highest score. The four variables are the input data size, speed in phase 1 for database pass and speed in phase 2 for the second pass into the database, and algorithm's accuracy in results.

TABLE 5
THE SCORE FOR EACH VARIABLE

Input Data Size		Speed Phase 1		Speed Phase 2		Algo-Accuracy	
Limited	1	Slow	1	Slow	1	Very Less	1
Less	2	Average	2	Average	2	Less	2
Large	3	Fast	3	Fast	3	Average	3
Very Large	4	Very Fast	4	Very Fast	4	High	4

Efficiency Results were calculated by taking the average of four different value input data sizes, speed phase 1, speed phase 2, and accuracy. The equation for this is given below:

ER = (IDS+SP1+SP2+AC)/4
ER = Efficiency Results
IDS = Input Data Size
SP1 = Speed Phase 1
SP2 = Speed Phase 2
AC = Accuracy

TABLE 6
SCORE TABLE FOR ASSOCIATION RULE MINING ALGORITHMS

Sr.	Name of Algorithms	Input Data Size	Speed Phase-1	Speed Phase-2	Accuracy	Efficiency Results
[1]	AIS	2	1	1	1	0.3125
[2]	SETM	2	1	1	2	0.3750
[3]	Apriori	1	3	1	2	0.4375
[4]	AprioriTID	3	1	3	3	0.6250
[5]	Apriori Hybrid	4	3	3	4	0.8750
[6]	Eclat	3	3	2	3	0.6875
[7]	Recursive Elimination	3	3	1	3	0.6250
[8]	TERTIUS	1	1	2	3	0.4375
[9]	FP-Growth	4	3	3	4	0.8750
[10]	Dyn-FP Growth	4	3	3	4	0.8750
[11]	PSO	4	4	4	4	1.0000

These algorithms were evaluated based on four variables that have been given in table 5. Then, we calculated the average of data size, speed phase 1 and speed phase 2, and accuracy, to convert the result into a standard format. The "Efficiency Results" field shows the overall efficiency of each Algorithm in the list.

4.5 Graphical Representation of Results

The following is a graphical representation of algorithms' performance according to the four variables discussed in the previous section. Each variable has been represented with a color bar in the Bar Chart, fig 3. Input data size is embodied by blue color, Speed in phase 1 is epitomized with orange color, speed in phase 2 is represented with gray color, and accuracy of the algorithm is bared by yellow color.

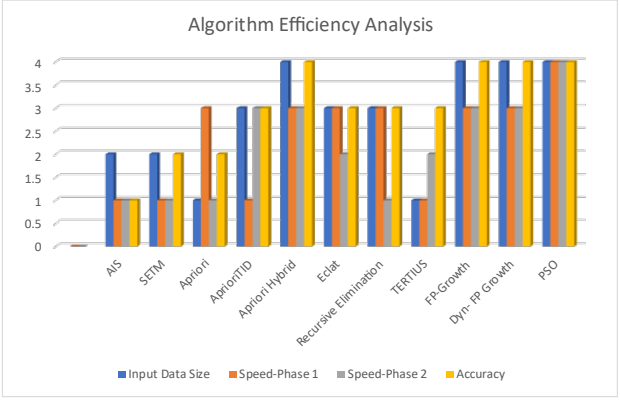

Fig 3. Algorithms Performance w.r.t Data, speed and Accuracy

Each algorithm is alienated into four vertical bars in the above-placed chart. Each bar epidictic a variable. An algorithm is considered to be good if all or most of the bars are at a great-

er point. The range of the bar is from 0 to 4 if the bar is near zero it has fewer points and if it is near to 4 it has higher points. If we see a graph, the AIS Algorithm is not good in all cases. But if we take a glance at PSO Algorithm, it is depicting the best results from the whole list and it is considered to be good. However, some other algorithms are on average [16]. Apriori Hybrid, FP-Growth, and Dyn-FP Growth are equal with each other in their performance.

If we glimpse each variable separately, input data size and accuracy of results both are same for Apriori Hybrid, FP-Growth, Dyn-FP Growth, and PSO. But if we see the speed in both phases, PSO takes advantage [17].

The overall performance of each algorithm is displayed in the following bar chart, fig 4. The graphical representation shows the performance of each algorithm clearly and unambiguously. The PSO Algorithm is at the top and AIS Algorithm is

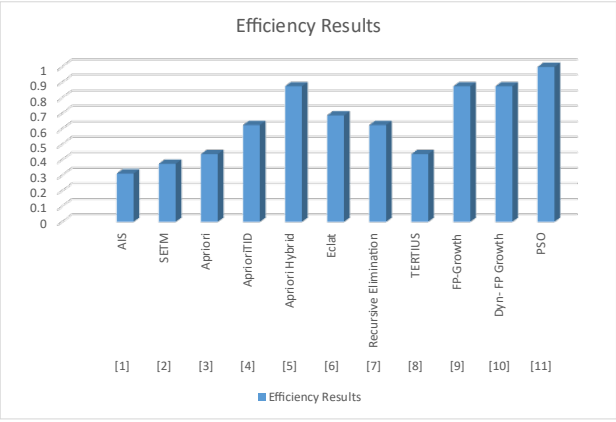

Fig 4. Algorithm's Efficiency Results

at the lowest position in its performance.

We can easily prioritize the algorithm in Ascending order according to their performance. The following is a list of algorithms prioritized in a way that the first one is the best at enactment and the last one is a little good at play-acting.

TABLE 7
ALGORITHMS PRIORITIZATION

PSO	*******	Recursive Elimination	****
Apriori Hybrid	******	Apriori	***
FP-Growth	******	TERTIUS	***
Dyn-FP Growth	******	SETM	**
Eclat	*****	AIS	*
AprioriTID	****		

AIS Algorithm works for small kinds of problems whereas the SETM is not used frequently. Apriori Algorithm is considered to be better for closed itemsets and AprioriTID is used for small problems. Apriori Hybrid is best in the case of closed itemsets and Eclat is used with free item sets. TERTIUS is not used mostly. FP growth is also used for large applications. Dyn-Fp Growth is an algorithm that needs candidate set generation. PSO Algorithm is used for free & closed itemsets and applied for a large application. If we consider the values calculated as the efficiency of each algorithm, PSO is better than all other algorithms. Next, AIS is not good enough in most cases when comparing with other above-mentioned algorithms.

5 VALIDITY OF LITERATURE REVIEW

Our review study is conducted in a systematically designed process. Whereas during the construction of search strings and selection of primary studies some assumptions have been considered, depending upon the common sharing of concepts among the authors. The documented procedure of the review process guaranteed the validity of the search, choice, and evaluation of articles. There may exist some threats to soundness in the subjective evaluation of the articles. However, these threats are weakened by reasoning the analysis process from multiple authors. Another threat to the validity of the review process is that it is based upon the results provided by the different search engines. The search is systematic and can be repetitive but each time the produced results may not be the same, due to the expanding nature of digital databases [18]. Here we may be overlooked few relevant articles in the results. However, we involved multiple authors in the selection process, thus reducing this possibility.

6 CONCLUSIONS

A systematic literature review has been conducted to complete this research work. The study of association rule mining was completely explored and reviewed systematically to gain the results. This research is evidence-based, so it presents all the pieces of evidence related to association rule mining. In the study, a research protocol was formulated for performing the literature review. The protocol was tested before executing the review to obtain the relevant data. The data was validated through assessment criteria before selecting the data for the primary study. Each primary study was extracted from the quality sources of digital databases.

Finally, the systematic literature review gave a list of algorithms currently working for association rule mining. These algorithms are well known and considered to be more common and efficient. This study provides us with eleven different algorithms to check their performance. The overall efficiency of the algorithm was calculated based on three major factors. These factors are the size of input data on which the algorithm is executed, the speed of completing phase one and the speed of completion of phase two, and finally the accuracy they provide in the results. The performance of the algorithms was evaluated concerning these three factors. Data size, speed, and accuracy of each algorithm were compared with the other algorithm to score them. Finally, the researcher came to the point that PSO was the best algorithm having a large amount of data. It was speedy in both phases and provided accurate results. This algorithm can be used for large applications with closed & free item sets.

Apriori Algorithm performed well in the case of closed itemsets when compared with other stated algorithms, but Eclat was better than all other mentioned algorithms in the case of free item sets. Recursive Elimination was better than the Apriori algorithm but less efficient than the Eclat. FP-growth's performance was far better than Apriori and Eclat in all the cases due to two passes in the database. Finally, the researcher concluded that the PSO algorithm was better than all the revealed algorithms in the case of a large application, closed item sets, and free item sets.

The outcomes of this research work will provide sound contribution in the understanding of association rule mining and working of algorithms related to them. There is a need for more improvements in the current algorithms to achieve optimal results. The optimal solution can be achieved in future work and distributed data mining (DDM) algorithms on a cluster can be concentrated. The implementation of algorithms can be done using different data

structures with good programming logic to improve efficiency.

REFERENCES

[1] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering a tertiary study," Information and Software Technology, Vol. 52, 2010, pp. 792-805.

[2] Bharathi, T., & Krishnakumari, P. (2016, January). A comparative analysis on efficiency of contemporary association rule mining algorithm. In Advanced Computing and Communication Systems (ICACCS), 2016 3rd International Conference on (Vol. 1, pp. 1-9). IEEE.

[3] Bhatnagar, V., Gupta, A., & Kumar, N. (n.d.). Algorithms for association rule mining. Encyclopedia of Artificial Intelligence, 76-84. https://doi.org/10.4018/978-1-59904-849-9.ch012

[4] Chi, X., & Fang, Z. W. (2011, May). Review of association rule mining algorithm in data mining. In Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on (pp. 512-516). IEEE.

[5] D. Evans, "Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions," Journal of Clinical Nursing, Vol. 12, 2003, pp. 77-84.

[6] Huria, S., & Singh, J. (2015, April). Implementation of Dynamic Association Rule Mining Using Back Navigation Approach. In Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on (pp. 1048-1050). IEEE.

[7] Kadir, A. S. A., Bakar, A. A., & Hamdan, A. R. (2011, November). Frequent absence and presence itemset for negative association rule mining. In Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on (pp. 965-970). IEEE.

[8] Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(2004), 1-26.

[9] Ölmezogullari, E., & Ari, I. (2013, June). Online association rule mining over fast data. In 2013 IEEE International Congress on Big Data (pp. 110-117). IEEE.

[10] Ozkural, E., Ucar, B., & Aykanat, C. (2011). Parallel frequent itemset mining with selective item replication. IEEE Transactions on Parallel and Distributed Systems, 22(10), 1632-1640.

[11] Prutzkow, A. (2018). Algorithms and data structures for association rule mining and its complexity analysis. https://doi.org/10.15405/epsbs.2018.11.02.62

[12] S. Vijayarani and S. Sharmila, "Comparative analysis of association rule mining algorithms," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-6, DOI: 10.1109/INVENTIVE.2016.7830203.

[13] SHARMA, K., & VASHISHT, S. (2013). An efficient approach using rule induction and association rule mining algorithms in data mining. Graduate Research in Engineering and Technology, 14-18. https://doi.org/10.47893/gret.2013.1021

[14] Sharma, S., & Toshniwal, D. (2015, December). Parallelization of association rule mining: Survey. In Computing, Communication and Security (ICCCS), 2015 International Conference on (pp. 1-6). IEEE.

[15] T. Rahman, M. M. J. Kabir, and M. Kabir, "Performance Evaluation of Fuzzy Association Rule Mining Algorithms," 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 2019, pp. 1-4, DOI: 10.1109/EICT48899.2019.9068771.

[16] W. A. K. Salman and S. B. Sadkhan, "Proposed Association Rules Mining Algorithm for Sensors Data Streams," 2021 1st Babylon International Conference on Information Technology and Science (BICITS), 2021, pp. 76-81, DOI: 10.1109/BICITS51482.2021.9509893.

[17] X. Ren, "Application of Apriori Association Rules Algorithm to Data Mining Technology to Mining E-commerce Potential Customers," 2021 International Wireless Communications and Mobile Computing (IWCMC), 2021, pp. 1193-1196, DOI: 10.1109/IWCMC51323.2021.9498773.

[18] Yang, T., Qian, K., Lo, D. C. T., Xie, Y., Shi, Y., & Tao, L. (2016, April). Improve the Prediction Accuracy of Naïve Bayes Classifier with Association Rule Mining. In Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on (pp. 129-133). IEEE.

Appendix A

Table 8: Quality Assessment Report of Primary Studies for RQ-1

S/n	Quality Assessment Question	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
1	**Research Problem**											
	Does the study explicitly state its problem?	1	1	1	1	1	1	1	1	1	1	1
	Does it recommend future work?	0	0	0	1	0	0	1	1	0	1	0
2	**Literature Reference**											
	Do the study presents a satisfactory literature review?	0	0	1	1	0	1	1	1	1	1	1
3	**Research Methodology**											
	Is the research comprehendingon a scientific research methodology?	1	1	1	1	1	1	1	1	1	1	1
	Is the evaluation technique stated?	0	0	0	1	1	1	1	1	1	0	0
	Is any statistical technique applied?	0	0	0	0	0	0	0	1	1	1	0
4	**Outcomes**											
	Has a handful of evidence provided after the analysis?	1	1	1	1	1	1	1	1	1	1	1
	Does the extracted evidence justify the conclusion?	1	1	1	1	1	1	1	1	1	1	1
	Total Score	**4**	**4**	**5**	**7**	**5**	**7**	**7**	**8**	**7**	**7**	**5**

Table 9: Quality Assessment Report of Primary Studies for RQ-2

S/n	Quality Assessment Question	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]
1	**Research Problem**														
	Does the study explicitly state its problem?	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Does it recommend future work?	0	0	0	0	0	0	0	0	0	1	1	0	0	1
2	**Literature Reference**														
	Do the study presents a satisfactory literature review?	1	0	1	0	1	0	1	0	1	1	1	1	1	1
3	**Research Methodology**														
	Is the research comprehending on ascientific research methodology?	1	1	1	1	1	1	1	1	1	0	1	1	1	1

	Is evaluation technique is stated?	0	0	1	0	0	0	1	0	1	0	1	1	1	1
	Is any statistical technique is applied?	0	0	1	0	0	0	1	0	0	0	0	0	1	0
4	**Outcomes**														
	Has a handful evidence provided after the analysis?	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Does the extracted evidence justify the conclusion?	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Total Score	**5**	**4**	**7**	**4**	**5**	**4**	**7**	**4**	**6**	**5**	**7**	**6**	**7**	**7**

Appendix B

Table 10: General Information from Primary Studies for RQ-1

Study ID	Study Title	Authors	Journal/ Conference	Publication Year	Keywords	Review Date	Organization Type
[1]	Wu et al., (2008)	Wu, Xindong and Kumar, Vipin and Quinlan, J Ross and Ghosh, Joydeep and Yang, Qiang and Motoda, Hiroshi and McLachlan, Geoffrey J and Ng, Angus and Liu, Bing and Philip, S Yu and others	Journal	2008	---	20-08-2020	Academic
[2]	Kotsiantis et al., (2006)	Kotsiantis, Sotiris and Kanellopoulos, Dimitris	Journal	2006	---	23-08-2020	Academic
[3]	Lucchese et al., (2006)	Lucchese, Claudio and Orlando, Salvatore and Perego, Raffaele	Journal	2006	*Data mining, association rules, frequent itemsets, condensed representations, closed itemsets, high-performance algorithms*	24-08-2020	Academic
[4]	Lin et al., (2010)	Lin, Xiaoyong and Zhu, Qunxiong	Journal	2010	*data mining, association rules mining, frequent patterns mining, pattern growth, share-inherit*	27-08-2020	Academic
[5]	Hall et al., (2009)	Hall, Mark and Frank, Eibe and Holmes, Geoffrey and Pfahringer, Bernhard and Reutemann, Peter and Witten, Ian H	Journal	2009	---	03-09-2020	Academic
[6]	Chaure et al., (2016)	Chaure, Tushar M and Singh, Kavita R	Journal	2016	*Frequent itemsets, association rule*	11-09-2020	Academic

					mining, mining algorithms		
[7]	Anaissi et al., (2015)	Anaissi, Ali and Goyal, Madhu	Journal	2015	*SVM, data mining, machine learning, Apriori algorithm, association rules*	19-09-2020	Academic
[8]	Bakar et al., (2015)	Bakar, Wan Aezwani Bt Wan Abu and Abdullah, Zailani B and Saman,Md Yazid B Md and Jalil, Masita Masila Bt Abd and Man, Mustafa Band Herawan, Tutut	Journal	2015	*Association Rule Mining, relationaldatabase, MySQL, Frequent itemset, Eclat algorithm, association rule*	24-10-2020	Academic
[9]	Cong et al., (2002)	Cong, Gao and Liu, Bing	Conference	2002	---	05-11-2020	Academic
[10]	Ghani et al., (2002)	Ghani, Rayid and Fano, Andrew E	Conference	2002	---	13-11-2020	Academic
[11]	Hipp et al., (2000)	Hipp, Jochen, Ulrich and Nakhaeizadeh, Gholamreza	Journal	2000	---	17-11-2020	Academic

Table 11: General Information from Primary Studies for RQ-2

Study ID	**Study Title**	**Authors**	**Journal/ Conference**	**Public Ation Year**	**Keywords**	**Review Date**	**Organization type**
[1]	Ayubi et al., (2009)	Siyamand Ayubi, Maybin K. Muyeba, Ahmad Baraani, John Keane		2009	---	03-12-2020	Academic
[2]	Mazid et al., (2009)	Mazid, Mohammed M and Ali, ABMShawkat and Tickle, Kevin S	Conference	2009	*Association Rule Mining, Classification, Apriori, Partial Decision Tree (PART), DARPA (Defense Advanced Research Projects Agency)*	05-12-2020	Academic
[3]	Kuo et al., (2011)	Kuo, Ren Jie and Chao, Chie Min and Chiu, YT		2011	*Association rule mining, Particleswarm optimization algorithm*	09-12-2020	Academic
[4]	Chung & Mangamur (2004)	Chung, Soon Myoung and Mangamuri,Murali	Conference	2004	*data mining, association rules,parallel database the system, performance analysis*	13-12-2020	Academic
[5]	Singh & Sethi (2015,December)	Singh, Tarinder and Sethi, Manoj	Conference	2015	*Frequent itemsets, minimum support,bound, candidate itemset*	14-12-2020	Academic
[6]	Li et al., (2005, August)	Li, Zhi-Chao and He, Pi-Lian and Lei, Ming	Conference	2005	*Association rule, Apriori-Tidalgorithm, HEA algorithm, data mining, KDD*	19-12-2020	Academic
[7]	Chadokar et al., (2013, July)	Chadokar, Surendra Kumar and Singh, Divakar and Singh, Anju	Conference	2013	*Genetic Algorithm, Apriori Algorithm, frequent itemsets, Association rule mining*	22-12-2020	Academic
[8]	Comito (2015, October)	Comito, Carmela and Talia, Domenico	Conference	2015	---	24-12-2020	Academic
[9]	Min et al., (2010,	Min, Li and Chunyan, Wang and Yuguang, Yan	Conference	2010	*Apriori Algorithm, ModeL-FP model,FP-growth method, Supporting System of Medical Decision*	27-12-2020	Academic

	Decem-ber)						
[10]	Peng et al., (2009, March)	Peng, Sihua and Liu, Xiaoping and Yu,Jiyang and Wan, Zhizhen and Peng, Xiaoning		2009	---	04-01-2021	Academic

Appendix C

Table 12: Data Extracted for RQ-1

Study ID	Description	Study Year	Country of study	Database	Quality Assessment
[1]	Top 10 algorithms about data mining	2008	USA	Springer	4
[2]	There is an overview of an existing association rule mining techniques	2006	Greece	Google Scholar	4
[3]	This paper presents a new scalable algorithmfor discovering closed frequent itemsets.	2006	Italy	IEEE	5
[4]	In this study, a new data structure is proposed, Share-struct, which is derived but different from FPtree.	2010	China	IEEE	7
[5]	This paper provides an introduction to the WEKA workbench and reviews the history of the project.	2009	USA	ACM	5
[6]	In this paper, a survey of Hadoop is being presented in detail, which helps in storing data and parallel processing in distributed environment.	2016	India	IEEE	7
[7]	This paper delivers a strategy for the implementation of a systematic analysis of a framework built on the established principles used in data mining and machine learn-ing.	2015	Australia	IEEE	7
[8]	In this paper, MySQL is implemented as an association rule mining database engine in testing benchmark dense datasets which available from Frequent Itemset Mining (FI-MI) online repository.	2015	Malaysia	IEEE	8
[9]	The proposed technique is implemented in the contexts of two existing algorithms, FP-tree and Tree Projection.	2002	Singapore	IEEE	7
[10]	In this paper, a case study of a system that is trained to extract semantic features for apparel products and populate a knowledge base with these products and features are presented.	2002	Chicago	IEEE	7
[11]	In this paper, the fundamentals of association rule mining is explained.	2000	Germany	ACM	5

Table 13: Data Extracted for RQ-2

Study ID	Description	Year of Study	Country of study	Database	Quality Assess-ment
[1]	An algorithm to mine general association rules from tabular data.	2009	United King-dom	Science Direct	5
[2]	This research aims to compare the performance between the rule-based classifica-tion and association rule mining algorithms based on their rule-based classification performance and computational complexity.	2009	Australia	IEEE	4

[3]	This study intends to propose a novel algorithm for association rule mining to improve computational efficiency as well as to automatically determine suitable threshold values.	2009	Taiwan	Science Direct	7
[4]	In this paper, a new algorithm is proposed for mining association rules from relations.	2004	USA	IEEE	4
[5]	This paper proposes a new approach called Sandwich-Apriori which is a combination of both Apriori and Reverse-Apriori.	2015	India	IEEE	5
[6]	The experiments show that the new algorithm is more effective in decreasing data size and execution times than the AprioriTid algorithm.	2005	China	IEEE	4
[7]	Optimizing network traffic by generating association rules using a Hybrid Apriori-Genetic algorithm.	2013	India	IEEE	7
[8]	This paper presents an experimental study of the energy consumption behavior of representative data mining algorithms running on mobile devices.	2015	Italy	IEEE	4
[9]	The research of the FP-growth method is based on the Apriori algorithm in MDSS.	2010	China	IEEE	6
[10]	A new approach is proposed for gene selection and multi-cancer classification based on the step-by-step improvement of classification performance (SSiCP).	2009	China	IEEE	5
[11]	Four association rule algorithms are considered in this study: Apriori Association Rule, Predictive Apriori Association Rule, Tertius Association Rule & Filtered Associator. We compare the result of these four algorithms & present the result.	2012	India	Google Scholar	7
[12]	The paper presents a parallelized incremental FP-Growth mining strategy based on MapReduce, which aims to process large-scale data.	2014	China	IEEE	6
[13]	This paper is based on the association rules data mining technology. The advantages and disadvantages of the Apriori algorithm and FP- growth algorithm are deeply analyzed in the association rules, and a new algorithm is proposed, finally, the performance of the algorithm is compared with the experimental results.	2015	China	IEEE	7
[14]	The paper reviews the features, data sets variants, support, confidence, rule generation, and candidate generation of the algorithms that are employed to mine the association rules.	2016	India	IEEE	7

Appendix D

Following are the primary studies extracted in response to research question 1

Bakar, W. A. B. W. A., Abdullah, Z. B., Saman, M. Y. B. M., Jalil, M. M. B. A., Man,

M. B., & Herawan, T. (2015, August). Vertical Association Rule Mining: Case study implementation with relational DBMS. In Technology Management and Emerging Technologies (ISTMET), 2015 International Symposium on (pp. 279-284). IEEE.

Bayardo Jr, R. J. (1998). Efficiently mining long patterns from databases. ACM Sigmod Record, 27(2), 85-93.

Cong, G., & Liu, B. (2002). Speed-up iterative frequent itemset mining with constraint changes. In Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on (pp. 107-114). IEEE.

El-Shishiny, H., Soliman, T. H., & El-Asmar, M. (2008, July). Mining drug targets based on microarray experiments. In Computers and Communications, 2008. ISCC 2008. IEEE Symposium on (pp. 175-181). IEEE.

Ghani, R., & Fano, A. E. (2002). Using text mining to infer semantic attributes for retail data mining. In Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on (pp. 195-202). IEEE.

Hipp, J., Güntzer, U., & Nakhaeizadeh, G. (2000). Algorithms for association rule mining—a general survey and comparison. ACM sigkdd explorations newsletter, 2(1), 58-64.

Lin, X., & Zhu, Q. (2010, July). Share-Inherit: A novel approach for mining frequent patterns. In Intelligent Control and Automation (WCICA), 2010 8th World Congress on (pp. 2712-2717). IEEE.

Lucchese, C., Orlando, S., & Perego, R. (2006). Fast and memory-efficient mining of frequent closed itemsets. IEEE Transactions on Knowledge and Data Engineering, 18(1), 21-36.

Meng, T., & Shyu, M. L. (2011, August). Automatic annotation of drosophila developmental stages using association classification and information integration. In Information Reuse and Integration (IRI), 2011 IEEE International Conference on (pp. 142-147). IEEE.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. (2008). Top 10 algorithms in data mining. Knowledge and information systems, 14(1), 1-37.

Zaki, M. J., Parthasarathy, S., Ogihara, M., & Li, W. (1997, August). New Algorithms for Fast Discovery of Association Rules. In KDD (Vol. 97)

Following are the primary studies extracted in response to research question 2

Aher, S. B., & Lobo, L. M. R. J. (2012). A comparative study of association rule algorithms for course recommender systems in e-learning. International Journal of Computer Applications, 39(1), 48-52.
Ayubi, S., Muyeba, M. K., Baraani, A., & Keane, J. (2009). An algorithm to mine general association rules from tabular data. Information Sciences, 179(20), 3520-3539.

Bharathi, T., & Krishnakumari, P. (2016, January). A comparative analysis on efficiency of contemporary association rule mining algorithm. In Advanced Computing and Communication Systems (ICACCS), 2016 3rd International Conference on (Vol. 1, pp. 1-9). IEEE.

Chadokar, S. K., Singh, D., & Singh, A. (2013, July). Optimizing network traffic by generating association rules using a Hybrid Apriori-Genetic algorithm. In Wireless and Optical Communications Networks (WOCN), 2013 Tenth International Conference on (pp. 1-5). IEEE.

Chung, S. M., & Mangamuri, M. (2004, April). Mining association rules from relations on a parallel NCR Teradata database system. In Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on (Vol. 1, pp. 465-470). IEEE.

Gu, X. F., Hou, X. J., Ma, C. X., Wang, A. G., Zhang, H. B., Wu, X. H., & Wang, X.
M. (2015, December). Comparison and improvement of association rule mining algorithm. In Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015 12th International Computer Conference on (pp. 383-386). IEEE.

Comito, C., & Talia, D. (2015, October). Evaluating and predicting the energy consumption of data mining algorithms on mobile devices. In Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on (pp. 1-8). IEEE.

Kuo, R. J., Chao, C. M., & Chiu, Y. T. (2011). Application of particle swarm optimization to association rule mining. Applied Soft Computing, 11(1), 326-336.

Li, Z. C., He, P. L., & Lei, M. (2005, August). A high efficient AprioriTid algorithm for mining association rule. In Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on (Vol. 3, pp. 1812-1815). IEEE.

Mazid, M. M., Ali, A. S., & Tickle, K. S. (2009, October). A comparison between rule-based and association rule mining algorithms. In-Network and System Security, 2009. NSS'09. Third International Conference on (pp. 452-455). IEEE.

Min, L., Chunyan, W., & Yuguang, Y. (2010, December). The Research of FP-Growth Method Based on Apriori Algorithm in MDSS. In Digital Manufacturing and Automation (ICDMA), 2010 International Conference on (Vol. 2, pp. 770-773). IEEE.

Peng, S., Liu, X., Yu, J., Wan, Z., & Peng, X. (2009, March). A new implementation of a recursive feature elimination algorithm for gene selection from microarray data.
In Computer Science and Information Engineering, 2009 WRI World Congress on (Vol. 3, pp. 665-669). IEEE.

Singh, T., & Sethi, M. (2015, December). Sandwich-Apriori: A combined approach of Apriori and Reverse-Apriori. In In-

dia Conference (INDICON), 2015 Annual IEEE (pp. 1-4). IEEE.

Wei, X., Ma, Y., Zhang, F., Liu, M., & Shen, W. (2014, May). Incremental FP-Growth mining strategy for dynamic threshold value and database based on MapReduce. In Proceedings of Computer Supported Cooperative Work in Design (CSCWD), 2014.