Big Data Analytics in Association Rule Mining: A Systematic Literature Review

Mahtab SHAHIN*
Information Systems Group, Tallinn University of Technology, Tallinn, Estonia

Sijo Arakkal Peious
Information Systems Group, Tallinn University of Technology, Tallinn, Estonia

Rahul Sharma
Information Systems Group, Tallinn University of Technology, Tallinn, Estonia

Minakshi Kaushik
Information Systems Group, Tallinn University of Technology, Tallinn, Estonia

Sadok Ben Yahia
Software Science Department, Tallinn University of Technology, Tallinn, Estonia

Syed Attique Shah
Data Systems Group, Institute of Computer Science, University of Tartu, Tartu, Estonia

Dirk Draheim
Information Systems Group, Tallinn University of Technology, Tallinn, Estonia

ABSTRACT
Due to the rapid impact of IT technology, data across the globe is growing exponentially as compared to the last decade. Therefore, the efficient analysis and application of big data require special technologies. The present study performs a systematic literature review to synthesize recent research on the applicability of big data analytics in association rule mining (ARM). Our research strategy identified 4797 scientific articles, 27 of which were identified as primary papers relevant to our research. We have extracted data from these papers to identify various technologies and algorithms of using big data in association rule mining and identified their limitations in regards to the big data categories (volume, velocity, variety, and veracity).

CCS CONCEPTS
• **Big data**; • **Hadoop distributed file system**; • **frequent itemset**;

KEYWORDS
Big data analytics, Association rule mining, Spark, MapReduce, systematic literature review

ACM Reference Format:
Mahtab SHAHIN, Sijo Arakkal Peious, Rahul Sharma, Minakshi Kaushik, Sadok Ben Yahia, Syed Attique Shah, and Dirk Draheim. 2021. Big Data Analytics in Association Rule Mining: A Systematic Literature Review. In *2021 the 3rd International Conference on Big Data Engineering and Technology (BDET) (BDET 2021), January 16–18, 2021, Singapore, Singapore.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3474944.3474951

1 INTRODUCTION
Due to the rapid development of science and technology, a large scale of unstructured and semi-structured data has been formed. To find useful knowledge from large data sets, it is necessary to use data mining technology. At present, a variety of data mining technologies have been created, such as association rules mining, sequence pattern discovery, etc. Association rule mining (ARM) was initially proposed by Agrawal et al. [1] as a technique to detect and extract useful information from a massive amount of data and extract useful information. ARM is used in various applications, including recommender systems [2], customer relationship management (CRM) [3], and cross-selling [4].

Association rules are typically generated in a two-step process. In the first step of the process, all frequent itemsets [5-8], i.e., all itemsets that fulfill specified minimum support, are generated for a given dataset. In the second step, each frequent itemset is used to generate all possible rules from the dataset; and all rules which do not satisfy specified minimum confidence are removed. The major step of association rule mining is in identifying frequent itemsets. Several ARM algorithms are currently in use: three typical classic representatives are Apriori [10], FP-Growth [11], and Eclat [12].

Big data is a comprehensive word for any collection of data sets that are extremely big and complex, and plays a crucial function in all aspects of an organization, for instance, marketing, health science, and clinical information [13, 14]. As shown in Fig.1, big data is composed of four characteristic features (4Vs) [15], i.e., volume, velocity, variety, and veracity of the data.

Several big data analytic techniques are used to extract, analyze, and visualize complex and different data types. In recent years, data has grown rapidly. Analyzing this data is a complex [16] and challenging task for humans. For instance, over 175 million tweets including videos, images, texts, and social relationships are generated by millions of accounts [18]. Big data analysis (BDA) helps organizations in decisions by analyzing datasets from different

*mahtab.shahin@taltech.ee

BDET 2021, January 16–18, 2021, Singapore, Singapore

ACM ISBN 978-1-4503-8928-0/21/01…$15.00
https://doi.org/10.1145/3474944.3474951

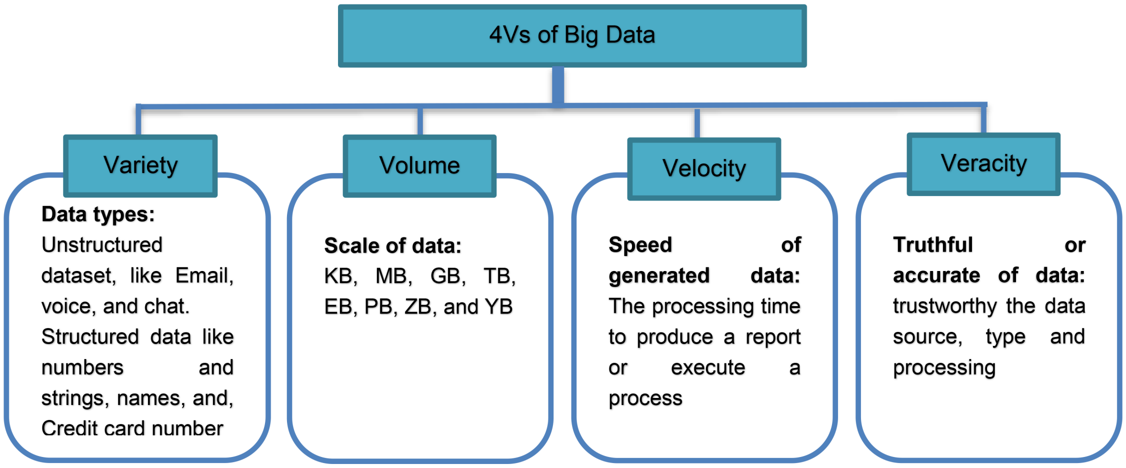

Figure 1: Big data features.

sources and developing valid information [18]. There are necessary tools for big data analysis that were examined. Each of these tools is focused on a specific field. Some are used for batch processing and others for real-time analysis. Apache Spark [19] is an open-source framework that has made a big splash since its introduction at AMP Lab at Berkeley University in 2009. Its core is a large-scale distributed processing engine that can be scaled well. Apache Spark supports four fundamental libraries for machine learning and data mining, including SparkQL, Spark Streaming, MLib, and GraphX [20].

Studies in big data have existed for over 15 years. However, there are a few studies that inquire about teaming together big data and association rule mining systematically. Therefore, in this paper, we decided to provide a systematic review of big data and association rule mining. In brief, the objectives of this research are as follows:

- Providing essential and useful information about big data and association rule mining.
- Providing a systematic review in this area.
- Qualify critical future challenges in this field and providing some suggestions for further research.
- Presenting a comparative summary of the selected articles concerning their main features.

In service of these research objective, we aim at answering the following research two concrete questions:

- RQ1: Which technologies have been used so far for association rule mining in big data scenarios?
- RQ2: What are the limitations of the found technologies in regards to the big data categories? (Volume, Velocity, Variety, and Veracity)

This paper is organized as follows: In Sec. 2, we describe the Systematic Literature Review (SLR) in more detail. In Sec. 3, each primary study is evaluated according to our evaluation criteria. Finally, Sec. 4 closes the paper with a conclusion and a brief discussion of the researchable issues.

2 METHODOLOGY

2.1 Review Method and Research Questions

Literature reviews, and in particular systematic literature reviews, have become popular in the software engineering research field to evaluate what we know in a particular topic and provide answers for specific research questions. This research has been accomplished by following Kitchenham and Charters [21] guidelines for conducting Systematic Literature Review (SLR) or Systematic Review (SR), which involves several activities such as the development of review protocol, the identification and selection of primary studies, the data extraction and synthesis, and reporting the results. We followed all these steps for the reported study as described in the following sections of this paper.

2.2 Search Strategy

The search strategy contains search terms, Academic resources, and search process, which are explained in the sequel.

2.2.1 Search Terms. The search string was expanded according to the following steps [21]:

- Identification of the search terms from research questions.
- Building an advanced search string using identified search terms, Boolean ANDs, and ORs.
- Identifying synonyms and antonyms of the search terms.
- Identifying the keywords from the related books or articles

The list of primary and secondary search terms is shown in Table 1

It should be considered that the word “technology” is usually not mentioned in the title of the articles and by including this search item in the search string, no additional relevant results can be achieved. Therefore, alternative search items, i.e., Hadoop and Spark, were included in the search string.

2.2.2 Academic Resources. Before starting the search, to increase the probability of finding relevant articles, it is necessary to select the appropriate set of data. The search for primary studies was

Table 1: Search terms used in this review

Primary Search Terms	Secondary Search Terms	Search String
big data, association rule	frequent itemset, Hadoop, spark, framework	("big data" OR Hadoop OR Spark) AND ("association rule" OR "frequent itemset" OR "frequent item set")

Table 2: Search results

Digital Library	Total Count	URL
ACM Digital Library	356	http://portal.acm.org
IEEE Xplore	217	http://ieeexplore.ieee.org
SpringerLink	2,638	http://springerlink.com
ScienceDirect	1,228	http://sceincedirect.com
Scopus	903	http://scopus.com/
Total	5,342	

conducted on the following digital libraries, ACM Digital Library, IEEE Xplore, ScienceDirect, and Springer.

2.2.3 Search Process. Table 2 presents the databases searched on October 27, 2020, and the number of relevant articles identified from each database. from the years 2012 to 2021. For this reason, we want to centralize in recent publications. As well, 2012 is when this research area in association rule mining and big data started to become popular and numerous studies have been conducted on it.

It is worth noting that there is a junction between information databases; therefore, some of the articles can appear in more than one database. Moreover, to avoid duplicate results, while searching through different databases, we manually selected other options. In total, 4,797 articles were identified after removing 363 redundant and duplicate articles (Fig. 2).

2.3 Study Selection

This section is used for selecting primary studies. Moreover, the Software package Mendeley (http://mendely.com) was used to store and manage the research results. To ensure that the articles were most likely related to our research questions, a two-phase selection process was conducted. Moreover, two researchers of this review independently analyzed the identified articles and selected the studies.

2.3.1 Selection Phase 1. In this phase, we studied the title and keywords and assessed them based on inclusion criteria as shown in the following list.

- Inclusion criteria
- IC1: Does the paper explain the theoretical foundation of association rule mining in big data?
- IC2: Is the paper about association rule mining in big data analysis?

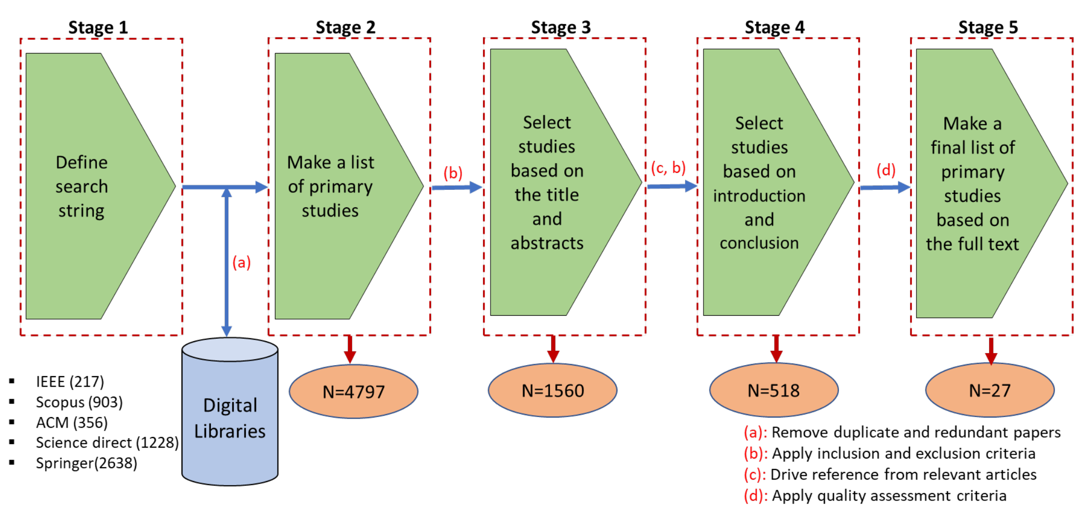

Figure 2: Search process and selection of primary studies.

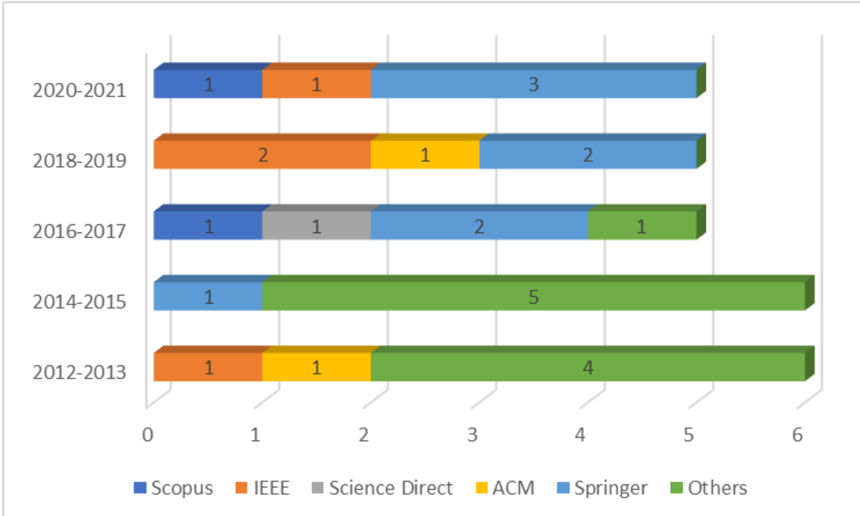

Figure 3: Distribution of selected articles by publisher.

- IC3: Is the paper discussing at least one big data technology or technique?
- IC4: Is the paper related to at least one aspect of the research questions?

We only selected papers that satisfied all of the items mentioned inclusion criteria. After this scanning, 1560 studies were found since their title and abstract be similar to searched keywords. Next, the introduction and conclusion of each study were read and their concepts were analyzed. During this phase, some studies were found to be precisely aligned with big data analysis research on the concept of association rule mining as discussed in Section 2, while others were found to be entirely out of context. At the end of this stage, 511 studies were found. Then, by scanning the references in the relevant articles, seven extra articles have been found, that were missed in the initial search. So, we added them to the list of primary studies and identified 518 relevant papers.

2.3.2 Selection Phase 2. In this phase, we applied the quality assessment to the selection of the primary studies. The quality assessment focused on researches that have enough information to answer the research questions. The questions of quality assessment are provided as follows.

- Quality assessment
- QC1: Is the objective of the study mentioned clearly?
- QC2: Does the study propose a new methodology or algorithm for big data or association rule mining?
- QC3: Are the simulations/experiments thoroughly analyzed and explain, and do the tests' results strongly support the work ideas?

All the articles were accessed by at least two researchers independently and the questions by answering "yes," "partly," and "no" to each of the established criteria. After the assessment was completed, we calculated a sum for each paper by giving one point for each "yes," 0.5 points for each "partly," and zero points for each "no." All papers that scored QC1 + QC2 + QC3 $\geq$ 2 points were accepted and included in the studies used in the data extraction and synthesis stage. The search process and selection of primary studies are shown in Fig.2. Moreover, Fig.3, depicts the number of primary studies based on the years and digital libraries. In the following, the author's name, the title of the studies, year, and type of publication are presented in Table 3

2.4 Data Extraction and Synthesis

In this stage of the review process, data extraction, a set of relevant data items was extracted from each primary study as shown in Table 4

As shown in Table 4, we have extracted data items beneficial for providing an overview of the primary studies, as well as those necessary for answering our research questions. After extracting the data, we further evaluated each primary study's relevance to our research objectives based on short descriptive summaries of primary studies prepared by each reviewer. Finally, during the data synthesis process, each of the primary studies was carefully analyzed to identify the suggested factors leading to the omission of quality practices.

3 RESULTS

This section summarizes the main obtained results and analyzes the collected data concerning the systematic literature review's research questions.

3.1 RQ1- Which Technologies Have Been Used So Far for Association Rule Mining in Big Data Scenarios?

We have identified 24 of 27 papers that can help us answer this research question. As a result, our SLR has found that big data uses various technologies for association rule mining. This review has identified and categorized these technologies. As shown in Table 5, since 2012, two and ten methods have been applied as the most

Table 3: The list of primary studies in the field of association rule mining and big data analysis

Primary Studies (PS)	Author(s) Name	Year	Study title	Publications
PS22	Yahia et.al	2012	An efficient implementation of the Apriori algorithm based on Hadoop-MapReduce model [22]	Journal
PS4	Yen Li et.al	2012	Apriori-based frequent itemset mining algorithm on MapReduce [23]	Conference
PS23	Li et.al	2012	Parallel implementation of Apriori algorithm based on MapReduce [24]	Conference
PS26	Rong et.al	2013	Complex statistical analysis of big data: Implementation and application of Apriori and FP-Growth algorithm based on MapReduce [25]	Conference
PS16	Moens et.al	2013	Frequent itemset mining for big data [26]	Conference
PS2	Thabtah, and Hammoud	2013	MR-ARM: A MapReduce association rule mining framework [27]	Journal
PS24	Qiu et.al	2014	YAFIM: A parallel frequent itemset mining algorithm with Spark [28]	Conference
PS1	Gui et.al	2015	A distributed frequent itemset mining algorithm based on spark [20]	Conference
PS12	Liang, and Wu	2015	Sequence-Growth: A scalable and effective frequent itemset mining algorithm [8]	Conference
PS19	Chavan et.al	2015	Frequent itemset mining for big data [29]	Conference
PS20	Zhang et.al	2015	A distributed frequent itemset mining algorithm using spark for big data analysis [19]	Journal
PS14	Gole et.al	2015	Frequent itemset mining for big data in social media using cluster Big FIM algorithm [30]	Conference
PS18	Chen et.al	2015	Mining association rule mining in big data with NGEP [13]	Journal
PS17	Kumar Seti, and Ramesh	2017	HFIM: A spark-based hybrid frequent itemset mining for big data processing [31]	Journal
PS10	Djenouri et.al	2017	Frequent itemset mining in big data with an effective single scan algorithm [32]	Conference
PS7	Singh et.al	2017	Performance optimization of MapReduce-based Apriori algorithm on Hadoop cluster [44]	Journal
PS9	Prasad et.al	2017	High-performance computation of big data: performance optimization approach toward a parallel frequent itemset mining algorithm for transaction data based on Hadoop MapReduce [33]	Journal
PS13	Chon, and Kim	2018	BIGMiner: A fast and scalable distributed frequent pattern miner for big data [34]	Journal
PS3	Rathee, and Kashyap	2018	Adaptive-Miner: An efficient distributed association rule mining algorithm on Spark [35]	Journal
PS25	Fu et.al	2018	Mining algorithm for association rule mining in big data based on Hadoop [36]	Journal
PS11	Bai et.al	2019	Association rule mining algorithm based on spark for pesticide transaction data analysis [37]	Journal
PS15	Gao et.al	2019	Mining frequent itemsets using improved Apriori or Spark [45]	Conference
PS8	Raj et.al	2020	EAFIM: Efficient Apriori-based frequent itemset mining algorithm on spark for big transaction data [38]	Journal
PS5	Senthilkumar et.al	2020	An efficient FP-Growth based association rule mining algorithm using Hadoop MapReduce [11]	Journal
PS21	Pal, and Kumar	2020	Distributed synthesized association rule for big transactional data [39]	Journal
PS6	Choi, and Chung	2020	Knowledge process of health big data using MapReduce-based association mining [40]	Journal
PS27	Dasgupta, and Saha	2021	Towards the speed enhancement of association rule mining algorithm for intrusion detection system [41]	Journal

Table 4: Data item extracted from primary studies

Data item extracted	Data item description	Related RQ
Study title	Table 3	Overview
Author(s) list	Table 3	Overview
Publication year	Table 3	Overview
Publication title	Table 3	Overview
The technology of big data	Table 5	RQ1
Algorithms of ARM	Table 5	RQ1
Size, and variety of big data	Table 6	RQ2

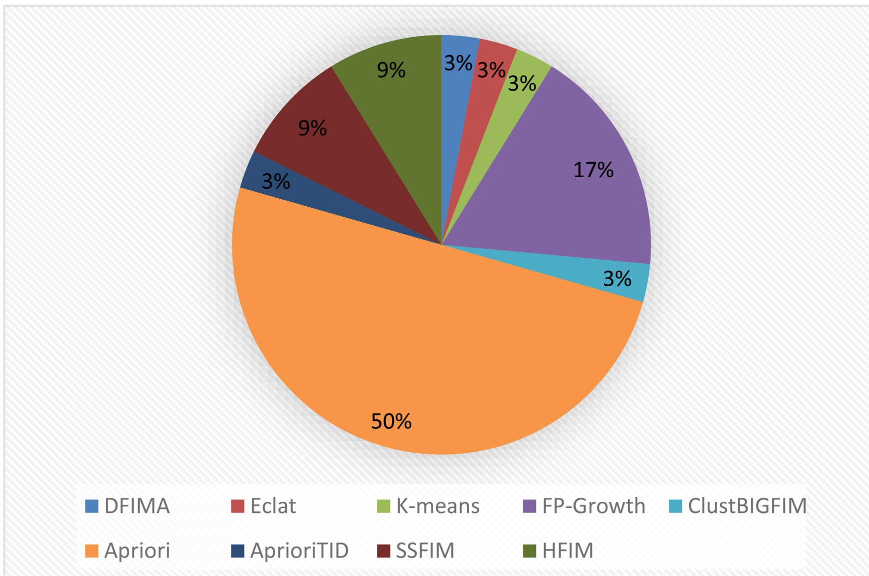

Figure 4: Distribution of used algorithm in association rule mining.

frequently used method, respectively, for big data and association rule mining.

Based on Table 5, Apriori is the most usable algorithms in ARM. The distribution of algorithms is shown in Fig.4.

Also, it can be observed that Apache Hadoop is the most used algorithm to compare Apache Spark. Fig. 4, shows this distribution. Moreover, As observed in Fig. 6, MapReduce and Ubuntu were frequently used.

3.2 RQ2: What Are the Limitations of the Found Technologies in Regards to the Big Data Categories?

To answer this research question, we extract and analyze information based on the experimental results and the datasets. Table 6 provided the details based on the feature of the applied big data set. As may be seen from the table, each primary study used various or specific datasets to test each algorithm. As mentioned before, big data has four primary features (Fig.1), where the datasets were classified based on them. The volume and Velocity in the table have been marked (✓) when the data set range satisfies the minimum of the defined value in each primary study. For example, KB, MB, GB,

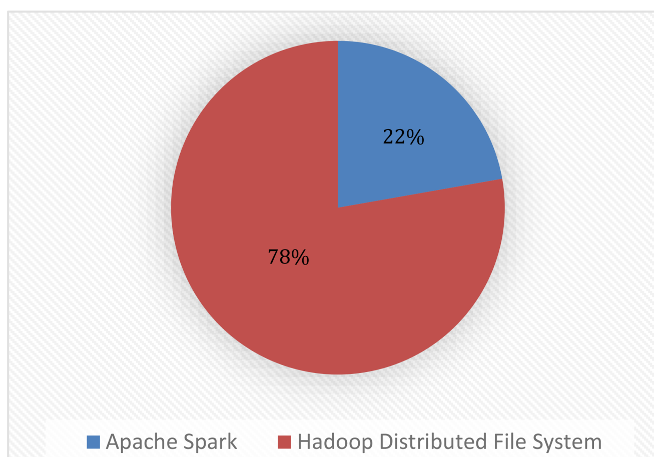

Figure 5: Distribution of used algorithm in big data.

etc., were the data set range for the volume feature. Variety has been chosen when the study applied various data sets, including

Table 5: Technologies and experimental environment used in selected primary studies

Primary Studies(PS)	Big Data Technology	ARM Algorithm	Experimental Environment
PS1	Apache Spark	DFIMA[1]	•MapReduce environment
PS2	Hadoop Distributed File System (HDFS)	FP-Growth	•Ubuntu •Java OpenJDK
PS3	Hadoop Distributed File System (HDFS)	•Apriori algorithm •FP-Growth	•Java OpenJDK •Weka
PS4	Hadoop Distributed File System (HDFS)	Apriori algorithm	•MapReduce environment
PS5	Hadoop Distributed File System (HDFS)	FP-Growth	•Ubuntu •MapReduce environment
PS6	Hadoop Distributed File System (HDFS)	Apriori algorithm	•MapReduce environment
PS7	Hadoop Distributed File System (HDFS)	Apriori algorithm	•Ubuntu •MapReduce environment
PS8	Apache Spark	Apriori algorithm	•MapReduce environment •Apache Spark framework
PS9	Hadoop Distributed File System (HDFS)	•ClustBigFIM •Apriori algorithm •FP-Growth	•MapReduce environment
PS10	Hadoop Distributed File System (HDFS)	•A new method (SSFIM[2]) •Apriori algorithm •Eclat •FP-Growth	•MapReduce environment
PS11	Hadoop Distributed File System (HDFS)	Apriori algorithm	•MapReduce environment •Ubuntu
PS12	Hadoop Distributed File System (HDFS)	•Apriori algorithm •New distributed FIM[3] algorithm (Sequence-Growth)	•MapReduce environment •Ubuntu •Java OpenJDK
PS13	Hadoop Distributed File System (HDFS),	•AprioriTid •FP-Growth	•MapReduce environment
PS14	Hadoop Distributed File System (HDFS)	•ClustBigFIM •K-means •Apriori algorithm	•MapReduce environment •Ubuntu
PS15	Apache Spark	•Apriori algorithm	•Apache Spark •Java OpenJDK •Hadoop •Ubuntu
PS16	Hadoop Distributed File System (HDFS)	•ClustBigFIM •Apriori algorithm	•Ubuntu •MapReduce environment
PS17	Apache Spark	•HFIM[4]	•MapReduce environment
PS19	Hadoop Distributed File System (HDFS)		
PS21	Hadoop Distributed File System (HDFS)	•Apriori algorithm	•MapReduce environment •Ubuntu
PS22	Hadoop Distributed File System (HDFS)	•Apriori algorithm	•MapReduce environment •Java OpenJDK
PS23	Hadoop Distributed File System (HDFS)	•Apriori algorithm	•MapReduce environment
PS25	Hadoop Distributed File System (HDFS)	•Apriori algorithm	•Ubuntu
PS26	Hadoop Distributed File System (HDFS)	•Apriori algorithm •FP-Growth	•MapReduce environment •Single-machine environment
PS27	Hadoop Distributed File System (HDFS)	FP-Growth	•Java OpenJDK •Ubuntu

Table 6: Used datasets and big data categories in selected primary studies

PrimaryStudies	Dataset	Size of dataset/Number of transactions	Volume	Velocity	Variety	Veracity
PS1	T10I4D100K[5]	3,84 MB	✓	✓		✓
PS2	Transaction dataset from FIMI Repository [43]	50-500 MB	✓	✓		✓
PS3	LastFM data	10-550K	✓	✓		
PS4	T10I4D100k, BMSWbView1, BMSPOS		✓	✓	✓	✓
PS5	IBM Quest Market-Basket Synthetic	17,5-63,7GB	✓	✓	✓	
PS6	Health big data set	Not mentioned specifically	✓	✓	✓	✓
PS8	Dense dataset (like Mushroom& Chess), T10104D100k, and Retail	10GB	✓	✓	✓	✓
PS11	The transaction information of agricultural inputs products[6]	150-400M	✓	✓	✓	✓
PS13	T10I4D100k	100,000 transaction	✓	✓		✓
PS15	Extended Bakery Dataset, and Retail Dataset	100000, 88163 transactions	✓	✓	✓	✓
PS16	Abstract [44], T10I4D100K, Mashroom, and Pumsb	158,029 Transactions	✓	✓	✓	
PS17	Chess, Mashroom, and T10104D100k	10,64 Transactions	✓		✓	
PS18	Iris[7], and ASD[8]	3000-10000 Transactions	✓	✓	✓	✓
PS19	C20d10k, Chess, Mushroom					
PS20	T40l10D100K[9], and T10I4D100K	14,8, and 3,84MB, Respectively	✓	✓	✓	
PS21	Accident, Chess, KDD99, Mushroom, PAMAPP, PowerC, Pumsb, Susy, US Cenus, and T10I4D100K	8416, 3196, 1000000, 8416, 1000000, 1040000, 49046, 5000000, 1000000, 100000, transaction	✓	✓	✓	✓
PS22	T10I4D100k, Quest Synthetic Data Generated by IBM		✓	✓		✓
PS23	T10I4D100K, T10I4D200K, T10I4D400K, and T10I4D800K	1, 2, 4, and 8GB	✓	✓	✓	
PS26	Real datasets	32-1024 MB	✓	✓		✓
PS27	Kyoto (real network traffic data)	128-708 MB	✓	✓	✓	✓

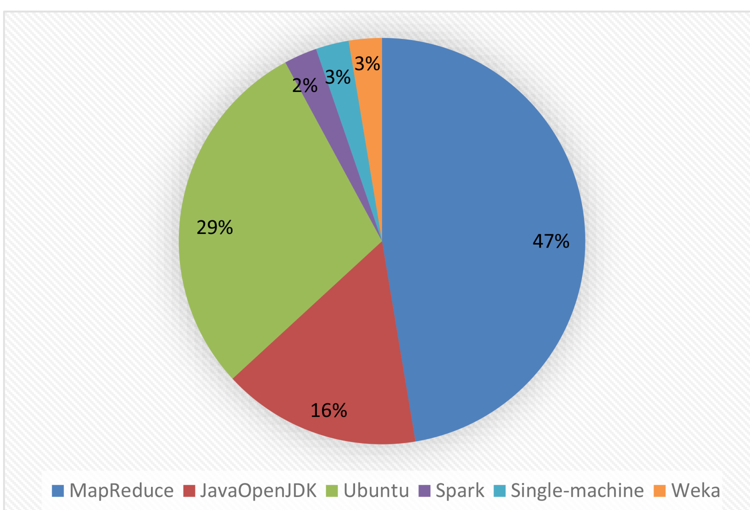

Figure 6: Distribution of used experimental environment in big data and association rule mining

both structured and unstructured, or a mix of some different structure datasets. Therefore, it has not been marked if the study used only one of the data sets. Veracity has been marked when the study has reported truthful results compared to other works with a similar approach. For instance, in PS27, in the recently published work [42], Kyoto used as the data set, where the volume of the data set was between 128 and 708 MB, velocity between 0.5 and 0.65 s, difference items as variety, and in a sum up, better results were reported to comparison the previous works.

4 CONCLUSION AND FUTURE WORK

This literature review aims to identify and analyze the trends, datasets, methods, and frameworks used in association rule mining and big data analysis between 2012 and 2021. Based on the designed inclusion and exclusion criteria, finally, 27 studies published between January 2012 and January 2021 remained and have been investigated. This literature review has been undertaken as a systematic literature review. The systematic literature review is defined as a process of identifying, assessing, and interpreting all available research evidence with the purpose to provide answers for specific research questions. Analysis of the selected primary studies revealed that focus on five topics: estimation, association, classification, clustering, and dataset analysis. Based on the primary studies, emerging data mining, big data with parallelization, and association rule to improve the usage of huge, complex datasets. Data mining literature already has sequential and parallel algorithms for finding frequent itemsets. Nine different methods have been applied to association rule mining. From the nine methods, the two most applied methods in association rule mining are identified. They are Apriori and FP-Growth. The results of this research also identified six experimental environments to execute experiments of association rule mining in big data analysis. They are MapReduce, Ubuntu, Java OpenJDK, Spark, single-machine, and Weka. Also, the total distribution of big data methodology is as follows. 78% of the research studies applied to Hadoop Distributed File System, and 22% of the studies applied to Apache Spark. Moreover, identified the kind of big dataset which applies in big data frameworks, and the most used dataset was T10I4D100k[22, 23, 24, 26, 31, 38, 20, 34, 19]. Based on Table 6, among all features of big data, veracity has the most limitations. Choosing the right algorithm can be very effective in solving this issue.

To enhance this review's finding, we intend to conduct a comprehensive survey of big data and association rule mining in real-world settings and identify the best experimental method for each data set concerning the big data categories.

ACKNOWLEDGMENTS

This work has been partially conducted in the project "ICT programme" which was supported by the European Union through the European Social Fund.

REFERENCES

[1] R Agrawal, T. Imieliński, and A. Swami. 1993. Mining Association Rules Between Sets of Items in Large Databases. ACM SIGMOD Rec. 22, 2, 207–216, doi: 10.1145/170036.170072.

[2] Lawrence, Richard D., George S. Almasi, Vladimir Kotlyar, Marisa Viveros, and Sastry S. Duri. 2001. Personalization of supermarket product recommendations." In Applications of data mining to electronic commerce, pp. 11-32. Springer, Boston, MA, 2001

[3] Seyed A. Shirkhorshidi, S. Aghabozorgi, T. Ying Wah, and T. Herawan. Big data clustering: a review. In International conference on computational science and its applications, pp. 707-720. Springer, Cham, 2014.

[4] T. Brijs, G. Swinnen, K. Vanhoof, and G. Wets. 1999. Using association rules for product assortment decisions. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 254-260. 1999.

[5] R. Agrawal, T. Imielinski, and A. Swami. 1993. Mining Association rules between set of items in Large Databases," In Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp. 207-216. 1993.

[6] Kaushik, M., Sharma, R., Peious, S. A., Shahin, M., Yahia, S. B., & Draheim, D. (2021). A Systematic Assessment of Numerical Association Rule Mining Methods. SN Computer Science, 2(5), 1-13.

[7] M. Kaushik, R. Sharma, D. Draheim, and M. Shahin. 2020. On the Potential of Numerical Association Rule Mining. International Conference on Future Data and Security Engineering. Springer, Singapore, 2020

[8] Shah, S. A., Seker, D. Z., Hameed, S., & Draheim, D. (2019). The rising role of big data analytics and IoT in disaster management: recent advances, taxonomy and prospects. IEEE Access, 7, 54595-54614.

[9] Draheim, D. (2017). FP Semantics of Jeffrey Conditionalization. In Generalized Jeffrey Conditionalization (pp. 33-39). Springer, Cham.

[10] W. Z. Cheng and X. Li Xia. 2014. A fast algorithm for mining association rules in image. Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, 513–516, doi: 10.1109/ICSESS.2014.6933618.

[11] A. Senthilkumar. 2020. An efficient FP-Growth based association rule mining algorithm using Hadoop MapReduce, Indian J. Sci. Technol., 13, 34, 3561–3571, doi: 10.17485/ijst/v13i34.1078.

[12] M. J. Zaki, S. Parthasarathy, M. Ogihara, W. Li, P. Stolorz, and R. MusickP. 1997. Arallel Algorithms for Discovery of Association Rules. Scalable High Perform. Comput. Knowl. Discov. Data Min., 373, 5–35, 1997, doi: 10.1007/978-1-4615-5669-5_1.

[13] Y. Chen, F. Li, and J. Fan. 2015. Mining association rules in big data with NGEP. Cluster Comput., 18, 2, 577–585, 2015, doi: 10.1007/s10586-014-0419-3.

[14] C. Yesheng, S. Kara, and Ka C. Chan. Manufacturing big data ecosystem: A systematic literature review. Robotics and computer-integrated Manufacturing 62 (2020): 101861.

[15] W. Inoubli, S. Aridhi, H. Mezni, M. Maddouri, and E. Mephu Nguifo. 2018. An experimental survey on big data frameworks," Futur. Gener. Comput. Syst., 86, 546–564, 2018, doi: 10.1016/j.future.2018.04.032.

[16] L. Wang, K. Lu, P. Liu, R. Ranjan, and L. Chen. 2014. IK-SVD: Dictionary learning for spatial big data via incremental atom update. Comput. Sci. Eng., 16, 4, 41–52, doi: 10.1109/MCSE.2014.52.

[17] H. S. Bhosale and D. P. GadekarA. 2014. Review Paper on Big Data and Hadoop," Int. J. Sci. Res. Publ., 4, 10, 1–7, 2014.

[18] S. A. Shah, D. Z. Seker, M. M. Rathore, S. Hameed, S. Ben Yahia, and D. Draheim. 2019. Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers? IEEE Access, 7, 91885–91903, 2019, doi:10.1109/ACCESS.2019.2928233.

[19] F. Zhang, M. Liu, F. Gui, W. Shen, A. Shami, and Y. Ma. 2015. A distributed frequent itemset mining algorithm using spark for big data analytics. Cluster Comput., 18, 4, 1493–1501, doi: 10.1007/s10586-015-0477-1.

[20] F. Zhang, M. Liu, F. Gui, W. Shen, A. Shami, and Y. Ma. 2015. A distributed frequent itemset mining algorithm using spark for big data analytics. Cluster Comput., 18, 4, 1493–1501, 2015, doi: 10.1007/s10586-015-0477-1.

[21] Barbara A. Kitchenham and S. Charters. 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007-01. Keele University. 2007.

[22] O. Yahya, O. Hegazy, and E. Ezat. 2012. An efficient implementation of Apriori algorithm based on Hadoop-Mapreduce model. Proc. of theInternational Journal of Reviews in Computing 12 (2012).

[23] M. Y. Lin, P. Y. Lee, and S. C. Hsueh. 2012. Apriori-based frequent itemset mining algorithms on MapReduce. In Proceedings of the 6th international conference on ubiquitous information management and communication, pp. 1-8. 2012.

[24] N. Li, L. Zeng, Q. He, and Z. Shi. 2012. Parallel implementation of apriori algorithm based on MapReduce. Proc. - 13th ACIS Int. Conf. Softw. Eng. Artif. Intell. Networking, Parallel/Distributed Comput. SNPD 2012, 236–241, doi: 10.1109/SNPD.2012.31.

[25] Z. Rong, D. Xia, and Z. Zhang. 2012. Complex statistical analysis of big data: Implementation and application of apriori and FP-growth algorithm based on MapReduce. Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, 2012, 968–972, 2013. doi:10.1109/ICSESS.2013.6615467.

[26] S. Moens, E. Aksehirli, and B. Goethals. 2013. Frequent Itemset Mining for big data. Proc. - 2013 IEEE Int. Conf. Big Data, Big Data 2013,1, 111–118, doi: 10.1109/BigData.2013.6691742.

[27] Sh, Ahsan, and Z. Halim. On efficient mining of frequent itemsets from big uncertain databases. Journal of Grid Computing 17, no. 4, 2019: 831-850.

[28] H. Qiu, R. Gu, C. Yuan, and Y. Huang. 2014. YAFIM: A parallel frequent itemset mining algorithm with spark. Proc. - IEEE 28th Int. Parallel Distrib. Process. Symp. Work. IPDPSW 2014, 1664–1671, 2014, doi: 10.1109/IPDPSW.2014.185.

[29] K. Chavan, P. Kulkarni, P. Ghodekar, and S. N. Patil. Frequent itemset mining for Big data. Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015, 1365–1368, 2016, doi: 10.1109/ICGCIoT.2015.7380679.

[30] S. Gole and B. Tidke. 2015. Frequent itemset mining for Big Data in social media using ClustBigFIM algorithm. 2015 Int. Conf. Pervasive Comput. Adv. Commun. Technol. Appl. Soc. ICPC 2015, c, doi: 10.1109/PERVASIVE.2015.7087122.

[31] K. K. Sethi and D. Ramesh. 2017. HFIM: a Spark-based hybrid frequent itemset mining algorithm for big data processing. J. Supercomput., 73, 8, 3652–3668. doi: 10.1007/s11227-017-1963-4.

[32] Y. Djenouri, D. Djenouri, J. C. W. Lin, and A. Belhadi. 2018. Frequent itemset mining in big data with effective single scan algorithms," IEEE Access, 6, 68013–68026, doi: 10.1109/ACCESS.2018.2880275.

[33] M. S. Guru Prasad, H. R. Nagesh, and S. Prabhu. 2017. High performance computation of big data: Performance optimization approach towards a parallel frequent item set mining algorithm for transaction data based on hadoop mapreduce Framework. Int. J. Intell. Syst. Appl., 9, 1, 75–84, 2017, doi: 10.5815/ijisa.2017.01.08.

[34] K. W. Chon and M. S. Kim. 2018. BIGMiner: A fast and scalable distributed frequent pattern miner for big data. Cluster Comput., 21, 3, 1507–1520, doi: 10.1007/s10586-018-1812-0.

[35] S. Rathee and A. Kashyap. 2018. Adaptive-Miner: an efficient distributed association rule mining algorithm on Spark. J. Big Data, 5, 1, 2018, doi: 10.1186/s40537-018-0112-0.

[36] C. Fu, X. Wang, L. Zhang, and L. Qiao. 2018. Mining algorithm for association rules in big data based on Hadoop," AIP Conf. Proc., 1955, no. April, doi: 10.1063/1.5033699.

[37] X. Bai, J. Jia, Q. Wei, S. Huang, W. Du, and W. Gao. 2019. An association rule mining algorithm based on spark for pesticide transaction data analyses. Int. J. Agric. Biol. Eng., 12, 5, 162–166, 2019, doi: 10.25165/j.ijabe.20191205.4881.

[38] S. Raj, D. Ramesh, M. Sreenu, and K. K. Sethi. 2020. EAFIM: efficient apriori-based frequent itemset mining algorithm on Spark for big transactional data," Knowl. Inf. Syst., 62, 9, 3565–3583, doi: 10.1007/s10115-020-01464-1.

[39] A. Pal and M. Kumar. 2020. Distributed synthesized association mining for big transactional data. Sadhana - Acad. Proc. Eng. Sci., 45, 1, 2020, doi: 10.1007/s12046-020-01380-8.

[40] S. Y. Choi and K. Chung. 2020. Knowledge process of health big data using MapReduce-based associative mining. Pers. Ubiquitous Comput., 24, 5, 571–581, 2020, doi: 10.1007/s00779-019-01230-3.

[41] S. Dasgupta and B. Saha. 2021. Towards the speed enhancement of association rule mining algorithm for intrusion detection system, 1180 AISC. Springer International Publishing.

[42] A. K. Koliopoulos, P. Yiapanis, F. Tekiner, G. Nenadic, and J. Keane. 2015. A Parallel Distributed Weka Framework for Big Data Mining Using Spark. Proc. - 2015 IEEE Int. Congr. Big Data, BigData Congr. 2015, 9–16, doi: 10.1109/BigDataCongress.2015.12.

[43] T. De Bie. 2011. An information theoretic framework for data mining. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 564–572, doi:10.1145/2020408.2020497.

[44] Singh, S., Garg, R., & Mishra, P. K. 2018. Performance optimization of MapReduce-based Apriori algorithm on Hadoop cluster. Computers & Electrical Engineering, 67, 348-364.

[45] Inoubli, W., Aridhi, S., Mezni, H., Maddouri, M., & Nguifo, E. (2018, August). A comparative study on streaming frameworks for big data. In VLDB 2018-44th International Conference on Very Large Data Bases: Workshop LADaS-Latin American Data Science (pp. 1-8).