International Journal of Innovative Engineering Applications

Journal homepage: https://dergipark.org.tr/ijiea

ANOMALY DETECTION FOR GEAR MANUFACTURING DOWNTIME IN THE AUTOMOTIVE SECTOR USING RARE ITEMSET MINING

***Devrim Naz Akdas*[1], *Derya Birant*[2], *Pelin Yildirim Taser**[3]**

[1]Dokuz Eylül University, Graduate School of Natural and Applied Sciences, Izmir, Turkey
[2]Dokuz Eylül University, Department of Computer Engineering, Izmir, Turkey
[3]Izmir Bakırçay University, Department of Computer Engineering, Izmir, Turkey

Abstract

Original scientific paper

Downtimes in manufacturing significantly influence productivity, and their analysis is necessary for successful and flexible production. Although some classification and regression studies have been performed on the machine downtime in the manufacturing area, the rare itemset mining (RIM) technique has never been implemented in the existing downtime studies until now. Besides, anomaly detection for gear manufacturing downtime in the automotive sector using RIM is yet to be explored. To bridge this gap, this study proposes the application of the RIM method for detecting anomalies in gear manufacturing downtime of earth moving machinery for the first time. In this study, the Rare Pattern Growth (RP-Growth) algorithm was executed on a real-world dataset consisting of downtimes in gear manufacturing of earth moving machinery to discover rare itemsets that indicate anomalies in downtimes. In the experiments, the rare itemsets (anomalies) in the downtime data were detected using different minimum support (minsup) and minimum rare support (minraresup) threshold values. The obtained results were also evaluated in terms of the number of itemsets, execution time, and maximum memory usage. The experimental results show that the proposed approach, called *Anomaly Detection with Rare Itemset Mining* (ADRIM), is an effective method for detecting anomalies in machine downtimes and can be successfully used in the manufacturing area, especially in the automotive sector.

Keywords: *Anomaly detection, data mining, gear manufacturing, rare itemset mining.*

1 Introduction

Manufacturing downtime is described as any period of time during which a machine is not producing [1]. The downtime in manufacturing is classified into two types: planned and unplanned. Planned downtime occurs when production equipment is restricted or shut down to accommodate planned repairs, maintenance, testing, or upgrades. The unplanned downtime in manufacturing is any unexpected stop of equipment that is scheduled to be in operation. While unplanned downtimes are a much more costly process, all downtimes have a price. When equipment downtime occurs, quality, availability, and performance components of Overall Equipment Effectiveness (OEE) are affected negatively. It is vital to analyze machine downtimes in order to improve lean manufacturing processes. Because of this reason, manufacturing companies aim to discover anomalies in both planned and unplanned downtimes for reducing the disruption times of machines and increasing operational efficiency.

Anomaly detection, an active study field in a variety of research communities, is the process of discovering data samples that significantly deviate from the norm [2]. Recently, different data mining techniques(RIM) have been used in the field of anomaly detection. Association rule mining (ARM) is a rule-based data mining technique that discovers interesting relations and common patterns among itemsets in large databases [3]. Rare itemset mining (RIM) is a subfield of ARM that aims to reveal out a rare and low-rank set of items in a dataset. In this study, the RIM technique was implemented using the Rare Pattern Growth (RP-Growth) algorithm for detecting anomalies in gear manufacturing downtimes of earth moving machinery.

The novelty and main contributions of this study are as follows: (i) It gives a brief survey of ARM and RIM techniques, (ii) It proposes an approach, called *Anomaly Detection with Rare Itemset Mining* (ADRIM), (iii) This study is the first attempt to implement the RIM method for detecting anomalies in machine downtimes, (iv) This study is also original in that it applies the RIM method using RP-Growth algorithm on a real-world dataset which includes gear manufacturing downtimes of earth moving machinery, (v) This study evaluates the obtained experimental results in terms of the number of itemsets, execution time, and maximum memory usage with the help of charts and tables.

This article is structured as follows. In the following section, related literature work on the subject is given. Section 3 presents brief background information about ARM and RIM paradigms. This section also explains the

* Corresponding author.
E-mail address: pelin.taser@bakircay.edu.tr

Received 02 February 2022; Received in revised form 19 October 2022; Accepted 19 October 2022

 Doi: https://doi.org/10.46460/ijiea.1067365

general structure of the proposed anomaly detection system and the used RP-Growth algorithm. In Section 4, the real-world machine downtime dataset used in this study is described. The obtained experimental results with discussions are also presented in this section. Finally, concluding remarks and future directions are clarified in Section 5.

2 Related Work

In the literature, there are several machine learning-based studies that aim to detect [4-6] and reduce [7-9] downtimes in different application areas, such as wind turbine, geology, production, etc. Wang et al. [5] used Support Vector Regression (SVR) to reduce the time-consuming fitness evaluation process for predicting schedules in response to machine breakdown. In another study [7], the researchers applied the C4.5 decision tree algorithm to identify and reduce the downtime in a production system. Nwanya et al. [8] implemented a Multiple Regression algorithm for optimizing machine downtime in plastic manufacturing.

RIM technique has been frequently preferred in many areas, such as data streams [10], distributed systems [11], energy [12], healthcare [13-15], marketing [16,17], network [18], and web applications [19,20]. For example, Hemalatha and Lakshmi [10] proposed a Minimal Infrequent Pattern-based Outlier Detection (MIFPOD) algorithm for investigating outliers in data streams based on obtained minimal infrequent patterns. The authors also defined three different measures for this algorithm, such as Three measures, namely Transaction Weighting Factor (TWF), Minimal Infrequent Deviation Factor (MIPDF), and Minimal Infrequent Pattern-based Outlier Factor (MIFPOF). In the study of [16], rare itemsets with a high-utility (RIHU) algorithm was proposed for discovering rare itemsets with high utility for revenue analysis. Bakariya and Thakur [20] introduced a Rare Itemset Mining from Weblog Data (RIMWD) algorithm that follows top-down and level-wise approaches for discovering all rare itemsets with their subsets.

To the best of our knowledge, a RIM paradigm has not been performed in the existing downtime studies until now. To bridge this gap, this study applies the RIM method for detecting anomalies in gear manufacturing downtime of earth moving machinery.

3 The Proposed Approach (ADRIM)

3.1 Rare Itemset Mining (RIM)

Association rule mining (ARM), the most popular pattern discovery technique in data mining, aims to find interesting correlations, common patterns, or associations among a set of items in large databases or data repositories [21]. The associations among items are frequently described using association rules. Formally let the set of distinct items $I = \{i_1, i_2, \dots, i_n\}$, and the set of transactions (transactional database) $D = \{t_1, t_2, \dots, t_m\}$, where each transaction $t \subseteq I$, n is the number of items and m is the number of transactions. An association rule is denoted as $X \rightarrow Y$ such that $X \cup Y \subseteq I$ and $X \cap Y = \emptyset$, where X (antecedent) and Y (consequent) are frequent itemsets. The rule is interpreted as if itemset X be in a transaction, itemset Y will very certainly occur in the same transaction. These rules are generated by satisfying user-defined criteria, such as support and confidence. While support indicates the frequency of transactions that include both itemsets X and Y, confidence gives the proportion of transactions including itemset X that also include itemset Y. The equations of the support and confidence thresholds are as follows:

$$Support(X \rightarrow Y) = \frac{|X \cup Y|}{|D|} \tag{1}$$

$$Confidence(X \rightarrow Y) = \frac{|X \cup Y|}{|X|} \tag{2}$$

where $|D|$ is the number of transactions in the database D, $|X|$ is the number of X itemsets, and $|X \cup Y|$ is the number of transactions that contains X and Y. For example, an association rule $X \rightarrow Y$ with 60% support and 70% confidence values means that 60% of the transactions contain X and Y itemsets together, and 70% of the transactions contain X, in which item Y also appears.

The main objective of ARM is extracting only those itemsets that are frequent relative to a particular threshold, such as minimum support and minimum confidence values. However, in many real-life problems, there could be significant itemsets but have a low support value (infrequent itemsets). To discover the relation among those itemsets with a low frequency of occurrence, the rare itemset mining (RIM) paradigm has been proposed in the literature.

RIM is a special type of ARM that attempts to discover rare correlations among a set of items in a dataset. An itemset R is a rare itemset if its support value is lower than minimum support (minsup) threshold and higher or equal to minimum rare support (minraresup) threshold, as shown in Equation (3).

$$\begin{aligned} &Support(R) < minsup \, \Lambda \\ &Support(R) \geq minraresup \end{aligned} \tag{3}$$

The field of RIM has grown in popularity recently, owing to its applicability in a variety of applications such as anomaly detection, discovering rare diseases, intrusion detection, and identifying infrequently purchased items. In the literature, several RARM algorithms have been proposed. The RP-Growth algorithm has been preferred in this study because of its efficiency and scalability.

3.2 Rare Pattern Growth

RP-Growth is a variant of the FP-Growth algorithm that focuses on mining rare itemsets in the large transactional database [22]. A rare itemset is one that appears within the range indicated by user-defined minraresup and minsup thresholds. The main advantages of the RP-Growth algorithm are its speed and efficient memory usage.

3.3 Anomaly Detection with Rare Itemset Mining (ADRIM)

Figure 1 presents the general overview of the implemented ADRIM approach in this study. First, gear manufacturing data of earth moving machineries are obtained in the data collection step. Then, in the next step, the experimental dataset is passed through a data preprocessing step (feature selection, feature extraction, and data discretization) to make it suitable for applying the RIM algorithms. In the RIM step, the RP-Growth algorithm is implemented on the experimental data for discovering rare itemsets that indicate anomalies in gear manufacturing downtimes of earth moving machinery. After that, the obtained rare itemsets are evaluated using user-defined minsup and minraresup thresholds. Finally, the detected anomalies in gear manufacturing downtimes are analyzed for decision making.

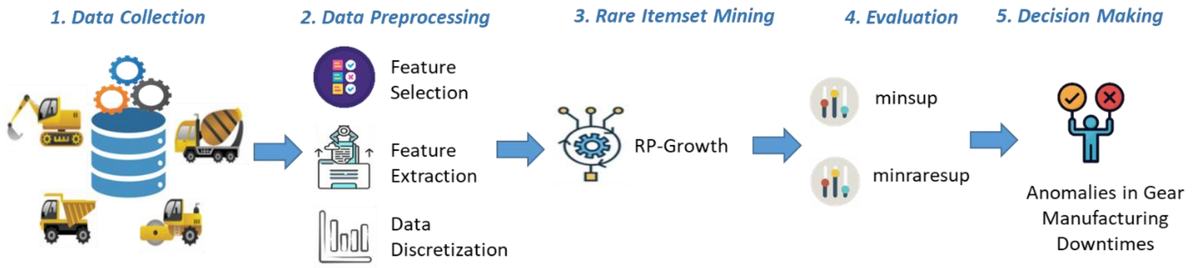

Figure 1. The general overview of the ADRIM approach for detecting anomalies in gear manufacturing downtimes.

4 Experimental Study

In the experiments of this study, the RP-Growth algorithm was executed on real-world experimental data for detecting anomalies in gear manufacturing downtimes of earth moving machinery. This algorithm was used for discovering rare itemsets that indicate anomaly downtimes. The proposed ADRIM approach was developed using an open-source data mining library specialized for pattern mining, written in the Java programming language [23]. The experiments were performed on a personal computer with an Intel Core i5-7200U 3.1-GHz processor and 8 GB of RAM.

4.1 Dataset Description

A real-world dataset considered in this study includes records of downtimes in gear manufacturing of earth moving machinery obtained from a factory in Izmir. The dataset consists of 11040 instances and eleven attributes that give information about occurred downtimes in machines.

The experimental dataset has passed through some data preprocessing steps (i.e., feature selection, feature extraction, and data discretization) for the implementation of the RP-Growth algorithm. First, in the feature selection step, the irrelevant and redundant attributes such as ID, name and surnames of operators, company names, etc. were extracted from the dataset. In the next step, the month and day attributes were derived from the date attribute to make it useful for analysis. In the final step (data discretization), the numeric attributes (time and total production) were discretized into five different intervals because the RP-Growth algorithm can work only on categorical data. Table 1 presents the attributes and their categorical values of the used dataset.

Table 1. The attributes and their categorical values of the dataset.

Month	**Day**	**WorkCenterCode**	**Downtime**	**DowntimeGroupName**	**OperationName**
January	Monday	IM01	Undefined	Breakdown	CNC turning
February	Tuesday	IM02	Failure / Maintenance	Quality	Hard turning
March	Wednesday	IM03	Launch	Personnel	Gear cutting
April	Thursday	IM04	Initial adjustment	Planned	Finishing and grinding
May	Friday	IM05	Settings correction	Workbench preparation	
June	Saturday	IM06	Changing tool holder	Other	
July	Sunday	IM07	Changing cutting tool		
August		IM08	Drill sharpening		
September		IM09	Warming the bench		
October		IM10	Assigning another task		
November		IM11	Carrying material		
December		IM12	Waiting for material		
		IM13	Personal need		
		IM14	Cleaning		
		IM15	No operator		
		IM16	Setting		
		IM17	Emending		
		IM18	Waiting for approval		
		IM19	Sampling		
DowntimeType	**Department**	**WorkCenterGroupCode**	**Time**	**TotalProduction**	
Planned Unplanned	CNC	IMG110	[0-3) = Very low	[0-30) = Very low	
	Gear Finishing	IMG130	[3-15) = Low	[30-80) = Low	
		IMG131	[15-30) = Medium	[80-130) = Medium	
		IMG134	[30-50) = High	[130-200) = High	
		IMG141	[>50] = Very high	[>200] = Very high	

4.2 Experimental Work

In this study, four different experiments were performed on the gear manufacturing downtime dataset for analyzing the followings:

(i) samples of the discovered rare itemsets (downtime anomalies),
(ii) the relationship between the number of rare itemsets and the minsup and varying minraresup values,
(iii) the execution time performances of the RP-Growth algorithm on different minraresup values,
(iv) the maximum memory usage of the RP-Growth algorithm on different minraresup values.

The first experiment implements the RP-Growth algorithm to discover rare itemsets that indicate detected anomalies in gear manufacturing downtimes. Table 2 represents some examples of the obtained rare itemsets with their lengths, support values, minsup thresholds, and minrare thresholds. For example, the itemset "Day = Saturday, WorkCenterCode = IM05, OperationName = Hard turning" having three items expresses that it occurs only in 0.12% of all downtime records. This means that On Saturdays, downtime at the work center IM05, where Hard turning operation is performed, is rarer than other downtimes. Another rare itemset obtained from the experiment is "DowntimeType = Unplanned, Time = Very high" with 7.25% support value. Considering this rare itemset, it is possible to decide that unplanned downtimes with very high durations are rare and undesirable. In other words, this itemset indicates that the downtimes are generally planned and of less duration.

Table 2. Examples of rare itemsets discovered by the RP-Growth algorithm.

Length	Rare Itemset	Support (%)	Minsup (%)	Minraresup (%)
1-Itemset	Day=Sunday	4.73	90	4
	Month=December	2.79	90	2
	WorkCenterCode=IM18	2.48	90	2
2-Itemsets	Department=Gear, WorkCenterGroupCode=IMG134	10.03	80	10
	DowntimeType=Unplanned, Time=Very high	7.25	80	5
	Day=Monday, DowntimeGroupName=Breakdown	5.21	80	5
3-Itemsets	WorkCenterCode=IM18, DowntimeType=Unplanned, TotalProduction=Very low	0.15	70	0.1
	Downtime=Changing tool holder, DowntimeGroupName=Workench preparation, OperationName=GearCutting	0.14	70	0.1
	Day=Saturday, WorkCenterCode=IM05, OperationName=Hard turning	0.12	80	0.1
4-Itemsets	Month=October, Day=Monday, WorkCenterCode =IM02, OperationName=CNC turning	2.33	50	2
	OperationName=Finishing and grinding, DowntimeType=Unplanned, WorkCenterGroupCode=IMG134, Time=Very high	1.68	30	1
	Month=April, OperationName=Gear cutting, Department=Gear, TotalProduction= Low	1.82	30	1
5-Itemsets	Downtime=Waiting for approval, DowntimeGroupName=Quality, OperationName=Gear cutting, DowntimeType=Unplanned, Department=Gear	2.07	10	1
	DowntimeGroupName=Personnel, OperationName=CNC turning, DowntimeType=Unplanned, Department=CNC, WorkCenterGroupCode=IMG110	1.98	10	1
	Downtime=Waiting for material, DowntimeGroupName=Personnel, OperationName=Gear cutting, DowntimeType=Unplanned, TotalProduction=Very low	0.18	80	0.1

In the second experiment, the RP-Growth algorithm was applied with a 90% minsup threshold and different minraresup threshold levels (%) ranging from 15 to 55 in increments of 5. Figure 2 presents the relation between the number of rare itemsets and the varying minraresup threshold values when the minsup value is kept constant. This figure shows that the higher the minraresup threshold value, the lower the number of rare itemsets discovered. This graph suggests that more rare itemsets are formed when the algorithm is conducted with higher minraresup values.

The third experiment evaluated execution times of the RP-Growth algorithm on the dataset with a constant minsup threshold value (90%) and various minraresup threshold levels (%) ranging from 15 to 55 in increments of 5. The obtained execution times are given in Figure 3 in milliseconds (ms). The varying minraresup values significantly affect the execution times of the experiments. This figure indicates that the execution time of the RP-Growth algorithm increases almost linearly as the minraresup threshold value increases. This result proves the RP-Growth algorithm's effectiveness in terms of speed.

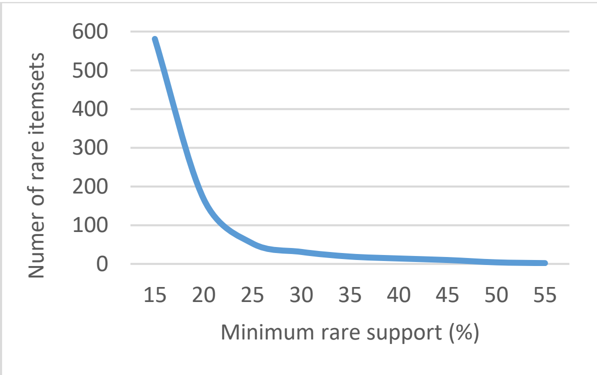

Figure 2. The number of rare itemsets with different minraresup thresholds.

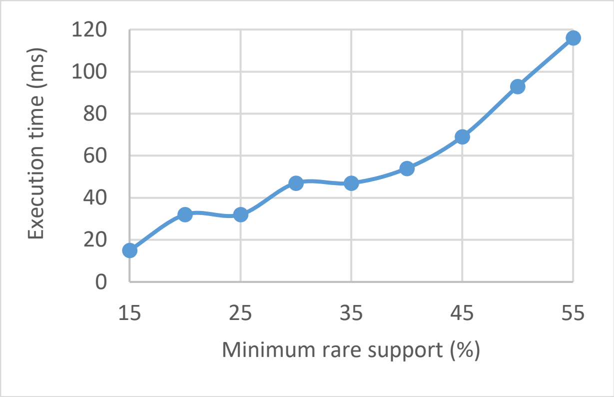

Figure 3. The execution time performance of the RP-Growth algorithm with different minraresup thresholds.

In the last experiment, the maximum memory usage of the RP-Growth algorithm was calculated in terms of a megabit (mb) with the same minsup and minraresup threshold values used in the first three experiments. The results are presented as a graph in Figure 4. It is obviously seen from this graph that The RP-growth algorithm provides efficient memory usage with acceptable levels.

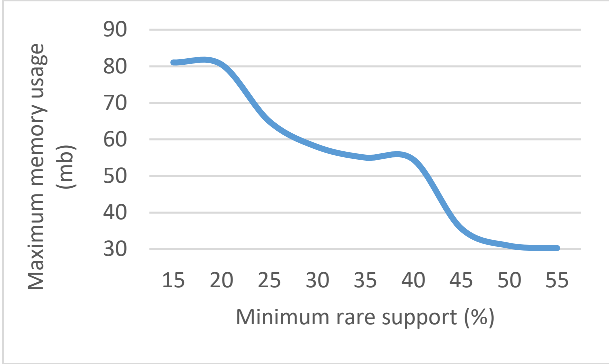

Figure 4. The maximum memory usage performance of the RP-Growth algorithm with different minraresup threshold.

5 Conclusion

Analyzing machine downtimes with detecting anomalies play an important role in improving lean manufacturing processes. For this purpose, this study proposes an approach, called anomaly detection with rare itemset mining (ADRIM). We implemented the RIM technique for detecting anomalies in gear manufacturing downtime of earth moving machinery. In the experiments, the RP-Growth algorithm, a modification of the FP-Growth algorithm, has been utilized because of its efficiency and scalability. This algorithm was executed on a real-world dataset, and the anomalies in the downtimes were detected to help experts make decisions about productivity and efficiency. Also, the experimental results were evaluated in terms of the number of itemsets, execution time, and maximum memory usage in this study.

As a future study, different popular anomaly detection algorithms (i.e., isolation forest, Local Outlier Factor, etc.) can be applied to the same dataset. In addition, the RIM method can be implemented for different domains with different algorithms apart from the RP-Growth.

Acknowledgements

The authors are deeply grateful to Çelikiş Spare Parts of Industry and Trade Company Limited, and Projesis Software Consulting Informatics Information Systems for providing the experimental dataset used in the study.

Declaration

Ethics committee approval is not required.

References

[1] Patale, G., Turukmane, R., Gulhane, S., Kolte, P., & Gupta, K. K. (2021). Causes of machine down time and its impact on nylon manufacturing process. *Man-Made Textiles in India*, *49*(5), 158-160.

[2] Böhmer, K., & Rinderle-Ma, S. (2020). Mining association rules for anomaly detection in dynamic process runtime behavior and explaining the root cause to users. *Information Systems*, *90*, 101438.

[3] Taşer, P. Y., Birant, K. U., & Birant, D. (2020). Multitask-based association rule mining. *Turkish Journal of Electrical Engineering & Computer Sciences*, *28*(2), 933-955.

[4] Mucchielli, P., Bhowmik, B., Ghosh, B., & Pakrashi, V. (2021). Real-time accurate detection of wind turbine downtime-An Irish perspective. *Renewable Energy*, *179*, 1969-1989.

[5] Wang, D. J., Liu, F., & Jin, Y. (2019). A proactive scheduling approach to steel rolling process with stochastic machine breakdown. *Natural Computing*, *18*(4), 679-694.

[6] Shafieezadeh, A., DesRoches, R., Rix, G. J., & Werner, S. D. (2014). A probabilistic framework for correlated seismic downtime and repair cost estimation of geo-structures. *Earthquake engineering & structural dynamics*, *43*(5), 739-757.

[7] Garcés, D. A., & Castrillón, O. D. (2017). Design of intelligent technology to identify and reduce downtime in a production system. *Información Tecnológica*, *28*, 157-70.

[8] Nwanya, S. C., Udofia, J. I., & Ajayi, O. O. (2017). Optimization of machine downtime in the plastic manufacturing. *Cogent Engineering*, *4*(1), 1335444.

[9] Mohan, T. R., Roselyn, J. P., Uthra, R. A., Devaraj, D., & Umachandran, K. (2021). Intelligent machine learning based total productive maintenance approach for achieving zero downtime in industrial machinery. *Computers & Industrial Engineering*, *157*, 107267.

[10] Hemalatha, C. S., Vaidehi, V. & Lakshmi, R. (2015). Minimal infrequent pattern based approach for mining outliers in data streams. *Expert Systems with Applications*, *42*(4), 1998–2012.

[11] Chandran, C. R., & Padmanabhan, A. (2016). An efficient algorithm for detecting outliers in a distributed environment using minimal infrequent item set pattern mining. *IIOAB Journal*, *7*(9), 22-25.

[12] Sun, C., Wang, X., & Zheng, Y. (2019). Data-driven approach for spatiotemporal distribution prediction of fault events in power transmission systems. *International Journal of Electrical Power & Energy Systems*, *113*, 726-738.

[13] Jin, H., Chen, J., He, H., Williams, G. J., Kelman, C., & O'Keefe, C. M. (2008). Mining unexpected temporal associations: applications in detecting adverse drug reactions. *IEEE Transactions on Information Technology in Biomedicine*, *12*(4), 488-500.

[14] Shrivastava, K., & Jotwani, V. (2020). Study to Determine Adverse Diseases Pattern using Rare Association Rule Mining. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *6*(3), 519-526.

[15] Reps, J. M., Aickelin, U., & Hubbard, R. B. (2016). Refining adverse drug reaction signals by incorporating interaction variables identified using emergent pattern mining. *Computers in biology and medicine*, *69*, 61-70.

[16] Selvarani, S., & Jeya Karthic, M. (2019). Rare Itemsets with high utility for Revenue Analysis. *Journal of Computational and Theoretical Nanoscience*, *16*(4), 1402-1407.

[17] Jeyakarthic, M., & Selvarani, S. (2019). An efficient approach using FM-weight for revenue prediction on rare itemsets. *International Journal of Recent Technology and Engineering*, *7*, 226-232.

[18] Rahman, A., Ezeife, C. I., & Aggarwal, A. K. (2010). Wifi miner: An online apriori-infrequent based wireless intrusion detection system. *Knowledge Discovery from Sensor Data* (Sensor-KDD 2008), *5840*, 76-84.

[19] Adda, M., Wu, L., White, S., & Feng, Y. (2012). Pattern detection with rare item-set mining. *International Journal on Soft Computing, Artificial Intelligence and Applications*, *1*(1), 1-17.

[20] Bakariya, B., & Thakur, G. S. (2016). Mining rare itemsets from weblog data. *National Academy Science Letters*, *39*(5), 359-363.

[21] Yildirim, P., Birant, D., & Alpyildiz, T. (2017). Discovering the relationships between yarn and fabric properties using association rule mining. *Turkish Journal of Electrical Engineering & Computer Sciences*, *25*(6), 4788-4804.

[22] Tsang, S., Koh, Y. S., & Dobbie, G. (2011, August). Rp-tree: rare pattern tree mining. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 277-288). Springer, Berlin, Heidelberg.

[23] Fournier-Viger, P., Lin, J. C. W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., & Lam, H. T. (2016, September). The SPMF open-source data mining library version 2. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 36-40). Springer, Cham.