IJARCCE
ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⪤ Vol. 11, Issue 1, January 2022
DOI: 10.17148/IJARCCE.2022.11129

A MODEL FOR EXPLORATION OF PERIODIC PATTERNS IN A DATABASE

[1]Johnny, F. A., [2]Bennett, E. O. [3]Sako, D. J. S.

[1,2,3]Department of Computer Science, Rivers State University of Port Harcourt, Nigeria

ABSTRACT: The Inability to determine better patterns in a group of events that will lead to particular deviations in customers' behaviour and difficulty in detection of a repetitive approach to a particular sales company or a group of companies in order to identify regularities in the business or in a particular sector of a business as regards to purchasing item at different seasons. which have become a problem in data mining. However, periodic mining patterns from transactional database require an exponential mining space to produce a huge number of patterns, the first priority for a mining algorithm is the efficient discovery of user-interest-based periodic patterns. It is often necessary to mine a limited, interesting representative subset of frequent trends in many real-world scenario. This paperpresents a model for efficient exploration of periodic pattern in big data. Suffix and prefix trees have been used to capture the contents of the database in a very compact way to generate the full set of periodic-frequent patterns in a database for frequency and support thresholds provided by the user. A periodic pattern algorithm was developed to efficiently list all periodic item-sets. Themodel was implemented in Jupyter notebook using python programming language. The results show that some of the patterns discovered in this database are appearing not only frequently within the database but also appearing at regular intervals within the database at minSup 0.1% and maxPrd 10%.

Keywords: Data mining, Support Vector Machine, Logistic Regression, Recognition

1. INTRODUCTION

An economical technique for exploration of periodic pattern in huge knowledge merely suggests that to mini knowledge sporadically victimization formular for a specific purpose. As this knowledge that is just too huge to maneuver around, it additionally moves too quick, it is unstructured and does not fit into regular architectures. To achieve price from such knowledge, the assorted fields that is generated from such massive amounts of knowledge within the e-commerce industries are banks, jumia.com, konga.com, jiji.com, e-bay, etc. periodic pattern mining, aim to find those frequent patterns that occur at regular intervals during a temporally ordered transactional information, this was studied in [1] with the aim of distinctive frequent periodic patterns since the shapes of a pattern's incidence in database cannot be determined by the attention-grabbing measures (such as support and closure) employed in frequent pattern-mining approaches. To boot, [2] planned a distinct live (regular frequent pattern mining), measured because the variance among frequent pattern periods, so as to sight periodic patterns in transaction-like database. On the opposite hand, [3] introduced associate economical approach to sight and establish regular behavior patterns that exhibit complete cyclic repetitions from Body Sensing element Networks (BSN)

2. RELATED WORKS

[4] Projected economical mining of partial periodic patterns in statistic info. They adoptedassociate Apriori-inspired search through pattern area employing a novel prefix-based organization known as a max-sub pattern tree. [5] Planned a general model for mining asynchronous periodic patterns in temporal database. Hinge upon this in their description of a collection of four algorithms for mining periodic patterns. The basic plan remains to conduct a level-wise search through pattern area, however increased with a lot of economical information structure and algorithms than earlier approaches, every formula enumerate lot of advanced patterns from the output of an earlier stage. [6] Projected effective periodic pattern mining in statistic database is the purpose of interest during this technique was the necessity to beat the limitation of generating versatile patterns by permitting event skipping in between attention-grabbing events. employing a suffix tree, it's inconceivable to skip a specific character in an exceedingly generated pattern wherever the pattern could be a combination of many characters and everyone is an illustration of freelance event in an exceedingly statistic info. [7] projected objective and subjective algorithms for grouping association rules. A pattern agglomeration algorithmic program, referred to as Subjective Grouping (SG), was developed which contains domain data and teams the foundations in keeping with the linguistics info of objects within the rules. [8] projected representative association rules and minimum condition most consequence association rules. The notion of Representative Association Rules (RAR) was introduced. RAR may be a least set of rules that covers all association rules. later on, a user is also supplied with the set of RAR's rather than the full set of association rules. However, once required, all usual association rules are often generated from the set of RAR's by suggests that of a canopy operator. [9] projected mechanical phenomenon

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⪤ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11129

pattern mining. They notice frequent movement patterns that represent additive behavior of moving objects wherever a pattern, referred to as T-pattern was outlined as a sequence of points with temporal transitions between consecutive points. A T-pattern is discovered if it's special and temporal elements or so correspond to the input sequences (trajectories). Which means of those patterns a completely different objects visit identical places with similar time intervals. Once the patterns square measure is discovered, the classical sequence mining algorithms is applied to seek out frequent patterns. [10] planned mining graph evolution rules. the matter of extracting graph evolution rules is completed in label evolving graphs. Evolving graph are outlined as a straight forward graph with initial strategy to find frequent isomorphs subgraphs that are connected and whose support is larger than a user outlined threshold. Then graph evolution rules are outlined from these patterns that describe edges rising within the future. as an example, during a social network wherever the perimeters represent co-authorship and therefore the labels represent the degree of the node, a pattern that describes an extremely label author connected to many authors with medium labels and that is later connected to a replacement author with a medium label might describe an area discriminatory attachment. [11] planned mining periodic behavior in dynamic social networks. They sought for regular patterns, i.e., patterns that occur at regular or near regular time intervals, notwithstanding they're infrequent. as an example, during a dynamic graph representing the “social” interactions of animals, a pattern might describe seasonal association. To work out the support of a pattern, they outlined a Periodic Subgraph Embedding (PSE), additionally named as periodic support set as an associate in nursing ordered set of your time steps separated by a relentless range of timestamps.

3 SYSTEM DESIGN

Architectural design specifies the structure, views and action of a system. The proposed system uses Suffix and Prefix methods to identify periodic frequent purchase pattern. Figure 1 shows the component of periodic pattern mining system.

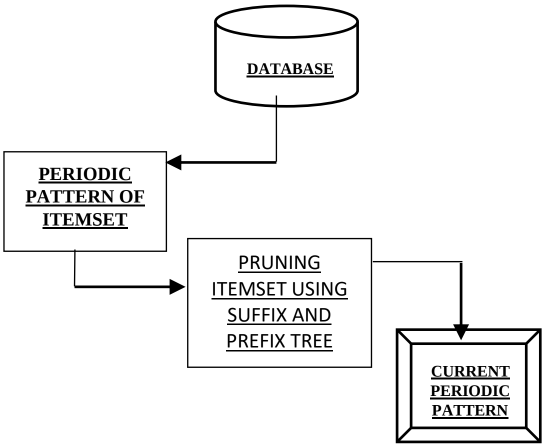

Figure 1: Architectural Design of the Proposed System

Periodic pattern involves deciding all those patterns that are unveiling either complete or partial cyclic repetitions during a statistic. cyclicity exposure, that's a vicinity of periodic pattern mining, it is a method for checking regularities of patterns’ occurrences at intervals the statistic.

The periodic pattern mining system is predicated on the suffix and prefix tree pruning methodology. during these strategies, a price is appointed to nodes within the tree, if the values matched for the tree, then derived set of periodic patterns are hold on, otherwise the patterns are made with acceptable price.

TABLE 1 SYMBOLS AND NOTATIONS USED

S/N	Symbols	Notation
1	tid	represents a transaction-id (or timestamp)
2	Y	Represents a pattern
3	TDB	represents the size of TDB in total number of transactions.
4	P^X	Is the complete set of periods of X in TDB
5	$Per(X)$	Is the periodicity of a pattern X

DOI: 10.17148/IJARCCE.2022.11129

6	Tid	Is the transaction Id
7	TID^X	Is the set of all transaction-ids where X occurs in TDB

3.1 PERIODIC PATTERN

The model of periodic-frequent patterns described as follows:

Let I = $\{i_1, i_2, \ldots, i_n 1 \leq n$, be a set of items. A set X = $\{i_1, \ldots, i_k\} \subseteq$ I, where j $\leq$ k and j, k $\in [1,n]$ is called a pattern (or an item-set).

A transaction t = (tid , Y) is a tuple, where tidrepresents a transaction-id (or timestamp) and Y is a pattern. A transactional database (TDB) over I is a set of transactions.

TDB = $\{t_1, t_2, \ldots, t_n\}$ where TDB represents the size of TDB in total number of transactions.

If X $\subseteq$ Y, it is said that t contains X and such transaction ids is denoted as tid_j^x, j $\in [1,m]$.

Let $TID^x = \{tid_j^x, \ldots, tid_k^x\}$, $j, k \in [1,m]$ and j ≤ k, be the set of all transaction-ids where X occurs in TDB. The support of a pattern X is the number of transactions containing X in TDB, which is denoted as Sup(X).

Therefore, Sup(X) = TID^x. Let tid_j^x and tid_j^x, i,j$\in [1,m-1]$ be two consecutive transactionids where X has appeared in TDB.

The period of a pattern X is the number of transactions or the time difference between tid_i^x and tid_j^x.

Let $P^x = \{P_1^x, P_2^x, \ldots \ldots, P_k^x\}$,r = Sup(X) + 1, be the complete set of periods of X in TDB.The periodicity of a pattern X is the maximum difference between any two adjacent occurrences of X, denoted as Per(X) = $max(P_1^x, P_2^x, \ldots \ldots, P_k^x)$. A pattern X is a periodic-frequent purchase pattern if Sup(X) ≥ minSup and Per(X) ≤ maxPer, where minSup and maxPer represent the user-defined thresholds on support and periodicity respectively. Both support and periodicity of a purchase pattern can be described in percentage of Transaction Database (TDB).

Example, let dataset I = {A,B,C,D,E,F,G,H} and transaction id 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 be represented in Table 1.

Table 2: Identification of Periodic purchase Pattern

Tid	Items
1	CDE
2	EGH
3	CDEH
4	ABCH
5	CDEG
6	AFGH
7	CDEF
8	CDE
9	DCG
10	CDEH
11	CDE
12	ABDF
13	CDEA
15	CDEB

The periods for the purchase pattern are: (1) = 1, (3-1) = 2, (5-3) = 2, (7-5) = 2, (8-7) = 1, (10-8) = 2, (11-10) = 1, (13-11) = 2, (15-13) = 2. The set of items containing ‘C’, ‘D’ and ‘E’. Therefore, ‘CDE’ is a purchase pattern. This purchase pattern contains 3 items. The periodicity of CDE, Per(CDE) = maximum (1,2,2,2,1,2,1,2,2) = 2.

frequent pattern because Support(CDE) ≥ minSup and Periodicity(CDE) ≤ maxPer.

Periodic pattern mining could also be helpful in analysing the client looking patterns; could also be a client is first off reaching to get a watch, shoes and phone at intervals six months. If this forms a pattern, then it can be accustomed advertise for different customers. For instance; a component might contain a collection of item conjointly known as events. However, things at intervals associate degree components aren't ordered as shown in Table 2.

© IJARCCE This work is licensed under a Creative Commons Attribution 4.0 International License

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ※ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11129

Table 3: Periodic Purchase of Items

PID	EID	ITEMS
1	1	a
1	1	bc
2	2	bcd
3	3	bc
3	2	abc
3	3	de
4	4	ef
4	1	egh
4	5	bce
5	6	ab

The identification of item or event <a> with its associated periodic id and event id is depicted in Table 4. Periodic Identity (PID) indicates period or season when an item is purchased. Event Identify (EID) indicates event or item number.

Table 4 Periodic Pattern of Item (Event) <a>

PID	EID
1	1
3	2
5	6

The periodic and event number of <a> are all the occurrence of <a> which includes PID (1) EID (1), PID (3) EID (2) and PID (5) EID (6).
The identification of item or event <b> with its associated periodic id and event id is depicted in Table 4.

Table 5 Periodic Pattern of Item <b>

PID	EID
1	1
2	2
3	3
3	2
4	5
5	6

The periodic and event number of<b> are all the occurrence of <b> which includes PID (1) EID (1), PID (2) EID (2) , PID (3) EID (3), PID (3) EID (2), PID (4) EID (5) and PID (5) EID (6).
The identification of item or event <ab> with its associated periodic id and event id is depicted in Table 5.

Table 6 Periodic Pattern of Item <ab>

PID	EID (a)	EID (b)
1	1	1
2		2
3	2	3
4		4
5	5	5

The periodic and event number of <ab> are all the occurrence of <ab> which includes PID (1) EID(a) (1) EID(b) (1), PID (2) EID(b) (2), PID (3) EID(a) (2) EID(b) (3), PID (4) EID(b) (4) and PID (5) EID(a) (5) EID(b) (5).

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⪤ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11129

3.2 SUFFIX AND PREFIX TREE

Suffix tree pruning for periodic pattern mining as shown in Figure 2. A dataset consists of patterns (A, B, C). The routine expand matches makes positive that the quantity of mismatches to the illustration string doesn't transcend a definite limit. At some node, if the mining tree determines that the support is lesser than a definite limit, it snips away that sub-tree within the suffix tree and persists its traversal.
However, if it discovers a length having the mandatory support, it flags the result.

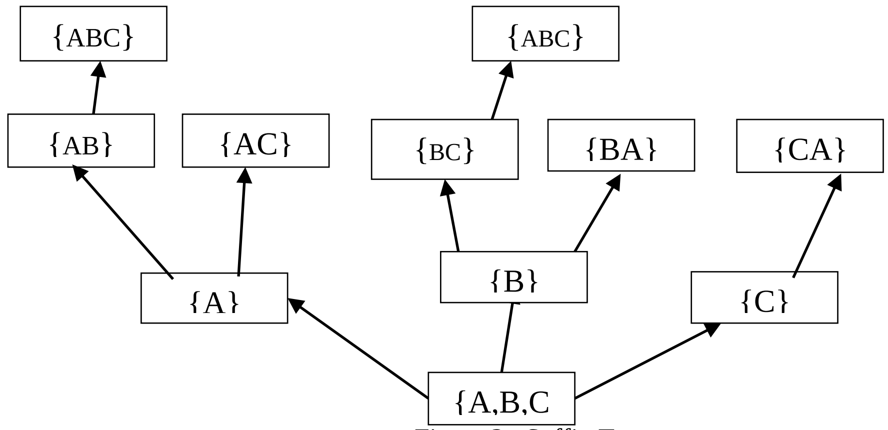

Figure 2: Suffix Tree

A pattern is termed as a recurrent pattern if its support is no less than a user specified smallest amount support threshold.A prefix tree is a prearranged tree construction which can symbolize a transaction database as shown in Figure 3. Every node in the tree symbolizes an item. All nodes at root point might be at level 0 and its offspring at point 1.

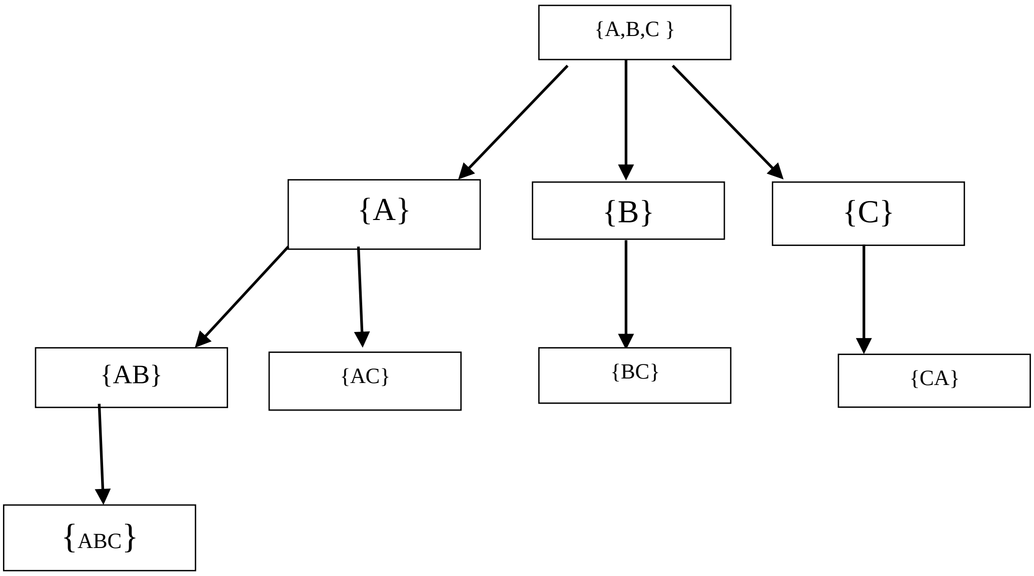

Figure 3: Illustration of Prefix Tree

The children of a nodes are generated and organized in step with the chosen authorship ordering. Prefix traverses the authorship tree in a very depth-first manner and prunes the completesubtrees nonmmoving at infrequent periodic patterns. Once the entire set of frequent sequences within the transactional info is mined, we want to separate out such patterns that are frequent, so as to get the set of distinction patterns. this may be done by taking every pattern, specified support(s) $\geq$ min_sup, this may be incontestable in Figure 4.
For example, given a sequence of patterns, find the complete set of subsequences; that is satisfying the min_sup threshold.

IJARCCE
ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ※ Vol. 11, Issue 1, January 2022
DOI: 10.17148/IJARCCE.2022.11129

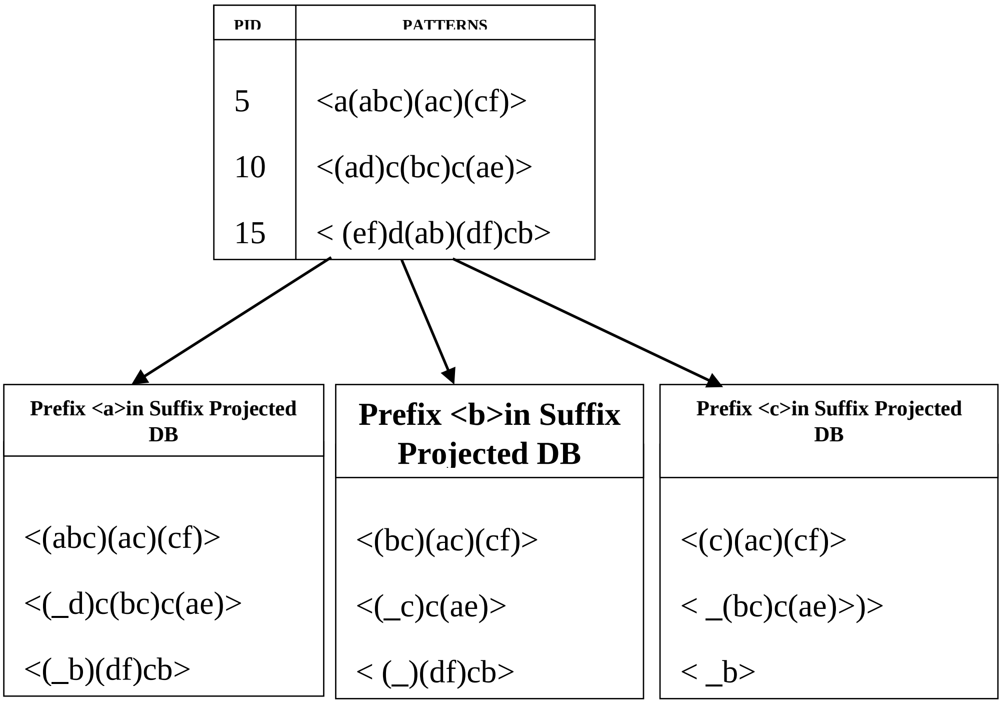

Figure 5: Periodic Pattern with Prefix and Suffix (Projected Database)

Given <a(abc)(ac)(cf)>, the prefix are as follows: <a>, <aa>, <a(ab)>, <a(abc)>. In prefix, if the item in the front is frequent, it captures it and their projection becomes the suffix. In prefix <a>, the position holder (_) shifts to d, in prefix <b> position holder shifts to c and in prefix <c> the position holder shifts to (. However, this method divides search space and mine each projected DB; <a> projected DB, <b> projected DB, <c> projected DB. The items (bc) appears in periodic id (PID) 5 and 10, therefore Item (bc) satisfies a minimum support of 2.

3.3.3 ALGORITHM TO MINE PERIODIC PATTERN

Input: A, as prefix itemset

DB_A, as suffix (project database of A) k, as length of itemset A

Output: set of LES with prefix A

1. generate the (k+1) itemset A consisting of the prefix itemset of A, each i,j in LES
2. put A' into the set of (k+1) B_{k+1}
3. for each (k+1) itemset A' ϵ Bk+1 do
4. scan DB_A to calculate the projected sub-database $DB_{A'}$ of A' and support of A'
5. calculate the PID(A')
6. if (PID(A') ≥ min_sup
7. then $LES_{k+1} \leftarrow LES_{k+1} \cup$ A'
8. call LES(A', $DB_{A'}$, k+1)
9. LES = (A', $DB_{A'}$, k+1);
10. return LES with the prefix A

The prefix A is recursively determined to directly turn out the whole set of LES. Firstly, it generates the (k+1)-itemset A consisting of the prefix itemset A and every i,j in LES. After that, place them into the set of Bk+1 (which may be a potential candidate of k+1). every itemset in Bk+1 is then processed and therefore the DBA' is scanned to calculate the projected sub-database as DBA' of A', PID(A') (which is that the Periodic Identity) and support(A'). A constraint is then applied to any confirm whether or not its extensions ought to be explored for the later projection search. If one itemset is thought to be a LES, the LES procedure is unceasingly dead. All projections of A' that is AN extension of A ought to be place into set of LES. Lastly, the LES procedure returns LES with a prefix A.

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⋇ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11129

4. RESULTS AND DISCUSSION

We implemented an approach to mining the periodic transaction of data patterns from a bakery serving Scandinavian, hot chocolate, coffee, pizza, jam, muffin, cookies and pastry. Getting such a transaction database, Table 6 shows the number of items sold from October 2018 to April 2019 (7 months) in retails dataset. There are 94 unique items sold by the Bakery, total 20507 items sold in 159 days throughout 7 months, with an average of 128.9748427672956 items sold daily. Bakery sold an average of 129 items daily.

Table 6: Number of Item Sold from October 2018 to April 2019 Discovered in Retail Data

Dates in months/Year & Time	Number of Item Sold
October 2018. 01:034am to23:31pm.	369
November 2018. 01:20am to 22:48pm.	4436
December 2018. 01:31am to 20:05pm.	3339
January 2019.05:45am to 19:31pm.	3356
February 2019. 02:28am to 23:58pm.	3906
March 2019. 00:31am to 22:31pm.	3944
April 2019. 03:04 to23:30pm.	1157

Finding patterns appearing more or less everyweekend in the transaction database of a customer. Table 7 consists of all items sold in weekday transactions beginning from 1st to 7th day, the average number of items sold were mined. Coffee is the highest sold item found in 6thweekday.

Table 7: average Number of Items Sold Discovered in Retail Data.

Weekday	Item	Average Number of Items Sold
1	pastry	110.6666
2	sandwich	104.0000
3	Brown-bread	100.9130
4	cake	115.0434
5	bread	135.8260
6	coffee	200.2173
7	tea	134.5652

The periodicity values of the top 10 itemsets is depicted in Figure 8. All itemsets that have support value over 1%.

Table 8: Periodic Pattern Discovered in Retail Data

S|N	Antecedents	Consequents	Antecedent Support	Consequent Support	Support	Periodicity
	Pattern					
1	Toast	Coffee	0.033597	0.478394	0.023666	0.07593
2	Sliced Bread	Coffee	0018172	0.478394	0.010882	0.02189
3	Brown-bread	Coffee	0.061807	0.478394	0.035182	0.05614
4	Pastry	Coffee	0.086107	0.478394	0.047544	0.06351
5	Alfajores	Coffee	0.036344	0.478394	0.019651	0.02264
6	Juice	Coffee	0.038563	0.478394	0.020602	0.02154
7	Sandwich	Coffee	0.071844	0.47839 4	0.038246	0.03877
8	Cake	Coffee	0.103856	0.478394	0.054728	0.05044
9	Scone	Coffee	0.034548	0.478394	0.018067	0.01539
10	Cookies	Coffee	0.054411	0.478394	0.028209	0.02179

The proposed model suffix+prefix tree has mined 673kb within 12 seconds and has performed better than Apriori algorithm as depicted in the Table above.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⋇ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11129

Table 9: Comparison

Models	Parameters		
	Records Mined	Data Size (KB)	Mining Time in Seconds
Suffix+prefix Tree	20507	673	7
Apriori algorithm	20507	673	23

5. CONCLUSION

We have presented our findings on mining of periodic patters. However, discovered a connected work titled "Efficient mining of partial periodic patterns", exploring some fascinating properties associated with partial cyclicity, like the Apriori property.The limitation of the approach lies on finding long periodic patterns of components and high procedure price. Technology has introduced new behavioral buying habits for shoppers, and also the emergence of e-commerce has offered retailers new ways in which to satisfy those customers request so as to seek out opportunities for development and facilitate shoppers product offering, retailers square measure more and more turning to knowledge analytics. Periodic pattern mining has been one in all the key approaches employed by major retailers to find product interconnections. It works by looking for mixtures of products that usually occur in transactions and buying patterns *between customers'. A periodic pattern algorithmic program was developed to expeditiously list most periodic item-sets. Results from associate experimental associate experimental assessment of real database have shown that the algorithm program developed is effective with a restricted variety of periodic patterns has been found.

REFERENCES

1) An, A., Khan, S. and Huang, X. (2003). Objective and Subjective Algorithms for Grouping Association Rules. *Proceedings Third IEEE International Conference on Data Mining (ICDM)*, pp.477-480.
2) Berlingerio, M., Bonchi, F., Bringmann, B. and Gionis, A. (2009). Mining Graph EvolutionRules". In *ECML/PKDD*, (1), pp. 115–130
3) Giannotti, F., Nanni, M., Pinelli, F. and Pedreschi, D. (2007). Trajectory pattern mining. In *ACM SIGKDD*, pp. 330–339.
4) Han, J., Yin, Y. and Dong, G. (1999) Efficient mining of partial periodic patterns in time series database. In proceedings of the 15th international conference on data engineering. *IEEE Computer Society*, pp. 106–115.
5) Huang, K. Y. and Chang, C. H. (2005). SMCA: a general model for mining asynchronous periodic patterns in temporal databases. *IEEE Trans Knowledge Data Engineering*, 17(6), pp. 774– 785.
6) Lahiri, M. and Berger-Wolf, T. Y. (2008). Mining Periodic Behavior in Dynamic Social Networks. In *ICDM*, pages 373–382.
7) Nishi, M. A. and Ahmed, F. A. (2017). Effective periodic pattern mining in time series databases. *Expert Systems with Applications*, 40, pp. 3015-3027.
8) Rashid, M. M., Karim, M. R., Jeong, B. S. and Choi, H. J. (2012). Efficient Mining Regularly Frequent Patterns in Transactional Databases. In Database Systems for Advanced Applications, *Proceedings of the International Conference on Database Systems for Advanced Applications*, pp. 15–19.
9) Tanbeer, S. K., Ahmed, C. F., Jeong, B. S. and Lee, Y. K. (2008). Mining regular patterns in transactional databases. *IEICE Trans. Information System*, 91, pp. 2568–2577.
10) Tanbeer, S. K., Hassan, M. M., Almogren, A., Zuair, M. and Jeong, B. S. (2016). Scalable regular pattern mining in evolving body sensor data. *Future Generational Computer System*.