

# Customer Purchasing Behavior using Sequential Pattern Mining Technique

Ashish Goel

Research Scholar

(Department of Computer Science of Engineering)  
Galgotia Collage of Engineering Technology,  
Uttar Pradesh Technical University, Gr.Noida, U.P

Bhawana Mallick, PhD

Dean Academics & HOD (CSE)

(Department of Computer Science of Engineering)  
Galgotia Collage of Engineering Technology,  
Uttar Pradesh Technical University, Gr.Noida, U.P

## ABSTRACT

In this competitive world, wherever each organization or company must improve yourself for collaborating within the market. For this improvement, we'd like to grasp the client purchasing behavior [2]. Presently data processing provides several techniques to enhance it. Aim of this paper is to shortcoming of the "Frequent Pattern Mining Technique" and use the "Sequential Pattern Mining Technique" to enhance the client purchasing Behavior. Because Apriori generate millions of candidate sets [19] and scan the group action information repeatedly and FP-Growth generate the Large No. of Projected database. As we all know that one amongst the foremost fashionable data processing approaches is "Clustering Analysis", that is truly helpful for divide the information attribute into similar variety of teams that take the high intra similarity and low lay to rest similarity and "Sequential Pattern" technique to seek out the co-relations between attributes of a relation and have applications in promoting, monetary and retail sector and it's unremarkably applied to investigate market baskets to assist organizers to work out that things are consecutive purchased along by customers.

This paper proposes an efficient technique to extract information from transactions records that is extremely helpful for increasing the client Satisfaction. Client details are divided victimization K-means [1] and sequential Pattern[4] Mining Technique "Prefix" algorithm is applied to spot client behavior over Banking knowledge.

Firstly, we divide into the database into a n number of partition with the help of Clustering technique Then Sequential algorithm provide the sequential patterns over banking data of Indian bank. According to the result of the both technique, the processing time of mining is decreased and the efficiency of algorithm has improved.

## Keywords

Data Mining, K-Means algorithm, Clustering, Sequential Pattern Mining Technique, FP-Growth Algorithm, Prefix Span Algorithm, User behavior.

## 1. INTRODUCTION

In the times data mining is incredibly vital to extract the vital patterns from the large repository due to information size is simply too large and each one got to search the most effective patterns from the large data, data mining technique is employed by the various organizations to optimize the data for locating the most effective pattern that followed by the shoppers thus data mining technique is employed by the various fields.

The main purpose of the data mining is 1st, our capability to gather and store the massive quantity of knowledge is speedily increasing day by day. As a result of the decrease within the price of storage devices and increase within the process power of computers, currently days it's possible to store vast quantity of structure knowledge and method it. The second however a lot of vital reason is that they got to turn such knowledge into helpful info and data. The data that's acquired through the assistance of knowledge mining are often applied into numerous applications like business management, retail and market research, engineering style and scientific exploration.

In the field of knowledge mining, the cluster technique play a really vital role that divide the massive quantity of knowledge into similar sort of cluster on the essential of demand. The partitioning clustering algorithm divides the "n" objects into "k" clusters with maximum intra-class similarity or minimum inter-class similarity. These clusters represent the groups of data and provide the representation of many data objects by fewer clusters. Clustering is very useful technique to deal with the statistical data and unsupervised learning.

Sequential Pattern Mining is an important research content in the field of data mining. It helps to find the sequential pattern in place of frequent pattern among the large number of database items and its most typical application is to find the new useful information about the Customer who have an account in the Bank in which region, which provide the Bank planning to improve the Customer satisfaction or business .

## 2. LITRETURE SURVEY

The data mining provide the many technique and algorithms that uses to extract the knowledge from the database or data repositories. The aim of this literature survey is to search out the two best techniques that efficiently extracted the Customer behavior data on banking data.

### 2.1 Cluster Analysis

Cluster analysis or cluster is that the task of "grouping a group of objects" in such the simplest way that objects within the "same group" (called cluster) square measure additional similar (in some sense or another) to every apart from to those in alternative teams (clusters). it's a main task of explorative data processing, and this can be a standard technique for applied math information analysis employed in several fields, as well as machine learning, pattern recognition, image analysis, data retrieval, and bioinformatics.

## 2.2 Apriori

A large range of distinct single things during a typical dealings database, and their mixtures could type a really immense range of item sets, it's difficult to develop ascendable strategies for mining frequent item sets during a giant dealings info. Agrawal and Srikant(1994)[9] ascertained a motivating downward closure property, known as Apriori, among frequent k item sets: A k-item set is frequent provided that all of its sub-item sets area unit frequent. this suggests that frequent item sets is strip-mined by first scanning the info to find the frequent 1-itemsets, then victimization the frequent 1-itemsets to come up with candidate frequent 2-itemsets, and check against the database to get the frequent 2-itemsets. This method iterates till no additional frequent k-itemsets is generated for a few k[6]. this can be the essence of the Apriori formula.

## 2.4 FP-Growth

FP-growth technique that mines the entire set of frequent item sets while not candidate generation. FP-growth works during a divide-and-conquer method. first, scan of the information derives a listing of frequent things within which things square measure ordered by frequency falling order. Consistent with the frequency-descending list, the information is compressed into a frequent-pattern tree or FP-tree that retains the item set association data. The FP-tree is strip-mined by ranging from every frequentlength-1 pattern (as Associate in Nursing initial suffix pattern), constructing its conditional pattern base (a "sub database", that consists of the set of prefix methods within the FP-tree co-occurring with the suffix pattern), then constructing its conditional FP-tree, and acting mining recursively on such a tree. The pattern growth is achieved by the concatenation of the suffix pattern with the frequent patterns generated from a conditional FP-tree. The FP-growth rule transforms the matter of ending long frequent patterns to sorting out shorter ones recursively and so concatenating the suffix. It uses the smallest amount frequent things as a suffix, giving smart property. Performance studies demonstrate that the strategy well reduces search time.

## 3. PROBLEM STATEMENT

The technique which is use by the previous researchers want some modification like take the large amount of data and leave the multiple scanning of data every time when a new data is occurs.

The problem is:

1. Small data size
2. Multiple scan
3. Large No. of candidate set generation
4. Large No. of Projected Dataset

Small data size means, the data size that handled by the Apriori algorithm is small because of the candidate set generation problem because Apriori algorithm[13] generate the large no. of candidate set generate for every data.

Multiple scan problem is a very critical problem with the Apriori Algorithm if we use the Apriori algorithm then for every search it generate candidate generation set and scan all the set for find the sequential. So every time a no. of scan needs for the find a pattern.

The main problem is large no. of candidate set generation. Apriori algorithm generates the no. of candidate set generation for 1 frequent pattern, 2 frequent patterns, and 3 frequent patterns. It depends on the minimum support.

## 4. PROPOSED WORK

### 4.1.Limitation of the Recent Work

As per the literature survey & recent trends, we tend to already realize the principally partitioning primarily based clustering rule utilized by the association rule mining that are supported the frequent pattern mining technique like Apriori[8].

Apriori may be a classic rule for learning association rules [18]. Apriori is intended to control on databases containing transactions. Different algorithms are designed for locating association rules in knowledge having no transactions, or having no timestamps. Different algorithms ar designed for locating association rules in knowledge having no transactions, or having no timestamps. The aim of the Apriori rule is to search out associations between totally different sets of information. It's generally brought up as "Market Basket Analysis". Every set of information includes a variety of things and is named dealing. The output of Apriori is sets of rules that tell North American country however typically things are contained in sets knowledge of as a result of Apriori rule straightforward to use over that data and straightforward to implementation. however Apriori rule generate the high quantity of candidate and takes the such a lot time to execution and additionally not excellent for the massive quantity of information.

So FP-growth utilized by the various users to search out the frequent pattern from the massive data [8]. as a result of FP-Growth not generate the candidate set and use the FP-tree primarily based structure over knowledge and generate the pattern while not generate any candidate set. FP-Growth scans the information just one occasion within the place of multiple scanning of information like Apriori. Supported Associate in Nursing analyzing an association rule mining rule known as FP tree. A brand new association rule mining rule known as increased FP was given. Because the main disadvantage of FP-Growth is that it's terribly tough to implement as a result of its complicated arrangement. during this FP-tree takes plenty of your time to create and additionally wants a lot of memory for storing the transactions.

FP-Growth approach is predicated on divide and conquers strategy for manufacturing the frequent item sets. FP-growth is especially used for mining frequent item sets while not candidate generation. Major steps in FP-growth isStep1- It first of all compresses the info showing frequent item set in to FP-tree. FP-tree is made victimization two passes over the dataset.Step2: It divides the FP-tree in to a group of conditional info and mines every individually, so extract frequent item sets from FP-tree directly. It include one root labeled as null, a group of item prefix sub trees because the kids of the basis, and a frequent .item header table. Every node within the item prefix sub tree consists of 3 fields: item-name, count and node link where---item-name registers that item the node represents; count registers the quantity of transactions depicted by the portion of path reaching this node, node link links to successive node within the FP- tree. Every item within the header table consists of 2 fields---item name and head of node link those points to the primary node within the FP-tree carrying the item name.

If the information have several same values like gender that have solely 2 instance male or feminine then if we tend to use the frequent pattern then we tend to cannot analysis that during which manner the male or feminine getting the merchandise as a result of each are available the resultant.

So this can be the limitation of the recent work.

## 4.2 Proposed Work

The aim of this thesis is to evaluate the performance of sequential mining over frequent mining by using partitioning clustering technique K-Means. This thesis focus on the best clustering partitioning based technique with sequential pattern mining like prefix. Here we use the K-means for divide the Indian bank data into groups on the basis of region or age group and further apply the sequential algorithm Prefix Span algorithm for search out the sequential pattern over them.

Proposed system is designed to overcome of Frequent Pattern mining Technique and K-means [5] [7] with an advantage hard clustering means. We now that the k means does not deal with the outliers but this problem arise when data have missing values and this problem is not arise with this data and otherwise the k means is one of the best technique for clustering so we use K means because it takes the less execution time in comparison the partitioning technique.

The drawback of the FP-Growth algorithm [8] like generates the 1.FP-tree structure. 2. Very complicated with implementation and provide the Frequent Patterns. Our proposed system provide the sequential pattern with less projected data use the suffix technique with pseudo projection technique that takes the less memory and less execution time and provide the Sequential pattern that helps to find out the customer behavior more perfectly.

## 5. PROPOSED METHOD

### 5.1 Proposed Method

- Step 1: Select Clustering Algorithm K-Means.
- Step2: Input the data file.
- Step 3: Select the output file for Clustered data.
- Step 4: Set the No. of Cluster
- Step 5: The data is divided into clustered data set.
- Step 6: Select the Output file as Input file again.
- Step 7: Select the Improved sequential FP-Growth Algorithm.
- Step 8: Select an output file.
- Step 9: Set the MIN-Support value.
- Step 10: Sequential pattern generate of the clustered data.

### 5.2 K-Means

K-Means is based on the hard clustering. K-means is very commonly used partitioning technique [18] that is mostly used for analyze data and trends in the large amount of data-Means is one of the most unsupervised learning methods. K-Means referred to the Hard clustering means data points belongs from the one cluster never belongs any other cluster. In other word we can say that Partitioning the objects into mutually exclusive clusters (K) is done by it in such a fashion that objects within each cluster remain as close as possible to each other but as far as possible from objects another clusters. K-Means divides the given set of “N” objects into “K” cluster,

where the K is the no. of cluster. Each cluster has a centre point i.e. centroid. All the data points are placed in a cluster having centroid nearest (or most similar) to that data points. After processing all data points, k-means, or centroids, are recalculated, and the entire process is repeated. On the basis of the new centroids, all the data points are bound to the clusters. In the each iteration centroid moves from different data points. This process is continuous until no any centroid move. At last we found the K cluster with N data points. In other words centroids do not move anymore. Finally, this algorithm aims at minimizing an *objective function*, in this case a squared error function. The objective function

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

Where  $\|x_i^{(j)} - c_j\|^2$  is a chosen distance measure between a data point  $x_i^{(j)}$  and the cluster center  $c_j$ , is an indicator of the distance of the  $n$  data points from their respective cluster centers.

#### 5.2.1 Algorithm of the K-Means

- Step1. Choose the K Number of clusters to partition N data objects.
- Step2.Generate K clusters and determines the cluster's center.
- Step3. Assign each object to the cluster to which the object is the most similar; based on the given similarity function.
- Step 4. Update the cluster means (centroid).
- Step5.Repeat steps 3 and 4 until no change occurs in the clusters.

#### 5.2.1 Advantages

- Easy to understand and implement.
- With a large number of variables, K-Means may be computationally fast.
- K-means is based on the exclusive clustering then its produce tighter clusters.

#### 5.2.3 Disadvantages

- Not applicable to categorical data.
- Unable to handle noisy data and outliers.
- It does not work well with clusters when original data have Different size and Different density of data.
- Result and total run time depends upon initial partition

The proposed model is designed for the clustering technique with sequential patterns that overcome the drawbacks of the frequent pattern mining technique or other sequential pattern mining technique like Apriori Based GSP that is the generalized method of Apriori Algorithm.

Here we apply the clustering algorithm k-means at the first step and then generate the sequential pattern apply the sequential pattern algorithm prefix span over them we know that this algorithm already in use with different place but never use this type of data like clustering over bank data to find out the craze of the customer in which sequence they

purchase the account or open an account in the bank and how to improve the business. So following subsections present the details of these steps.

### 5.3 Prefix Span Algorithm

Based on the analysis of the Free Span [9] algorithm, one can see that one may still have to pay high cost at handling projected databases. Is it possible to reduce the size of projected database and the cost of checking at every possible position of a potential candidate sequence? [3] To avoid checking every possible combination of a potential candidate sequence, one can first fix the order of items within each element. Since items within an element of a sequence can be listed in any order, without loss of generality, one can assume that they are always listed alphabetically. For example, the sequence in  $S$  with Sequence id 10 in our running example is listed as  $\langle a(abc) (ac)d(cf) \rangle$  instead of  $\langle a(bac) (ca)d(fc) \rangle$ . With such a convention, the expression of a sequence is unique. Then, we examine whether one can fix the order of item projection in the generation of a projected database. Intuitively, if one follows the order of the prefix of a sequence and projects only the suffix of a sequence, one can examine in an orderly manner all the possible subsequences and their associated projected database. Thus, we first introduce the concept of prefix and suffix

**Definition 1 (Prefix).** Suppose all the items within an element are listed alphabetically. Given a sequence  $a = \langle e_1, e_2, e_3, \dots, e_n \rangle$  (where each  $e_i$  corresponds to a frequent element in  $S$ ), a sequence  $B = \langle e_1', e_2', \dots, e_m' \rangle$  is called a prefix of  $A$  if and only if 1)  $e_i' = e_i$  for ( $i \leq m-1$ ); 2)  $e_m'$  subset of  $e_m$ ; and 3) all the frequent items in  $(e_m - e_m')$  are alphabetically after those in  $e_m'$ .

**Definition 2 (Suffix).** Given a sequence  $A = \langle e_1, e_2, \dots, e_n \rangle$  (where each  $e_i$  corresponds to a frequent element in  $S$ ). Let  $B = \langle e_1, e_2, \dots, e_{m-1} \rangle$  ( $m \leq n$ ) be the prefix of  $A$ . Sequence  $C = \langle e_m, e_{m+1}, \dots, e_n \rangle$  is called the suffix of  $A$  with regards to prefix  $B$ , denoted as  $C = A/B$ , where  $e'' = (e_m - e_m')^2$ . We also denote  $A = B.C$ . Note, if  $B$  is not a subsequence of  $A$ , the suffix of  $_$  with regards to  $A$  is empty.

**Definition 3 (Projected database).** Let  $A$  be a sequential pattern in a sequence database  $S$ . The  $_$ -projected database, denoted as  $S|_A$ , is the collection of suffixes of sequences in  $S$  with regards to prefix  $A$ . To collect counts in projected databases, we have the following definition:

**Definition 4 (Support count in projected database).** Let  $A$  be a sequential pattern in sequence database  $S$ , and  $B$  be a sequence with prefix  $A$ . The support count of  $B$  in  $A$ -projected database  $S|_A$ , denoted as  $\text{support}_{S|_A}(B)$ , is the number of sequences in  $S|_A$  such that  $B$  is a subset of  $A.C$ .

#### 5.3.1 Pseudo projection

The above analysis shows that the major cost of Prefix Span is database projection [10], i.e., forming projected databases recursively. Usually, a large number of projected databases will be generated in sequential pattern mining [15]. If the number and/or the size of projected databases can be reduced, the performance of sequential pattern mining can be further improved. One technique which may reduce the number and size of projected databases is pseudo projection. The idea is outlined as follows: Instead of performing physical projection, one can register the index (or identifier) of the corresponding sequence and the starting position of the projected suffix in the sequence. Then, a physical projection of a sequence is replaced by registering a sequence identifier and the projected position index point. Pseudo projection reduces the cost of

projection substantially when the projected database can fit in main memory. This method is based on the following observation: For any sequence  $s$ , each projection can be represented by a corresponding projection position (an index point) instead of copying the whole suffix as a projected subsequence.

#### 5.3.2 Algorithm

- Algorithm Input parameters :
  1. A sequential database  $S$
  2. Minimum support threshold.
- Algorithm Output parameters:
  1. A bunch of sequential patterns.
- Subroutine: Prefix Span ( $P, L, S|P$ ).
- Parameters used in the algorithm:
  2.  $P$  is denoted sequential pattern.
  3.  $L$  is denoted the length of  $P$  (sequential pattern).
  4.  $S|P$  is denoted the  $P$ -projected database. Where  $P \neq \text{set of patterns}$ . Otherwise the sequence database  $S$ .
- Call Prefix Span ( $\langle \text{set of patterns} \rangle, 0, S$ ).
- Scan  $S|P$  once,
  1. find the set of frequent items  $b$  such that:
  2.  $b$  can be assembled to the last element of  $P$  to form a sequential pattern; or
  3.  $\langle b \rangle$  can be appended to  $P$  to form a sequential pattern.
- For each frequent item  $b$ :
  1. Append it to  $P$  to form a sequential pattern  $P$ .
  2. output  $P$
- For each sequential pattern  $P$ 
  1. Construct  $P$ -projected database  $S|P$ .
- Call Prefix Span ( $P, L+1, S|P$ ).

## 6. PROPOSED DESIGN

This paper develops a method to design Banking information, informed by Bank associations observed in the same region of bank customers. Clustering and Sequential Pattern are used to identify customer behavior. We can easily predict the how to increases the customer also Table 1 shows the customer data. The details can be fetched from Bank database.

	A	B	C	D	E	F
1	Region	Age Group	Gender	Marital Status	Have a car	
2	INTER_CITY	30	FEMALE	MARRIED	YES	
3	RURAL	60	FEMALE	MARRIED	NO	
4	SUBRURAL	40	MALE	MARRIED	YES	
5	INTER_CITY	30	FEMALE	UNMARRIED	YES	
6	TOWN	30	FEMALE	MARRIED	NO	
7	SUBRURAL	50	FEMALE	MARRIED	NO	
8	TOWN	60	FEMALE	MARRIED	NO	
9	INTER_CITY	30	FEMALE	MARRIED	NO	
10	TOWN	30	FEMALE	MARRIED	NO	
11	INTER_CITY	30	MALE	MARRIED	NO	
12	INTER_CITY	30	MALE	MARRIED	NO	
13	SUBRURAL	30	MALE	MARRIED	YES	
14	RURAL	20	MALE	UNMARRIED	YES	
15	TOWN	60	MALE	MARRIED	YES	
16	INTER_CITY	30	MALE	UNMARRIED	YES	
17	INTER_CITY	40	MALE	MARRIED	YES	
18	TOWN	30	FEMALE	MARRIED	NO	
19	INTER_CITY	40	MALE	MARRIED	YES	
20	INTER_CITY	50	MALE	MARRIED	NO	
21	RURAL	40	FEMALE	UNMARRIED	YES	
22	SUBRURAL	20	MALE	UNMARRIED	YES	
23	TOWN	30	FEMALE	MARRIED	NO	

Ashish Goel On (Customer Purchasing Behaviour using Sequential Pattern Mining Technique T...)

Choose an algorithm: Clustering KMeans

Choose input file bank-data 11.CSV Input File

Set output file after clustering.bt Output File

Choose K: 4 (e.g. 4)

☒ Open output file when the algorithm terminates

Run algorithm

```

===== KMEANS - STATS =====
Total time ~: 33 ms
Max memory 4 0010986328125 mb
=====
  
```

[illegible]

Ashish Goel On (Customer Purchasing Behaviour using Sequential Pattern Mining Technique T...

Choose an algorithm: Sequential\_Pattern\_Mining

Choose input file after clustering.bt Input File

Set output file after sequential.bt Output File

Choose minsup (%): 3 (e.g. 0.4 or 40%)

Enter the Region INTER\_CITY, TOWN, RURA...

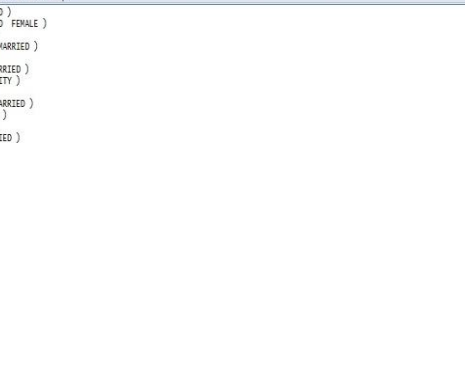
☒ Open output file when the algorithm terminates

Run algorithm

```

===== Sequential Pattrn Algorithm =====
Total time ~ 28 ms
Sequential sequences count : 12
Max memory (mb) : 2 890419006347656212
=====

```



```

after sequential - Notepad
File Edit Format View Help

(MARRIED )
(MARRIED FEMALE )
(MALE )
(MALE MARRIED )
(NO )
(NO MARRIED )
(INTENSITY )
(YES )
(YES MARRIED )
(FEMALE )
(30 )
(UNMARRIED )

```

Ashish Goel On (Customer Purchasing Behaviour using Sequential Pattern Mining Technique T...)

Choose an algorithm: Frequent\_Pattern\_Mining

Choose input file after clustering.bt Input File

Set output file after frequent.bt Output File

Choose minsup (%): 4 (e.g. 0.4 or 40%)

Enter the Region  INTER\_CITY, TOWN, RURA...

☒ Open output file when the algorithm terminates

Run algorithm

===== FP-GROWTH - STATS =====  
 Frequent itemsets count: 94  
 Total time ~ 47 ms  
 =====

28

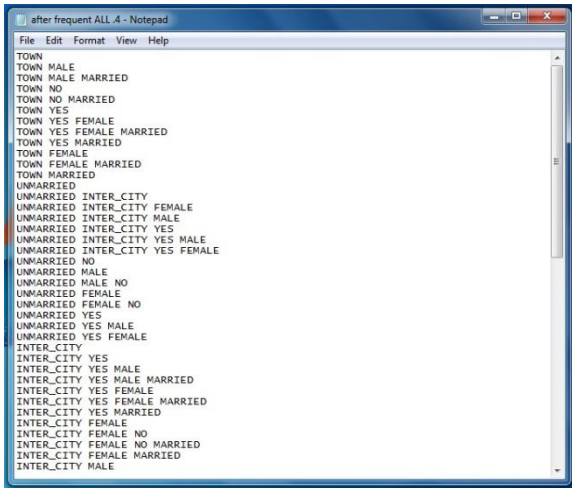


Figure 6- Output of Frequent Pattern Mining Algorithm after Processing Under All Regions

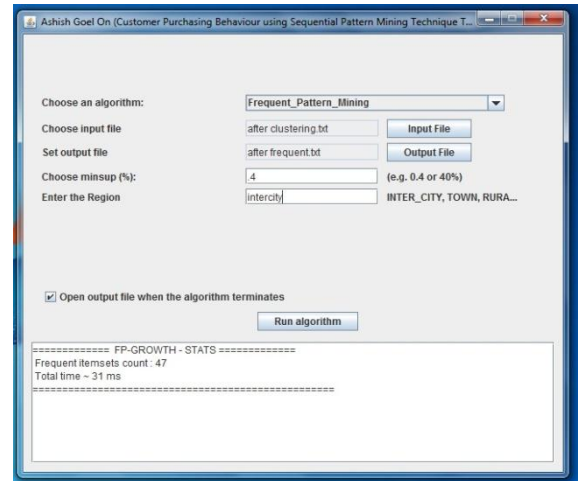


Figure 9- Frequent Pattern Mining Algorithms with Min. Support 40% under Intercity

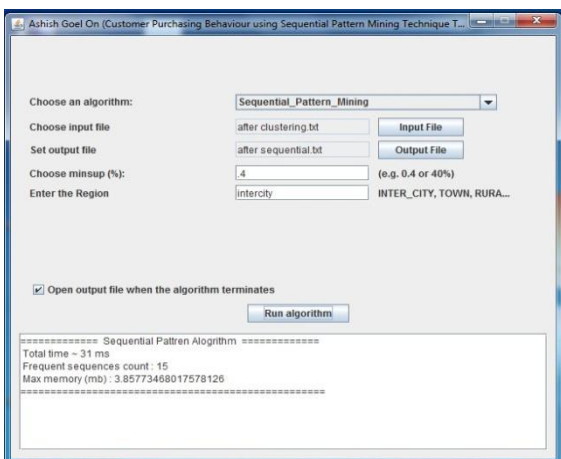


Figure 7- Sequential Pattern Mining Algorithms with Min. Support 40% under Intercity

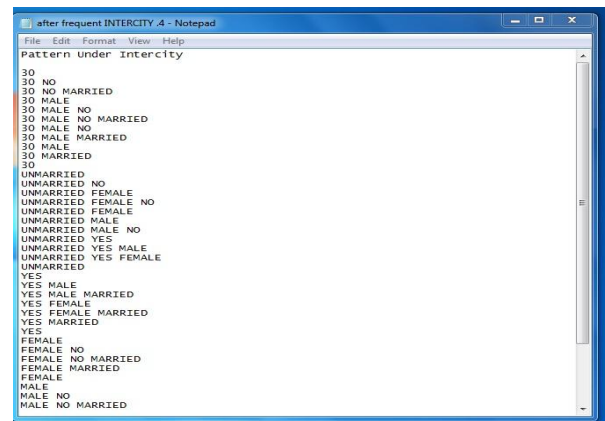


Figure 10- Output of Frequent Pattern Mining Algorithm after Processing Under Intercity

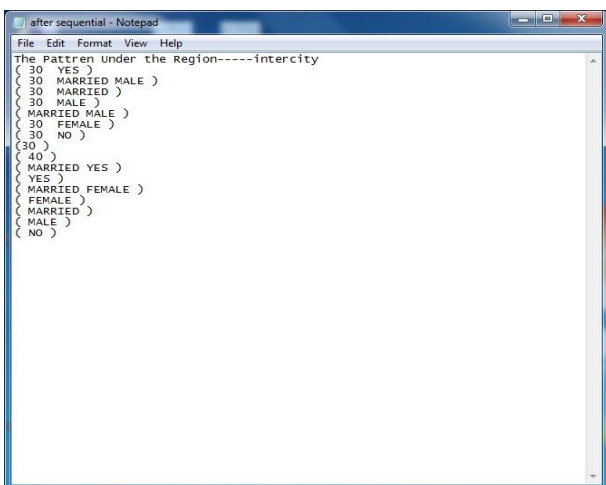


Figure 8- Output of Sequential Pattern Mining Algorithm after Processing Under Intercity

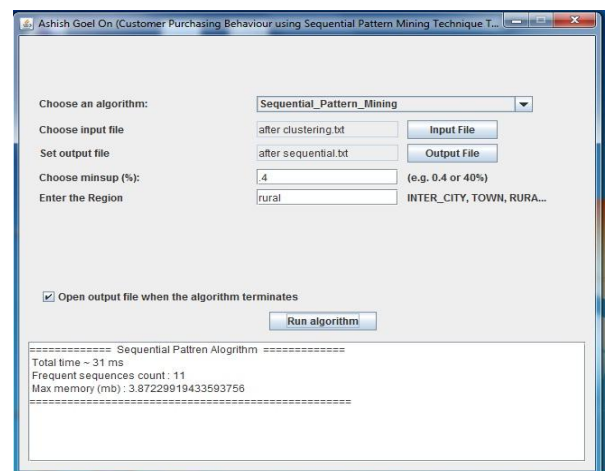


Figure 11- Sequential Pattern Mining Algorithms with Min. Support 40% under Rural



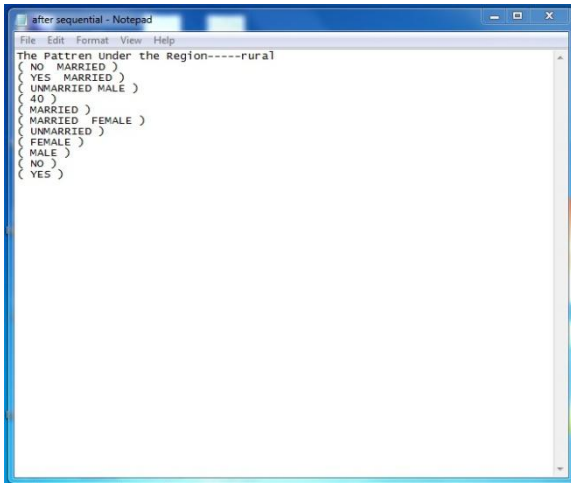


Figure 12- Output of Sequential Pattern Mining Algorithm after Processing Under Rural

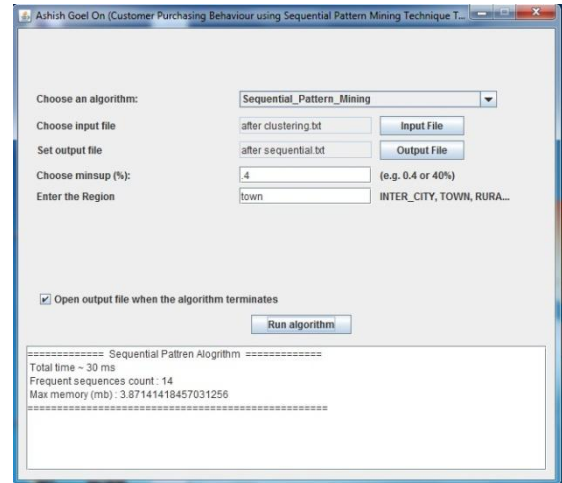


Figure 15- Sequential Pattern Mining Algorithms with Min. Support 40% under Town

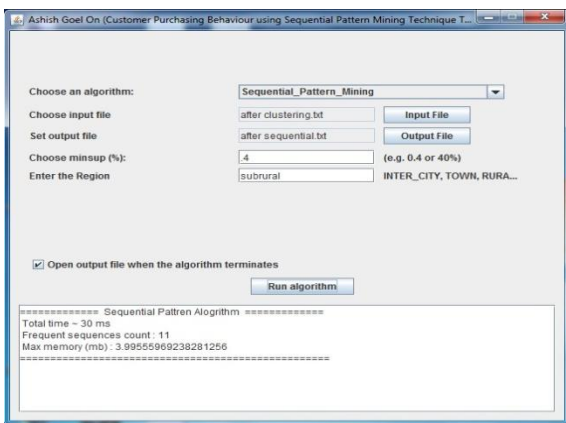


Figure 13- Sequential Pattern Mining Algorithms with Min. Support 40% under Subrural

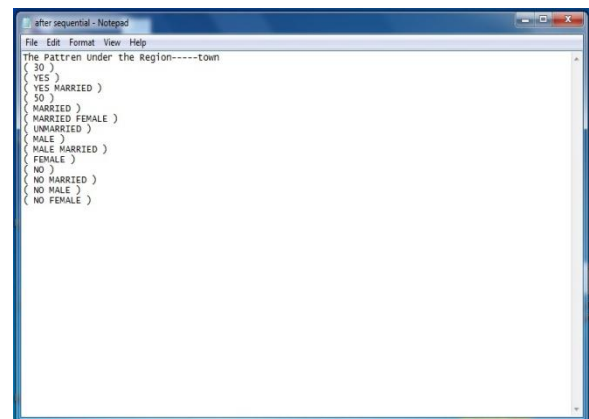


Figure 16- Output of Sequential Pattern Mining Algorithm after Processing Under Town

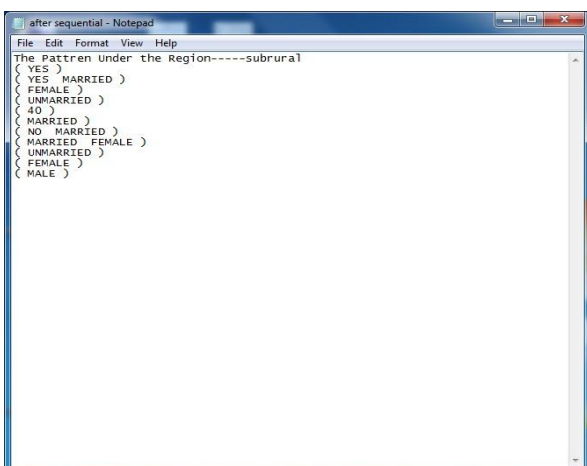


Figure 14- Output of Sequential Pattern Mining Algorithm after Processing Under Sub rural

Here we compared the two algorithm FP-Growth (Frequent Pattern Based) and Prefix Algorithm (Sequential Pattern Based) over the Indian bank data to find out the interesting pattern over customer purchasing behavior in a bank. Mostly Frequent pattern based algorithm used in this manner like Apriori and FP-growth that provide the only frequent patterns and does not takes the right information because this type of data have many similar type of instance like male-female, married-unmarried. So in the frequent item set all data is overlapped and we don't get the right information but in the sequential pattern we get only sequential data means second instance always follow the first. So Sequential Pattern does not repeat the same pattern and extracts the use full information.

## 7. CONCLUSION & FUTURE WORK

In this paper we already analysis the customer behavior over bank data to search out the sequential pattern to improve the performance of the bank to suggest his employee, agents which type of customer mostly open the account in the bank or which type of customer remain & not interested to open the account in the bank as per the bank data. So agents or employee of the bank target those customers to improve the

performance of bank. So data mining methodology has tremendous contributions for researchers to extract the hidden knowledge and information. The research described in this paper also identified significant sequential pattern within each segment. Sequential pattern can be used to motivate customers to increase their purchases and keep loyal to the organization. The behavior of customers can be easily identified.

Here we use the Frequent Pattern Algorithm and Sequential Pattern Algorithm Prefix span, the Frequent Pattern Algorithms provide the Frequent patterns This information is useful but when a data have many similar type of items in each row then we cannot get the right information about the customer because they select the common frequent items that is larger than the sequential. So sequential Pattern mining technique means Prefix algorithm only provide the sequence item that always done is a sequence. So, based on our view, the implication of this method is far beyond yet another efficient sequential pattern mining algorithm. It demonstrates the strength of the pattern growth mining methodology since the methodology has achieved high performance in both frequent-pattern mining and sequential pattern mining. Moreover, our discussion shows that the methodology can be extended to mining multilevel, multidimensional sequential patterns, mining sequential patterns with user-specified constraints, and so on. Therefore, it represents a promising approach for the applications that rely on the discovery of frequent patterns and/or sequential patterns.

## 8. FUTURE WORK

There square measure several attention-grabbing problems that require to be studied additional, like mining closed and largest successive patterns, mining approximate successive patterns, and extension of the tactic toward mining structured patterns. Especially, the developments of specialized successive pattern mining strategies for explicit applications, like deoxyribonucleic acid sequence mining which will admit faults, like permitting insertions, deletions, and mutations in deoxyribonucleic acid sequences, and handling industry/engineering successive method analysis square measure attention-grabbing problems for future analysis.

## 9. REFERENCE

- [1] A.K. Jain, June, 2010, "Data Clustering: 50 Years Beyond K-Means", Volume No. 31, Issue 8, pp: 651-666.
- [2] Abdullah Al-Mudimigh, 2009, "Efficient implementation of data mining: improve customer's behavior", pp 7-10.
- [3] Ayres, J., 2002, "Sequential pattern mining using a bitmap representation", In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [4] Agrawal R and Srikant R, 1995, "Mining Sequential Patterns", Volume No. 9.
- [5] Bharati R. Jipkate, 2012 "A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms", pp: 2250-3005.
- [6] Dawn E. Holmes, 2012, "Data Mining Techniques in Clustering, Association and Classification, Data Mining", Volume 23, pp: 1-6.
- [7] Gianfranco Chicco, MAY (2006), Comparisons Among Clustering Techniques for Electricity Customer Classification", Volume No. 21, Issue No. 2, pp: 933-940.
- [8] Han J., Dong G., 2000 "Free-span: Frequent pattern-projected sequential pattern mining", pp: 355-359.
- [9] Horng-Jinh Chang, 2007 "An anticipation model of potential customers' purchasing behavior based on clustering analysis and association rules analysis" Volume No. 32, pp: 753-764
- [10] Jian. Pei, 2001, "Prefix-Span: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth".
- [11] Lijuan Huang, DECEMBER (2011), "Analysis on E-consumers' Purchasing Behavior Based on Data-driving Model", Volume No. 6, Issue No. 12, pp: 1713-1717.
- [12] M. Zaki, 2001 "SPADE: An efficient algorithm for mining frequent sequences, Machine Learning,.
- [13] P. Isakki alias Devi, (2012), " Analysis of Customer Behavior using Clustering and Association Rules " Volume 43- No.23, pp 0975 – 8887.
- [14] Rakesh Agrawal and Ramakrishnan Srikant, (1994) "Fast algorithms for mining association n rules in large databases". Volume No. 20, pp: 487-499.
- [15] Srikant R. and Agrawal R., (1996), "Mining sequential patterns: Generalizations and performance improvements", Volume No.5.
- [16] Shruti Aggarwa, 2008 "Comparative Study of Various Improved Versions of Apriori Algorithm".
- [17] Shalini S Singh, 13-14 May 2011 "K-means v/s K-medoids: A Comparative Study".
- [18] T. Velmurugan, 2011, "A Survey of Partition based Clustering Algorithms in Data Mining: An Experimental Approach", Volume No. 10, Issue No .3, pp: 478- 484.
- [19] V.Thanuja, 2011, "Applications of Data Mining in Customer Relationship Management", Volume No. 2, Issue 3, pp: 423-433