Journal of Theoretical and Applied Information Technology
28th February 2018. Vol.96. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USING K-MEANS ALGORITHM AND FP-GROWTH BASE ON FP-TREE STRUCTURE FOR RECOMMENDATION CUSTOMER SME

[1]MUHAMMAD ALI SYAKUR, [2]BAIN KHUSNUL KHOTIMAH, [3]EKA MALA SARI ROCHMAN [4]BUDI DWI SATOTO

[1,2,3,4]Faculty of Engineering, University of Trunojoyo Madura

E-mail:[1]syakurali@yahoo.co.id, [2]bain@trunojoyo.ac.id, [3]ekamalasari3@gmail.com, [4]budids@yahoo.com

ABSTRACT

The market basket has been finded patterns of purchase customer in SME. Purchase patterns can help to make recommendations and product promotions. This research used K-Means algorithm for sales data clustering and uses FP-Growth Algorithm to know the relation of each cluster. K-Means clustering to classify customer data based on the same attribute, then determined the relationship between patterns in each group with FP-Growth Algorithm. K-Means to do customer segmentation based on background, customer characteristic and level of purchasing power. To facilitate the analysis of customer relationships with products purchased, then each cluster profilling customer will be processed data record by FP Growth to know the relevance of goods purchased. The research presents a discussion of the comparison of time complexity between FP-Growth algorithms and Apriori Algorithms. This research would been done the development and application of the use of Trees ie FP-Tree (Frequent Pattern Tree). They are an extension of the use of Trees in the data structure. FP-Tree is used in conjunction with the FP-Growth algorithm to determine the frequent itemset of a database, in contrast to the a priori paradigm of scanning the database repeatedly to determine the frequent itemset. In this study, the number of transactions with many items of goods and consumer purchasing power are varied, grouped first by using K-Means algorithm, cluster results formed into several groups including five customer groups based on customer profile. The result of the test is average on minsupp = 60 and minconf = 40, so the average processing time is 957 ms.

Key Word: *K-Means, Fp-Growth, FP-Tree, SME, Profilling Customer, Pattern*

1. INTRODUCTION

SME (Small Medium Enterprises) are companies that collaborate with the market, so as to be able to compete should be able to understand the behaviour of the client in the purchase [1]. One technique is to use the data processing Frequent Pattern Mining Engineering and using Sequential Patterns Mining Engineering grouping so that it can analyse customer behaviour [2]. Apriori method to analyse many sets of candidates set to get information each group set in repeatedly. Utilization of large amounts of data contained in an information system to support decision-making activities is not enough to simply rely on operational data as recording and materials to make a report, but also required the excavation potential that exists in the transaction data to be processed using specific algorithms, the association rules Apriori algorithm is a method that can be used to explore the potential of data related to each other [2-3].

Selection of sales transaction was used as objects to analyse the data of stored and to be extracted useful information for the company. Receipt of any listed types and price of goods purchased the number of items purchased, the total expenditure, and the date and time of the transaction. The sale SME of the potential analyse need of association between the types products that are related to the ability of purchasing power [4]. The association carried out using Apriori algorithm to seek the establishment of item sets of each cluster type of goods to calculate how much the support cluster type of goods purchased simultaneously with a cluster other items and how much value the certainty or the value of confidence cluster type of goods purchased together with cluster type of goods both generally and more based cluster purchasing power. Values support and confidence generated from each cluster is evaluated to obtain the greatest value that will serve as the conclusion of research conducted. The use apriory

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

algoritma is equal Market basket analysis to explaint system in retail market as recommendation the placement of goods, strategic of promotions, and profit of the supermarket. Market Basket Analysis utilizing the sales transaction data to be analyzed and then find the pattern of the items together in a transaction. One of the benefits of the Market Basket Analysis designing sales or marketing strategy by utilizing the sales data that exist in the company, namely: 1. By changing the layout of the store, placing items in adjacent goods are often purchased together by the consumer. 2. Giving discounts to items of goods that are rarely bought and expensive [5-7]. The implementation of e-commerce then clustering techniques required to classify products based on customer tastes. Grouping these customers based on several parameters including the number of products, quality, price, rating, and type. There are many variety intuitions on measurement objects can be similar with usage many different clustering algorithms, show the mark clusters from data. The approaches clustering show similarity of spatial closeness data. That clustering is a smaller Euclidean distance between two points which shown data is relatively similar [6].

Usage data mining contain data processing techniques with clustering techniques is a function to classify data. Clustering techniques in e-commerce applications to classify a particular problem based on the trend, which is to classify the customer data [8]. The research use K-Means clustering which determined number of clustering as shown performance of this algorithm [9-10]. With clustering, a set of very large datasets, grouped into several groups based on similarities owned by each - each member of the group. In a cluster hierarchy starts with making m clusters where each cluster is comprised of one object and ends with a cluster where members are m objects. At each stage of the procedure, a cluster in combination with one cluster to another. In hierarchical clustering, the distance of each - each object is calculated by any other object that will be found a couple objects located nearby. So, that each object will be paired with an object or group of objects other closest distance. Thus, hierarchical clustering method to calculate the distance similarity between objects.The combined use of K-Means method with other methods to gain knowledge of each cluster identification relationship [11]. Strehl and Ghosh (2003) have described a method cluster for high-dimensional data requiring analysis to represent clusters of each member. The researchs have developed a clustering algorithm for high-dimensional data processing a long time. These data have some attributes so that the difficulties in assessing the similarity of objects significantly [12]. They were applied the K-Means method with the concept of hierarchical clustering to improve the quality of the cluster [13].

Association Rule Mining methods are also used to detect the similarity of products sold and those that occur in one group [14]. The algorithm used in this study to discover the rules of associative between combinated of items is the algorithm FP-Growth algorithm is very efficient in the search for frequent itemset. FP-Growth takes a different approach from the paradigm that has been commonly used, namely a priori algorithm. This algorithm stores information about frequent itemset in the form prefix-Tree structure or often called FP Tree. In the FP Tree formed can compress the data transactions that have the same item, so as to reduce the scan database repeatedly in the mining process and can happen faster [15-16]. In the study will be discussed regarding the application of FP Growth algorithm to determine the association between the products on transaction data minimarket. It can be seen that the line is obtained from association rule pattern can be used for product promotion strategy. In the study will be discussed regarding the application of FP Growth algorithm to determine the association between the products on transaction data. It can be seen that the line is obtained from association rule pattern can be used for product promotion strategy [17].

The problem in this research is how to classify data of purchasing of goods into certain group based on consumer characteristic. Where Objects are distributed in a particular group by taking into account the degree of correlation between cluster member targets and the variables that affect cluster members. The variables used are Age, Profession, Income, Education, Quality of Batik and Gender. Cluster analysis using FP Growth in Market Analysis to find knowledge of customer habits in shopping batik product. The combination of these methods is expected to be able to analyze the needs of consumers who later become recommendations in producing batik. This study proposed a hybrid method of clustering and K-Means Assotiation rule to make predictions and simulation grouping the customer data. The system can provide knowledge of existing customer data set. Results using frequent pattern is in a group to get a pattern - a pattern that is often encountered. The existing pattern used very

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

effectively to provide recommendations to customers. Cluster analysis on data sets that have the same (similar) with other data in the same cluster and have nothing in common (dissimilar) to the other objects in the different clusters. By using clustering method, the object - the same object can be classified into one for easier identification.

2. LITERATURE REVIEW

The main purpose of this study is to improve customer satisfaction and revenue increase super market TS. The following algorithm used in this study is the combination of K-Means the algorithm to produce cluster analysis and Association Rule fp Growth for mining Market Basket analysis is taken from SMEs. In the research transaction was observed from the purchase data of batik based a copy of an invoice containing items purchased by different customers. Improve by K-Means hybrid model with FP-Growth is used to improve the method and facilitate the analysis as needed.

Fp Growth based on the new Tree can save a lot of memory in terms of noise reduction. In this study can reduce scanning database scanned. test results on the test dataset and found that the techniques developed are more efficient and provide more clear correlation between items. The concept of dynamic tree restructuring can achieve an increased frequency a prefix tree structure with a single pass to reduce the mining time. The researchers concluded, the use of tree restructuring achieved remarkable performance advantages in terms of overall runtime. The proposed a counter-defeating algorithm of FP-Growth by taking the same path repeatedly and failing to release the processed node just in time. This requires a solution by adopting a strategy to remove the FP tree and releasing the node in order to reduce the algorithm's embedded memory, thus increasing the efficiency of the algorithm [24].

2.1 K-Means Clustering Algorithm

K-Means (MacQueen, 1967) is one of the unsupervised learning algorithms and simplest way to solve the problem of clustering. This procedure follows the simple and convenient way to define a set of specific data through a certain number of clusters k predetermined. A set of vectors will be x_j, **j = 1, … n,** divided into groups *i* (G_i) for i=1,… *c*. The cost function is based on the Euclidean distance between vectors vector xk in group j and c cluster centers. This procedure follows the simple and convenient way to define a set of specific data through a certain number of clusters assumes k cluster that was previously set. The main idea is to define the centroid *k*, one for each cluster. The next step is to take every point include in a particular data set to connect it to the nearest centroid. If, *k* is new centroids as the points the same data set and the nearest new centroid. They can know that the centroid k change their location step by step until there are no more changes were made [18].

$$j = \sum_{i=1}^{c} j_i = \left(\sum_{k,x_k} \in G_i \ ||x_k - \ c_i||^2\right) \quad (1)$$

$j_i = \sum_{k,x_k} \in G_i \ ||x_k - c_i||^2$ is the cost function in single group divided groups defined by the membership matriks. The groups are defined by the membership matrix, binary *cxn*, *U*, in which elements u_{ij} dalah 1 if the point data *j*, and x_j belongs to a class i and 0 or vice versa. Once, the cluster centers c_i determined by minimizing uij accordance with equation:

$$1, if \ \left||x_j - c_i|\right|^2 \leq \ ||x_j - c_k||^2 \ , for \ each \ k \ \neq \ i \quad (2)$$

$$0, \qquad otherwise$$

Which means that x_j belongs to a class *i* if c_i is the closest center among all centers. Conversely, membership matrix set, uij set, then the optimal center c_i which minimizes the equation is the average of all vectors in the group *i*.

$$c_i \frac{1}{|G_i|} \sum_{k,x_k \in G_i}^{N} X_k \ ,$$

where $|G_i|$ is value from

$$G_i, or \ |G_i| = \sum_{j=1}^{n} u_{ij} \quad (3)$$

The algorithm was presented with a collection of data $\mathbf{x}_i = \mathbf{1,\ldots, n}$; then determine the center of the cluster C_i and membership matrix **U** iteratively [19].

2.2 Association Rule

Association rule is a data mining process to determine all the associative rules that meet the minimum requirements minimum support (minsup) and minimum confidance (minconf) on a database. Both of these conditions will be used for interesting association rules in accordance with the limits defined, namely minsup and minconf [7,14]. Association Rule Mining is a procedure to look for relationships between items in a dataset. Starting with the search for frequent itemset, namely the combination that most often occurs in an itemset and must meet minsup. This phase will be conducted searches combinations of items that meet the minimum requirements of the value of the support in the database. The calculation of the value

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of support an item A and item A and B can be obtained by the following formula [14].

$$Support\ (A) = \frac{Number\ of\ transactions\ containing\ item\ A}{Total\ transaction} \quad (4)$$

$$Support\ (A,B) = P(A \cap B)$$
$$P(A \cap B) = \frac{Number\ of\ transactions\ containing\ item\ A}{Total\ transaction} \quad (5)$$

After all frequent items and large itemets are obtained; you can find the minimum confidence (minconf) condition by using the following formula:

$$Confidence(A \rightarrow B) = P(A|B)$$
$$P(A|B) = \frac{Number\ of\ transactions\ containing\ A\ and\ B}{Number\ of\ transactions\ containing\ A} \quad (6)$$

2.3 FP Growth

The pattern of association of the functionality associated with the data transactions such as e-commerce activities for extracting data. The pattern of association will provide an overview of a number of attributes, or certain properties that often appear together in a given data set. Paradigm priori developed by Agrawal and Srikan (1994), which is Apriori Heuristic: Every patterns with long pattern k that does not often appear (not frequent) in a data set, the pattern of length (k + 1) containing sub k pattern will not often appear also (not frequent). The basic idea of this a priori paradigm is to find the set of candidates to the length (k + 1) of a set of frequent patterns of length k, then match the number of occurrences of these patterns with the information contained in the database [8]. Apriori algorithm will scan database repeatedly, especially if the amount of data is large enough. So, the FP-Growth algorithm which only requires twice scans the database to determine frequent itemset. The data structure used to find frequent itemset with FP-Growth algorithm is an extension of the use of a prefix Tree, commonly called is FP-Tree.

FP-Growth algorithm can directly extract frequent itemset of FP-Tree that has been formed by using the principle of divide and conquer. In the second part will discuss the process of formation of a set of data FP-Tree transaction. FP-Growth is one of the alternative algorithm that can be used to specify the data set that appears most frequently (frequent itemset) in a data set. FP-Growth takes a different approach from the paradigm that has been frequently used, that paradigm apriori [20].

2.4 FP-TREE Algorithm

FP-Tree is a compressed data storage structure. FP-Tree is built by mapping each transaction data into every particular path in FP-Tree. Every transaction is mapped, there may be transactions that have the same item, and the path may be overwritten. The more transaction data that has the same item, then the compression process in the FP-Tree data structure will be more effective [21]. FP-Tree requires twice the scanning of transaction data proven to be very efficient. Let $\mathbf{a_{1...n}} = \mathbf{\{a_1, a_2, ..., a_n\}}$ be a collection of items. Each transaction database = $\mathbf{\{T_1, T_2, ..., T_n\}}$, where T_i (i € [1..n]) is a set of transactions containing items n. Whereas support is the counter of the frequency of occurrence of transactions containing a pattern. A pattern is said to occur frequently (frequent pattern) if the support of the pattern is not less than a constant ξ (minimum support threshold) that has been defined previously. The problem of finding frequent patterns with minimum support threshold support count ξ is what FP-Growth attempts to solve with the help of FP-Tree Structure.

The FP-Tree development stage with FP-Growth algorithm consists of a set of transaction data to search for significant frequent itemsets. FP-Growth algorithm is divided into three main steps, namely [22]:

1. Conditional Pattern Base Generation Phase
   It is a subdatabase containing a prefix path (a prefix path) and a suffix pattern. Generation of conditional pattern base obtained through FP-Tree that has been built before.
2. Conditional FP-Tree Generation Stage.
   This stage, the support count of each item is summed, and then each item that has a support count is greater than the minimum support count ξ.
3. Search on frequent itemset
   If the Conditional FP-Tree is a single path, then it is a frequent itemset by combining items for each conditional FP-Tree. If it is not a single track, then a recursive Fp-Growth generation is performed.

2.5 Hybrid K-Means clustering dan Algortma FP Growth

Apriori algorithm is a method that can be used to explore the potential of data related to each other. Hybrid model K-Means clustering and Fp-Growth provided an overview of the number of transactions

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as evidenced by the large number of receipts every day that affect the acceptance. K-Means for profiling customer and Fp-Growth for product purchased. Attributes are grouped by the same attribute and the relationships between patterns in each group are related to what products they buy. This study will analyze the potential of every receipt of connection or association between the types of goods that one with another type of goods that are related to the ability of purchasing power. The association carried out using Apriori algorithm to seek the establishment of two itemsets or three itemsets of each cluster type of goods to calculate how much the support cluster type of goods purchased simultaneously with a cluster other items and how much value the certainty or the value of confidence cluster type of goods purchased together with a cluster of other types of goods both in general and by cluster purchasing power. Values support and confidence generated from each cluster is evaluated to obtain the greatest value that will serve as the conclusion of research conducted.

System design consists of several stages. First, clustering the input data to cluster using K-Means input data member and data criteria that have been determined. Second, do the association rule process. After the data has been grouped by using K-Means, then Fp-Growth method of data mining to get the rule of relationship between the products that have been sold. When scanning, If the number of data frequencies calculated is less than the frequent item set and the value of the relationship strength is measured by the lift ratio<1 then the item or combination of items will not be included in the next calculation. If it meets the result of rule can be determined.

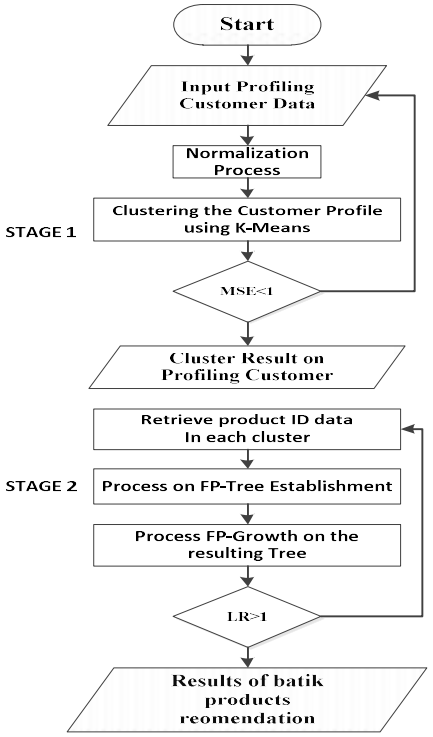

Figure 1: Flowchart Hybrid Model K-Means Clustering and FP Growth

Figure 1. The data will be processed with normalization for the process of K-Means method, namely member data and criteria data. From the normalization result will be processed cluster using K-Means method. The result of K-Means will be analyzed for each cluster to know the active buyer of the transaction. The result of the combination of K-Means and Fp-Growth contains a combination of batik purchase transaction, it are systems rule for recommendation. Determine the number of clusters and cluster centers randomly, data taken from the normalized customer profiling data.

1. Calculate the minimum distance of each object to the cluster center using Euclidean distance.
2. Grouping objects into clusters based on minimum distance.
3. The calculation will stop if the resulting cluster is the same as the previous cluster result.
4. The next process uses fp-Growth algorithm by determining minimum support and minimum confident.
5. The process is done by forming k-itemset from the data that has been entered before with the item that will be formed exceeds the minimum support.
6. The k-itemset forming process will stop if no items have met the minimum support. Then

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

after the last k-itemset is formed the process of generating minimum confidence.

7. Rule that has the highest confidence called strong association means the rule that has the relationship or the most powerful association.

2.6 Lift Ratio

Elevator ratio is used to evaluate the strength of an association rule. Lift ratio is the ratio between the confidence of a rule and the benchmark confidence value. Benchmark confidence is the ratio between amount of all items consequent to the total number of transactions. Benchmark confidence and lift ratio formula according to equation [23].

$$Benchmark\ Confidence = \frac{N_c}{N} \quad (7)$$

$$Lift\ Ratio = \frac{Confidence\ (A,C)}{Benchmark\ Confidence\ (A,C)} \quad (8)$$

With variables Nc = Number of item transactions in consequent, N = Number of database transactions. If the value of the lift ratio is greater than 1, then it indicates the benefit of the rule. If the value of the lift ratio is higher, the association strength increases. Assessing the accuracy of association rule is obtained by comparing the association rule of the test data with the test results; it is expressed by the equation:

$$Error = \frac{|Support\ (test\ data - training\ |}{Support\ training\ data}\ x\ 100\% \quad (9)$$

3. RESULT AND DISCUSSION

Data collection techniques are carried out by taking the market basket analysis dataset. Data consists of two kinds, namely customer profile data and transaction itemsset. Customer profile data consists of 6 variables and 100 variations of batik. Sales transaction data consists of 2000 records taken for 1 month.

3.1 K-Means clustering process on customer profiling data

Profiling data are category data, so it requires a transformation process. The process of transforming the data will transform the variables containing the categories into numeric, by calculating the frequency of data that appears by sequencing the highest to lowest frequency for data initialization. Sample some categories of data that have been transformed according to the Table 2 and Table 3.

Table 1: Sampel Data Profiling of Customer

No	User Code	Age	Profession	Income (x) / (million)	Education	Quality of Batik	Gender
1	Id 1	25	Employee	x≤2	High School	3	Female
2	Id 2	47	Teacher	5≤x≤10	S1	1	Male
3	Id 3	60	Pension	2≤x≤5	S1	5	Female
4	Id 4	70	Entrepreneur	x≥10	Junior High School	6	Female
5	Id 5	56	Farmer	x≤2	primary school	4	Male
6	Id 6	63	Government Employees	5≤x≤10	S1	2	Female
7	Id 7	48	Trader	x≥10	Primary School	6	Female
8	Id 8	20	Studen	x≤2	High School	6	Male
9	Id 9	29	Employee	5≤x≤10	D3	3	Female
10	Id 10	42	Teacher	5≤x≤10	S1	1	Female

Table 2: Process of Data Transformation with Age Group

No.	Age	Frequency	Initials
1	15-25	67	4
2	25-35	139	3
3	35-45	178	2
4	45-55	182	1
5	55-65	35	5
6	x≥65	10	6

Table 3: Process of Data Transformation with Customer Income

No.	Income (x) / (million)	Frequency	Initials
1	x≤2	67	2
2	2≤x≤5	209	4
3	5≤x≤10	170	3
4	x≥10	54	1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4: The Results MSE on Each No. Cluster

Iterasi	Clusters 2	Clusters 3	Clusters 4	Clusters 5
100	3	17	14	16
500	7	32	37	24
1000	22	76	64	38
1500	27	121	97	55
2000	65	225	173	37
MSE	0.748	0.524	0.367	0.425

Determine the cluster center and test it using a different number of clusters. Then calculate the distance of each data to the center of the cluster by using the Euclidean distance and Group data into clusters with minimal distances. After all the data is placed into the closest cluster, then recalculate the new cluster center based on the average member present in the cluster. Setelah melakukan proses K-Means maka melakukan analisa cluster dengan menggunakan MSE (Means Square Error) sesuai pada table 4.

The data yields four clusters with the lowest MSE value, it can be concluded that cluster 4 has the highest cluster homogeneity level. Based on MSE results also shows that cluster 4 is the most optimal compared to other clusters. The system accuracy test uses data training as much as 2000 data with 4 clusters. The result of the trial will be used to analyze cluster differences based on their characteristic characteristics. System test results are shown by table 4. below:

1. Cluster-1: customer about 29 years old with high school education and women. Customers have income below Rp.2.500.000 and buy batik with an average price of Rp170.000.
2. Cluster-2: customer about 49 years old with S1 and female education, income above Rp.2.500.000, batik with average price Rp.460.000.
3. Cluster-3: customer with average age 55 years old with S1 education of men, income above Rp.2.500.000 with batik price Rp.500.000.
4. Cluster-4: customer with average age 43 years with junior high school education and female enthusiast, income below Rp.1.000.000 with batik price Rp.100.000.
5. Cluster-5: customer with average age of 43 years with high school education and male enthusiast, income below Rp.2.500.000 with batik price Rp.200.000.

3.2 FP-Growth process based on FP-Tree

Data taken from the Customer who make the purchase process of batik. So, the process of buying batik recommendations using FP-Growth begins with the formation of FP-Tree from the purchase transaction data. The sample of this study was using 10 data taken randomly in SMEs Batik. The initial stage is to filter there batik data using the value of support 0.3 is a value that is above the minimum value of support that has been determined. FP Tree by setting the assumption of support value more than the specified min support value. Table 6, by filtering the data by removing batik items that do not meet the minimum support values and sorting the order of items based on Table 7 for the manufacture of FP-Tree.

Table 5: Sample Transaction of Data Batik.

Transaction	Id of Batik Products
1	69, 70, 71, 72, 68
2	68, 71, 70, 69
3	45, 72
4	40, 44, 76,83, 84
5	61, 60, 52, 19, 72, 79
6	60, 61, 83, 84, 86, 71
7	79, 44,40
8	44, 70
9	70, 68
10	61, 70

Table 6: Frequency of items with min support ≥ 0.30

Id of batik products	Frequency	Support Value
70	5	0.5
71	3	0.30
72	3	0,30
68	3	0.30
84	2	0.25
79	2	0.25
44	2	0.25
40	2	0.25
69	2	0.25
83	2	0.25
61	2	0.25

Table 7: Transactions that have been sorted according to the frequency of items

Transaction	Id of Batik Products
1	70, 71, 72, 68, 69
2	70, 71, 68, 69
3	72, 45
4	40, 44, 76,83, 84
5	72, 61, 79, 60, 52, 19
6	71, 61, 83, 84, 86, 60
7	79, 44,40
8	70, 44
9	70, 68
10	70, 61

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Each node in the FP-Tree contains three informations: item labeling, support count, and a connecting pointer that connects the nodes with the same item label between the paths. The picture below is the result of the formation of transactions 1 and 2. The role of FP-Growth Algorithm is to find the pattern of frequent itemset of FP-Tree that has been formed before. After FP-Tree is complete, FP-Growth algorithm looks for all possible subsets by generating conditional FP-Tree and searching for frequent itemset.

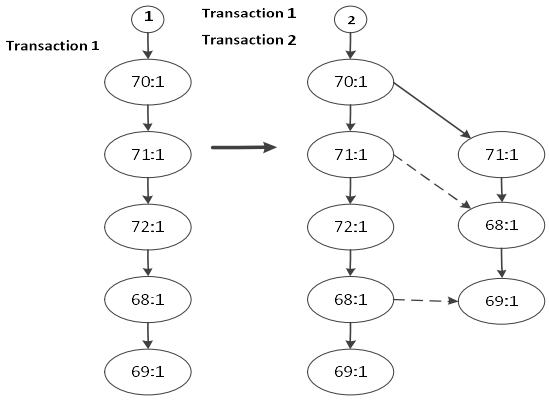

Figure 2: Results of FP-Tree establishment after reading of transactions 1 and 2

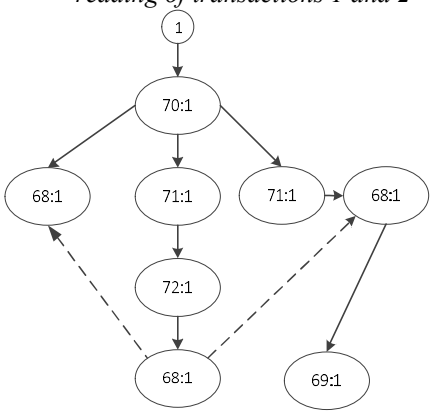

Figure 3: Tree ending of item 68

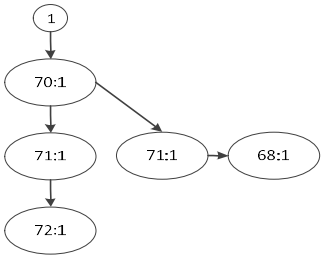

Figure 4: Tree after item 68 is omitted

The first stage takes a conditional pattern base on the FP-Tree by scanning the FP-Tree with the path prefix from the bottom up. Example item for 68, if removing items with support count does not meet the minimum support in figure 4 item 72 only has 1 occurrence with item 68. So it can be removed by generating conditional FP-Tree in table 8 and from the Tree ending in item 68 can be taken as a conditional pattern base with the results in the table 8.

Table 8: Conditional Pattern Base and FP-Tree

FP-Tree	Id of Batik Products
Conditional Pattern Base	{70:1},{70,71:1},{70,72,71:1}
Conditional FP-Tree	{270,271:1}

After the conditional FP-Tree is formed, the next step is to subsets from conditional FP-Tree to the item to generate frequent item sets. The result of the merger of K-Means Clustering and FP-Growth, where each cluster will be analyzed from the products purchased based on customer tastes. The result of analysis of purchased product will be analyzed by FP-Growth for each cluster according to Table 9. Test results use 100 iterations and minimum support 60 and minimum confidence 40. They have been done to see how the rules pattern in one month. The test results in each cluster have resulted confiddent. They depend on the number of iterations and data changes.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 9: Result of Rule Formed on The Test of Each Cluster

Amount (C)	Result Rules Every Cluster	Confidence (%)
2 Cluster	Batik Sekar Jagat →Batik Manokan/Batik bengkopi	0,62
	Batik Paraoh → Batik bengkopi/Batik Manokan/Batik Kucing Rindu	0,54
	Batik junjung derajat →Batik Sekar Jagat/Batik Pecah Batu/Batik Matahari Terbit	0,41
3 Cluster,	Batik Sekar Jagat → Batik Suramadu/Batik Manukan	0,46
	Batik Mega Mendung → Batik Tiga Dimensi Batik Ular Naga/Batik Pecah Batu	0,53
	Batik Suramadu → Batik Kapper/Batik Panca Warna/Batik Premium/ Batik Mega Mendung	0,61
4 cluster	Batik Pecah Batu → Batik bengkopi/Batik Manokan/Batik Kucing Rindu/Batik Beras Tumpah	0,44
	Batik Motif Ayam Jago → Batik Burung Merak/Batik Bangao/Batik Unggas/Batik Manokan	0,51
	Batik Ikan → Batik Ular Naga/Batik Suramadu	0,60
5 cluster	Batik Paraoh → Batik bengkopi/Batik Beras Tumpah/Batik Matahari Terbenam	0,53
	Batik Suramadu → Batik Ular Naga/ Batik Bangao/Batik Unggas/Batik Manokan	0,22
	Batik Ikan → Batik Ular Naga/Batik Suramadu/ Batik Serat Kayu	0,52

This suggests that the appearance of a combination of items more often does not necessarily lead to better recommendation results. Determination of variable data can determine the accuracy of FP-Growth made to determine minimum support and minimum confidence to find the interrelated frequent itemset as a strategy in the promotion of batik in SMEs. The highest value of precision is found on average confidence variations generated above 50%.

The test results on the FP-Growth algorithm is the smaller the minimum value of support then the more rules are generated but not all the rules generated are valid. The minimum value of support affects the formation of the rule but does not have much effect on the validity of the rule.

3.3 The Results of Time Analysis Test

The result show smaller the minimum confidence value, so the more rules generated and the result of the rule may be invalid but the minimum confidence value will affect the validity of the rule. The length of execution time in Fp-Tree allows faster access than the usual a priori according to table 10.

Table 10: Comparison of experimental time of Apriori algorithm and Fp Growth

	Min Support=60%, Min Confident =40% , Number of Dataset = 2000			
Cluster (C)	C=2	C=3	C=4	C=5
Apriori (ms)	203	1081	1120	1208
Fp Growth base on Tree (ms)	167	872	981	1021

The calculation of lift ratio in Table 11, all the rules formed have a good strength because the value of the lift ratio rule is greater than 1. The calculation of lift ratio values is based on the assumption that consequent and antecedent are independent. The value of the lift ratio greater than 1 indicates that the attachment between two items, consequent and antecedent is strong, since the trust value or confidence rule that exceeds the minimum confidence is tested again on a percentage with the antecedent combination. Rule generated with the value of confidence close with min support 60% and min confidence 40%. They used the test for lift ratio (LR) with the result according to Table 11. Test results using Minimum Support = 60% and Minimum Confident 40% as follows:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 11: Result on Calculation of LR of Batik SME Using Min Support 60% and Min Confidence 40%

Result rules every Cluster	Conf (%)	Freq	Benc	LR
Batik Sekar Jagat →Batik bengkopi	0,62	7	0,25	2,24
Batik Paraoh → Batik bengkopi	0,54	5	0,52	1,81
Batik junjung derajat →Batik Sekar Jagat	0,81	10	0,56	1,24
Batik Sekar Jagat → Batik Suramadu	0,46	12	0,23	1,32
Batik Mega Mendung → Batik Tiga Dimensi	0,53	8	0,35	1,65
Batik Suramadu → Batik Kapper	0,61	6	0,19	2,10
Batik Pecah Batu → Batik bengkopi	0,64	9	0,24	1,76
Batik Motif Ayam Jago → Batik Burung Merak	0,51	10	0,31	1,32
Batik Ikan → Batik Ular Naga	0,60	10	0,48	1,98
Batik Suramadu → Batik Ular Naga	0,53	8	0,22	1,11
Batik Ikan → Batik Ular Naga	0,22	17	0,72	1,30

4. DISCUSSION

This suggests that the appearance of a combination of items more often does not necessarily lead to better recommendation results. Determination of variable data can determine the accuracy of FP-Growth made to determine minimum support and minimum confidence to find the interrelated frequent itemset as a strategy in the promotion of batik in SMEs. The test result has been done by determining the centroid on K-Means Algorithm to produce the group members of the group are measured by the proximity of the object based on the mean value of the group performed. The number of cluster groups is almost the same as varying in MSE values. K-Means with optimal cluster could use to indicate pattern by the smallest MSE value close to 0. So, customer segmentation results FP-Growth rules to generate the number of associated rule associations. Where search results value Lift Ratio> 1 to determine a valid rule as a decision maker.

In this study using Apriori had longer time with execution time of 1208 ms compared with k-means hybrid time usage with Fp-Growth based on Tree using 1021 ms execution time on cluster 5 usage. In hybrid model usage grouping process of data set become some groups so that objects within a group have many similarities and have many differences with other group objects for the analysis of collected sales transaction data to generate useful knowledge for business-related management in increasing sales (Samizadeh, R., 2015). Apriori Algorithm is required generate candidate to get frequent itemsets. FP-Growth algorithm does not generate candidate because it uses the concept of Tree development in search of frequent itemsets. FP-Growth algorithm can directly extract frequent Itemset from FP-Tree. So the FP-Growth algorithm is faster than Apriori algorithm. This method is used to accelerate the process of determining frequent itemset before generating rule as decision recommendation. Associate Search Rules are done through two stages: frequent itemset search and rule shrinkage. In addition, this algorithm requires a large memory allocation to search itemsets. The formed of FP-Tree can compress transaction data that has the same item, resulting in less computer memory usage, and the process of finding frequent itemset becomes faster. FP-Growth requires only twice the scanning of the database in search of frequent itemsets so that the time required becomes relatively short and efficient.

5. CONCLUSION

This study relates two systems, where grouping customer prifiling to know the taste of batik buyers based on economic level and profile. While fp-Growth using item set on the historical data of buyers who have made batik purchases used to analyze the relationship between products. Customer clusters containing customers with a certain economic level will affect the level of purchases of items of goods that affect the taste of batik purchases. The result of cluster analysis in which cluster cosmos with high economic level hence high purchasing power give bigger certainty. Research on basketball market analysis results in higher values of support and confidence, the higher the degree of accuracy of the rule or the resulting pattern. The results of research with minimum support and minimum confidence aim to determine the product often purchased simultaneously on the previous purchase transaction. The result of rule is measured lift ratio with FP-Growth base on FP-Tree more than 1, then the usage average on minsupp = 60 and minconf = 40 and the average processing time is 957 ms.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCE

[1]. Firli Irhamni, Bain Khusnul Khotimah, Dewi Rahmawati, "Improvement Integrated Performance Measurement System (IPMS) For Small and Medium Enterprise Impact Of Information Technology", Journal Of Theoretical And Applied Information Technology, Vol.95, No.2, 31 January 2017, pp.319-327.

[2]. Jagmeet Kaur, Neena Madan, "Association Rule Mining: A Survey ", International Journal of Hybrid Information Technology, Vol.8, No.7, 2015, pp.239-242.

[3]. Bain Khusul Khotimah, Firli Irhamni, And Tri Sundarwati , "Genetic Algorithm For Optimized Initial Centers K-Means Clustering In SMEs", Journal of Theoretical and Applied Information Technology,Vol.90, No.1,15 August 2016, pp.23-30.

[4]. Minal G. Ingle, N. Y. Suryavanshi, "Association Rule Mining using Improved Apriori Algorithm", International Journal of Computer Applications, Vol. 112, No. 4, February 2015, pp.37-42.

[5]. Herman Aguinis, Lura E. Forcum, Harry Joo, "Using Market Basket Analysis in Management Research", Journal of Management, *Indiana University,* Vol. 39, No. 7, November 2013, pp.1799-1824

[6]. S.Balaji1 ,Dr.S.K.Srivatsa, "Customer Segmentation for Decision Support using Clustering and Association Rule based approaches", International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 3, No. 11, November, 2012, pp. 525-529.

[7]. Er. Anand Rajavat and Er. Pranjal singh solanki, "Modern Association Rule Mining Methods", International Journal of Computational Science and Information Technology (IJCSITY), Vol.2, No.3, November 2014, pp.1-9.

[8]. Karim K. Mardaneh, "Small to Medium Enterprises and Economic Growth: A Comparative Study of Clustering Techniques", Journal of Modern Applied Statistical Methods, Journal of Modern Applied Statistical Methods, Vol. 11, No. 2, November 2012, 469-478.

[9] Chris Ding, Xiaofeng He, 2004, "Principal Component Analysis and Effective K- Means Klasterisasi", Proceeding of SIAM International Conference on Penambangan data, Vol. 2, 2004, pp. 497 – 501.

[10] Dragut, Andrea B., "Stock Data Klasterisasi and Multiscale Trend Detection", Methodology and Computing in Applied Probability, Vol. 14, 2012, pp. 87 – 105.

[11]. Miss Jainee Patel, Mr. Krunal Panchal, "Frequent Item-sets Based on Document Clustering Using K-Means Algorithm", International Journal of Advance Research and Innovate Ideas in Education, Vol.2, No.3, 2016, pp.1927- 1934.

[12] Alexander Strehl, Joydeep Ghosh, "Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions", Journal of Machine Learning Research, Vol. 3, 2002, pp. 583-617.

[13]. Hu Ding, Yu Liu, Lingxiao Huang, Jian Li, "K-Means Clustering with Distributed Dimensions", Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 2016.

[14]. R. Z. Inamul Hussain, S. K. Srivatsa, "A Study of Different Association Rule Mining Techniques", International Journal of Computer Applications, Vol. 108, No 16, December 2014, pp.10-15.

[15]. Aiman Moyaid Said, P. D. D. Dominic, Dr. Azween B Abdullah, "A Comparative Study of FP-Growth Variations", International Journal of Computer Science and Network Security, Vol. 9, No.5, May 2009, pp. 266-272

[16]. Yi Zeng, Shiqun Yin, Jiangyue Liu, and Miao Zhang, "Research of Improved FP-Growth Algorithm in Association Rules Mining", Hindawi Publishing Corporation Scientific Programming, 2015, pp.1-6

[17]. Jagrati Malviya, Anju Singh, Divakar Singh, "An FP Tree based Approach for Extracting Frequent Pattern from Large Database by Applying Parallel and Partition Projection", International Journal of Computer Applications, Vol. 114, No. 18, March 2015, pp. 1-5

[18]. Shadi I., Abu Dalfa, Mohammad Mikki, "K-Means algorithm with a novel distance measure", Turkish Journal of Electrical Engineering & Computer Sciences, Vol. 21, 2013, pp. 1665-1684

[19]. Hadi A. Alnabriss, Wesam Ashour. Avoiding Objects with few Neighbors in the K-Means Process and Adding ROCK Links to Its Distance, Vol. 28, No.10, August 2011, pp. 12-17

[20]. Tri Thanh Nguyen, "A Compact FP-Tree for Fast Frequent Pattern Retrieval", 27th Pacific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Asia Conference on Language, Information, and Computation , 2013, pages 430－439

[21].E.R.Naganathan, S.Narayanan, K.Ramesh kumar, "Fp-Growth Based New Normalization Technique For Subgraph Ranking", International Journal of Database Management Systems (IJDMS), Vol.3, No.1, February 2011, pp.81-91

[22]. Xiao-jun Chen, Jia Ke, Qian-qian Zhang, Xin-ping Song and Xiao-ming Jiang, "Weighted FP-Tree Mining Algorithms for Conversion Time Data Flow", International Journal of Database Theory and Application, Vol.9, No.1, 2016, pp.169-184

[23].Zhang C, Zhang S, "Association rule mining: models and algorithms", chaps 1–2, Springer, Berlin Heidelberg, New York, 2002, pp 1–46

[24] Meera N. and Shafaque F. S., "An optimized algorithm for association rule mining using FP tree", Procedia Computer Science, vol. 45, 2015, pp. 101 – 110