Journal of Theoretical and Applied Information Technology
15th June 2021. Vol.99. No 11

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

UNSUPERVISED LEARNING AS A DATA SHARING MODEL IN THE FP-GROWTH ALGORITHM IN DETERMINING THE BEST TRANSACTION DATA PATTERN

[1]MUSTAKIM, [2]ULYA KHAIRUNNISA, [3]ALEX WENDA, [4]AHMAD ILHAM, [5]FOLKES E. LAUMAL, [6]ACHMAD DAENGS GS, [7]DARMA SETIAWAN PUTRA, [8]IDA BAGUS ARY INDRA ISWARA, [9]SRI RAHMAWATI FITRIATIEN, [10]ROBBI RAHIM

[1,2]Puzzle Research Data Technology (Predatech), Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Riau, Indonesia

[1,2]Department of Information System, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Riau, Indonesia

[3]Department of Electrical Engineering, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Riau, Indonesia

[4]Informatics Department, Universitas Muhammadiyah Semarang, Semarang, Indonesia

[5]Department of Computer Engineering, Politeknik, Negeri Kupang, Indonesia

[6]Faculty of Economic, Universitas 45 Surabaya, Surabaya, Indonesia

[7]Informatics Department, Politeknik Aceh Selatan, Indonesia

[8]STMIK STIKOM Indonesia, Denpasar, Bali, Indonesia

[9]Mathematics Education Departement, Faculty of Teacher Training and Education, Univeristas PGRI Adi Surabaya, Indonesia

[10]School of Computer and Communication Engineering, Universiti Malaysia Perlis, Pauh, Malaysia

E-mail: mustakim@uin-suska.ac.id

ABSTRACT

Market Basket Analysis is an analysis related to consumers and products in marketing. One of the successes of a company in the retail sector depends on promotion and shopping cart analysis. The data patterns generated from an association-based analysis are mostly applied by companies, one of which is the use of data mining technology. FP-Growth has been known as a reliable algorithm in terms of association, but some obstacles in its implementation in the field are often not finding a rule if using a diverse dataset. Unsupervised Learning or what is often known as grouping techniques such as K-Means, K-Medoid, and Fuzzy C-Means (FCM) can divide optimal data based on euclidean distances so that it finds better data patterns than without data sharing, especially in the case of FP-Growth. Comparisons are made by experimenting with the number of clusters 2 to 7, each of which is applied to the clustering algorithm. The results of these experiments, K-Medoid is the algorithm with the best validity value compared to other algorithms. Besides, the use of unsupervised learning techniques combined with FP-Growth can generate rules for each algorithm compared to simply applying FP-Growth.

Keywords: *K-Means, K-Medoids, FCM, FP-Growth, Data Sharing, Silhouette Index, Cluster Validity.*

1. INTRODUCTION

The development and business competition in world trade through the free market economy and advances in information technology have brought companies to an increasingly competitive level in meeting customer demands which are also getting higher. So that business people must find solutions and think of strategies or breakthroughs that can ensure the sustainability of their business. In the retail business, one way that can be done to determine market conditions is by observing sales transaction data. Sales transactions are stored in the company database and then this data is processed to produce sales reports and income statements. However, the sales transaction data can be further processed so that new information is obtained [1].

The currently growing business is 212 Mart, where the concept adopted by this company is the sharia model. In daily, monthly to yearly transactions, this company has very large data and can be analyzed using data mining techniques. Data Mining has various algorithms for analyzing data including clustering, association rules, classification, and others [2].

In retail business problems, the association rule technique is a suitable solution for analyzing market baskets. Association rules are used to identify correlations or patterns between objects [3][4]. There are many algorithms proposed to find association rules [5], one of which is the FP-Growth algorithm. FP-Growth is present as a development of the Apriori algorithm [6] by using a tree building concept called FP-Tree in the search for frequent itemsets, no longer using candidate generation [7]. Another advantage of Algotima FP-Growth is that there is no need to perform repeated database scans if a previously found itemset combination has been found [8], and the FP-Growth algorithm has faster performance than the Apriori algorithm [9].

However, a number of technical problems were found related to the most common recommendation techniques in analyzing market baskets using the association rules method. Analyzing market baskets with large itemsets tends to be ignored by association rules, and item recommendations are less precise because of the unavailability of information on retail products [10] so that for large data the results obtained are less accurate [11][12]. In overcoming this problem, it is necessary to group the attributes used to form the same group of attributes, and then proceed to determine the association pattern in each group [13], so that they can simplify the search process for product recommendations. Clustering aims to group objects in a cluster that have a high degree of similarity to each other and are not very similar to other objects in different clusters [14][15].

Based on previous research comparing sales transaction data that was clustered first, then after finding the best cluster, it was continued with the application of the FP-Growth algorithm to find association rules, with data on sales transactions that were not clustered first [16], obtained 5 rules in cluster 3, meanwhile there are no rules found in large data that are not clustered. The clustering algorithm used in this study is 3 using the K-Medoids algorithm. Where it is known in previous research comparing the K-Medoids algorithm and Fuzzy C Means (FCM) for grouping sales data [17], the Fuzzy C Means Algorithm is better than the K-Medoids Algorithm with a Silhouette Index value of 0.2159, while K-Medoids with a value of 0.2018. Meanwhile, research comparing the Fuzzy C Means (FCM) and K-Means algorithms on GPS-based transportation mode grouping [18], with the Silhouette Coefficient Index (SI) test shows that the K-Means algorithm is superior to FCM. The resulting average value of the K-Means algorithm is 0.458267. While the average value of the FCM algorithm is 0.440682. Then, in a study comparing the K-Means and K-Medoids algorithms on the grouping of disability distribution areas in children [19], the validity value of the Silhouette Coefficient Index (SI) was obtained in the K-Means algorithm. of 0.1443 and the K-Medoids algorithm of 0.5009. This shows that the K-Medoids algorithm is better than the K-Means algorithm.

Based on the description above and supported by several previous studies, this study will conduct research using the K-Medoids, K-Means, and FCM algorithms in the clustering process to then compare between the three algorithms so that it will produce which algorithm has the best cluster. based on the results of the best cluster validity values with a large number of sales transaction data sources. Then, the best cluster data results will be applied to the FP-Growth algorithm to find the pattern of association rules, so that it is expected to be able to provide more accurate product recommendations to customers because the dataset to be associated is smaller.

2. LITERATURE REVIEW

2.1. Clustering Algorithm

Cluster analysis is based on several objects that are compared using the distance function in building a model [20]. There are many types of clustering algorithms that are often used in the case of structural and non-structural data [21]. The K-Means algorithm is a well-known partition algorithm for clustering. The K-Means algorithm groups data based on their proximity to each other according to Euclidean distances [22]. In the K-Means algorithm, 'n' the number of observations is divided into 'k' clusters so that the observations in a cluster are closest to each other in reference values such as cluster mean and object distance [23]. The K-Means algorithm aims to minimize the objective function known as the squared error function given in Equation (1) [24].

$$J_{KMeans}(x;v) = \sum_{i=1}^{c}\sum_{j=1}^{n} D_{ij}^{2} \quad (1)$$

The K-Medoids algorithm or Partitioning Around Medoid (PAM) was proposed by Kaufman and Rousseeuw, as a development of the K-Means algorithm [23]. The K-Medoids algorithm uses a representative object (medoid) as the center of the cluster, not using the mean value as the center of the cluster [25]. The medoid for each cluster is calculated using the formula in Equation (2).

$$\sum_{j\in C_i} d(i,j) \quad (2)$$

The Fuzzy C Means (FCM) algorithm was presented in its original form by Dunn (1974) as an alternative to the classical k-means clusters and completed by Bezdek (1974) [25]. In the FCM algorithm, membership of the data is not explicitly assigned a value of 0 and a value of 1, but with a membership degree value with a value limit of 0 to 1 [24]. The FCM algorithm minimizes the objective function in Equation (3).

$$J_{FCM}(X;U,V) = \sum_{i=1}^{C}\sum_{j=1}^{n} u_{ij}^{m} D_{ijA}^{2} \quad (3)$$

2.2. Market Basket Analysis and FP-Growth

Market Basket Analysis is a mathematical modeling technique based on the theory that if you buy a certain group of goods, you are likely to buy another group of goods. It is used to analyze customer purchasing behavior and help in increasing sales and maintaining inventory by focusing on point of sale transaction data [3][26]. One of the well-known algorithms from market basket analysis is Apriori and FP-Growth. The FP-Growth algorithm is an algorithm of the association rules method with the concept of building a tree, which is commonly called the FP-Tree, in searching for frequent itemsets, and does not use generate candidates [27][28].

2.3. K-Means Clustering

K-Means clustering is one of the simplest and most common clustering techniques known. To do the clustering, the value of k must be determined first. Usually, users already have initial information about the object being studied; including how many clusters are most appropriate. It can use the measure of dissimilarity to group objects in detail. Dissimilarity can be translated into distance concepts. If the distance of two objects or point data is close enough, then the two objects are similar. The higher the distance value, the higher the dissimilarity [2][29].

2.4. K-Medoid Clustering

The K-Medoids algorithm is used to find medoids in a cluster. K-Medoids is stronger than K-Means in finding k as a representative object to minimize the number of data object inequality, reduce noise and outliers [22]. The basic strategy of this algorithm is to find k clusters in n objects first randomly. Each remaining object is grouped with the most similar Medoid. K-Medoids algorithm uses representative objects as representative points in retrieving the average value of objects in each cluster [30]. The distance between objects i and j is calculated using the dissimilarity measurement function, where one of them is the Euclidean Distance Function shown in equation 4 [31][32]:

$$\mathrm{dii} = \sqrt{\sum_{a=1}^{p}(x_{iu} - x_{iu})^2, i = 1, \dots\dots n; j = 1, \dots . n} \quad (4)$$

From the equation above, Xia is the a-variable of object i (i = 1, ..., n; a = 1, ..., p) and dij is the Euclidean Distance value. The algorithm also calculates the exchange probability of each object with another cluster center using criteria functions such as equation 5:

$$E = \sum_{j=1}^{k}\sum_{\rho\epsilon cj}^{n} |p - 0j| \dots \dots\dots\dots \quad (5)$$

The equation 3 above implies that E is the sum of absolute errors for all objects in the dataset; p is the point in the space that represents an object in the Cj cluster, and oj is the object in the Cj cluster.

2.5. Fuzzy C-Means Clustering

The FCM first was introduced by Jim Bezdek in 1981[33], that provided possibility of one data owned by two groups or more. The ability of FCM to separate data towards different group with certain score, hence the lost score could count the distance from complete dataset and used it [34][35]. FCM was one of the techniques to cluster the data which was the existence of every data in every cluster determined by the level of membership [36]. FMC belongs to supervising clustering method which is the number of the center of the cluster determined in the clustering process [33], with some main process that was shown in equality 6, 7, and 8 [37][38].

Calculate the center of the k-th cluster: Vkj, with k=1,2,....,c; and j=1,2,...,m:

$$V_{kj} = \frac{\sum_{i=1}^{n}\left((\mu_{ik})^{w} \times X_{ij}\right)}{\sum_{i=1}^{n}(\mu_{ik})^{w}} \quad (6)$$

Calculates an objective function on the t-iteration:

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(X_{ij}-V_{kj}\right)^2\right](\mu_{ik})^2\right) \quad (7)$$

Calculates the change in the partition matrix

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (8)$$

Bedzek filed fuzzy clustering validity by counting the Partition Coefficient Index (PCI) as the score evaluation of membership data in every cluster. The PCI value only evaluates the level of membership, without using the vector (data) that contains geometric data usually. The score in range [0,1], the bigger score is close to 1 means that the achievement of cluster's quality is better [12][39].

2.6. Silhouette Index

The silhouette method analysis can be used for k-means algorithm validation. The calculation of coefficient silhouette value can vary from -1 to 1. If SI = 1 means the object, i is already in the right cluster [40]. To simplify the analysis, the value of s is converted into two values, 1 if the silhouette value is greater than 0 and 0 if the silhouette value is less than 0 [41]. So with the result of previous converted s value calculated, the cluster is the best cluster because less s values under 0 [42].

3. RESEARCH METHODOLOGY

Data collection in this research was obtained from sales transaction data of 212 Mart Merdeka Street, Dumai City, Indonesia on May 1, 2019, to October 31, 2019. The clustering process using the K-Means, K-Medoids, and FCM algorithms was carried out to experiment product data sharing before the association rule was carried out using the FP-Growth algorithm. In this research, data is grouped into small, medium, and large volumes in units of the transaction period.

The hope of splitting datasets to find the best association relationship in each transaction. The splitting of dataset on time groups is also expected to provide the best conclusions, whether or not it is necessary to clustering. The experiment was carried out solely to find the best accuracy of the FP-Growth algorithm process as a company recommendation in preparing a market basket.

Data cleaning is done by deleting data containing less than 2 items, so that items used are product categories categorized by shelf and non-shelf. 12,916 data records were obtained after cleaning the data to be used in this study. This research is intended to compare the data association rule with the data association rule which is carried out by the clustering process first so that it can be analyzed to determine the marketing strategy for the main company.

4. RESULT AND ANALYSIS

Data grouping using the K-Means, K-Medoid, and FCM algorithms is done by experimenting with the number of clusters from 2 to 7 clusters. The search for data grouping was conducted, namely, the entire transaction data used (May 1, 2019 – October 31, 2019) with 12,916 data records, the first 3 months of transaction data (May 1, 2019 – July 31, 2019) with 6,748 data records, and the last 3 month of transaction data (August 1, 2019 – October 31, 2019) with 6,168 records. Then, it is continued with the determination of the cluster validity value using the measurement of the validity of the Silhouette Index (SI). The most homogeneous SI value is the highest SI value. The following is the cluster validity results obtained in each experiment in grouping data.

Journal of Theoretical and Applied Information Technology
15th June 2021. Vol.99. No 11

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

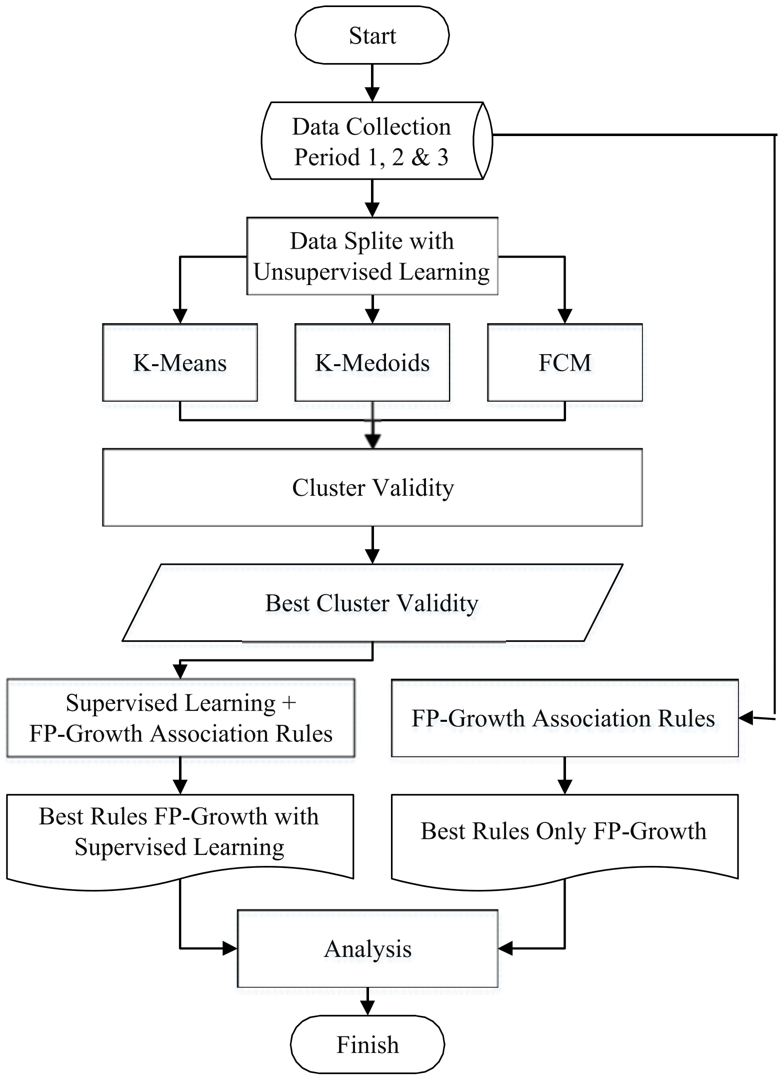

Figure 1. Research Methodology

4.1. Data Sales Transaction Data Grouping with All Data

Based on Figure 2 above, it can be seen that the number of groupings with 2 clusters has the closest average between the algorithms used. This shows that the resulting cluster validity has a very homogeneous data range. In contrast to the others, FCM tends not to produce members in each cluster which indicates that the data diversity in this algorithm is not very good. Apart from that, K-Medoid is an algorithm with the highest cluster validity value in each number of clusters, which means that K-Medoid for overall data is the best algorithm with the highest SI value of 0.8402 which is in the number of clusters 5, higher than K-Means and FCM with the highest SI values of 0.5697 and 0.4640.

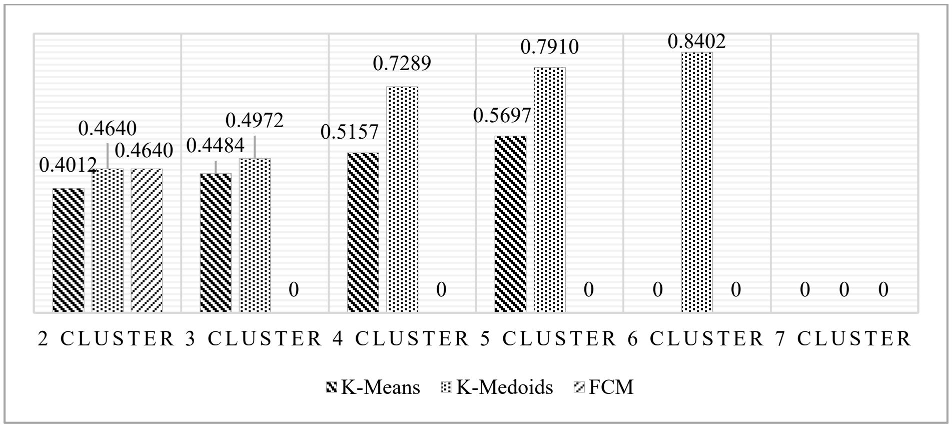

Figure 2. Comparison of SI Value Result of the K-Means, K-Medoid, and FCM Algorithms Against All Transaction Data (May 1, 2019 – October 31, 2019)

4.2. Sales Transaction Data Grouping with First Three Months Data

From Figure 3 above, it can be stated that the algorithm used for the first 3 months of data experiments resulted in the average data in each algorithm being in the number of clusters 2 and 6. Similar to the overall data used previously, the first 3 months of data also had a tendency that K-Medoid is the algorithm with the highest validity value in each number of clusters and FCM is the low validity algorithm that has no members in the experiment of 3 clusters, 4 clusters, and 5 clusters. In this data, it is also obtained that the experiment stops with the number of experiments 6. Therefore, in this data K-Medoid is stated as the best algorithm with an SI value of 0.8497 which is greater than K-Means and FCM.

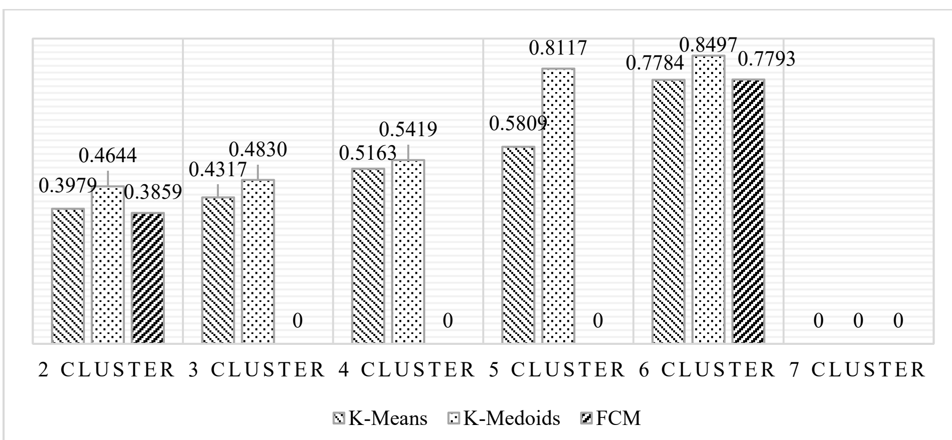

Figure 3. Comparison of SI Value Result of K-Means, K-Medoid, and FCM Algorithms Against Transaction Data for the First 3 Months (May 1, 2019 – July 31, 2019)

4.3. Sales Transaction Data Grouping with the Last Three Months Data

Based on Figure 4 above, it can be stated that the distribution of each data in the last 3 months is better than the entire data and data of the first 3 months. This is shown in the experiment, FCM has an increase in the validity of the cluster and the data homodicity. In addition, in some experiments, FCM has a better SI value compared to K-Means, namely in experiments with the number of clusters 2 and 4. These results can be concluded that the more the number of clusters formed in the K-Means and K-

Journal of Theoretical and Applied Information Technology
15th June 2021. Vol.99. No 11

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

medoid algorithms, the resulting SI value is also high, even if there is a decrease it is not too significant. Similar to the overall data and data for the first 3 months, the experiment with the number of clusters of 6 is the maximum experiment produced, namely with K-Medoid as the best algorithm compared to K-Means and FCM.

Based on the results of the cluster validity value of the 3 clustering algorithms, the K-Medoids algorithm is the algorithm that has the highest cluster validity value compared to the results of the cluster validity value of the K-Means and FCM algorithms, both for the overall trial of transaction data (May 1, 2019 – October 31, 2019), transaction data for the first 3 months (May 1, 2019 – July 31, 2019), transaction data for the last 3 months (August 1, 2019 – October 31, 2019), namely in cluster 6, with validity values of 0.8402, 0.8497 and 0.7234.

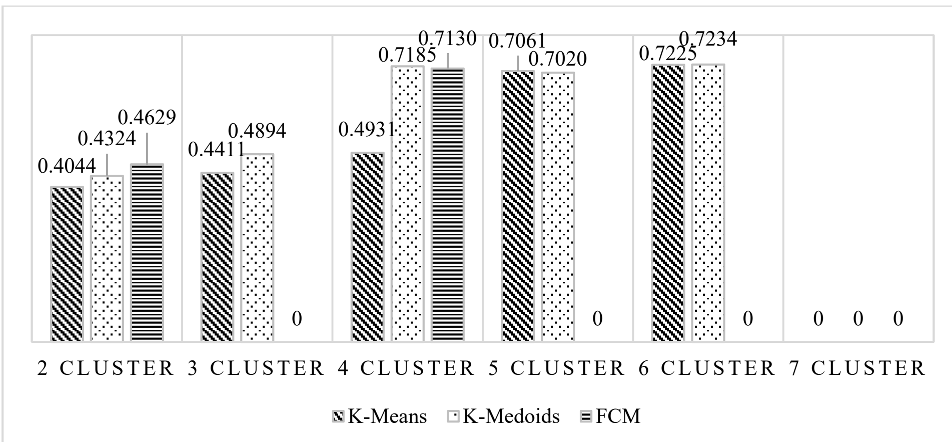

Figure 3. Comparison of SI Value Result of the K-Means, K-Medoid, and FCM Algorithms Against Transaction Data for the Last 3 Months (August 1, 2019 – October 31, 2019)

4.4. Data Association of K-Medoid Algorithm Cluster Results with the FP-Growth Algorithm

After getting the cluster with the best validity value, the next step is that the best cluster results are associated using the FP-Growth algorithm. The results of the association process will be measured using the support value and the confidence value is tested with a minimum support value of 5% and a minimum confidence value of 75%. Regarding the entire data (May 1, 2019 – October 31, 2019), rules were found in clusters 2, 4, and 6. Regarding the transaction data for the first 3 months (May 1, 2019 – July 31, 2019), rules were found in clusters 2, 3, and 4. Transaction data for the last 3 months (August 1, 2019 – October 31, 2019), rules were found in clusters 2, 3, and 4. In detail, the experimental rules of the FP-Growth algorithm on K-Medoid are shown in table 1.

Table 1. FP-Growth Experiment on K-Medoid Algorithm

Trial Month	Cluster	Implications	Support	Confidence	Number of Rules
May 1, 2019 - October 31, 2019	2	Packaged Beverages	5.00%	76.00%	1
	4	Spices	11.60%	86.00%	5
		Packaged Beverages	9.75%	85.75%	4
	6	Packaged Beverages	7.50%	76.50%	2
May 1, 2019 – July 31, 2019	2	Packaged Beverages	7.29%	78.57%	7
	3	Spices	5.00%	78.00%	1
	4	Spices	13.25%	89.00%	4
		Packaged Beverages	12.14%	86.29%	7
		Biscuit	14.00%	84.40%	5

Trial Month	Cluster	Implications	Support	Confidence	Number of Rules
August 1, 2019 – October 31, 2019	2	Spices	14.00%	87.80%	5
		Packaged Beverages	13.50%	83.25%	4
	3	Spices	6.00%	80.00%	1

The results of the search for association patterns using the FP-Growth algorithm from the transaction data that has been clustered, with a minimum support value of 5% and a minimum confidence value of 75%, in the search experiment for all transaction data, 12 rules were obtained.

4.5. Comparison of Data Associations Using the FP-Growth Algorithm (Without Clustering) and Data Sharing Using Clustering

Based on the results of the search for rules/patterns using the FP-Growth algorithm with transaction data without prior clustering, no rules/patterns were found from the minimum support and minimum confidence values that were predetermined for both the search for all transaction data (May 1, 2019 – October 31, 2019), transaction data for the first 3 months (May 1, 2019 - July 31, 2019), as well as transaction data for the first 3 months (August 1, 2019 - October 31, 2019).

Whereas in FP-Growth with the distribution of clustering data, the rules generated in this research are, in the search category of all transaction data used, the total rules obtained are 12 rules. In the search category for the first 3 months of transaction data, the total number of rules obtained is 19 rules. In the search category for the transaction data for the last 3 months, the total rules obtained are 24 rules. Then, there are 3 similarities of rules generated against the 3 data search categories. There are also 2 similarities in the generated rules to the search category for the entire transaction data used with the transaction data search category for the first 3 months. And there are 2 similarities in the rules generated against the search category for the entire transaction data used with the transaction data search category for the last 3 months. The rules that have been produced in this research can be a consideration for decision making by the company in determining product offering recommendations.

5. CONCLUSION

The conclusion regarding the results of this research is that after data clustering was carried out using the K-Means, K-Medoids, and FCM algorithms with the experimental number of clusters 2, 3, 4, 5, 6, and 7 as experimental modeling, an algorithm was obtained that had the best cluster validity namely the K-Medoid algorithm with a SI value of 0.8402 in the overall data, 0.8497 in the category of the first 3 months, and 0.7234 in the category of the last 3 months. The maximum experiment that resulted from mapping the maximum data clusterization was in the experiment with the number of clusters 6, for experiments with 7 clusters in the three algorithms resulted in a validity value of 0. Based on the Association Rule that has been formed for each category of search rules, there are 8 types of categories that are interrelated. Most frequently purchased by customers, namely the categories of Cooking Oil, Tea/Coffee, Cooking Seasoning, Packaged Beverages, Cold Snacks, Rice, Baking Ingredients, and Basic Ingredients. The maximum rule that is formed from the research experiment is 7 rules which are in the first 3 months of data. Of all the data used, rules were not found in the data that were not clustered first. Therefore, first doing the clustering process on a large amount of data is the right thing to do before searching for association rules.

REFERENCES:

[1] D. Listriani, A. H. Setyaningrum, and F. E. M.A, "Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro)," *J. Tek. Inform. Vol 9 No. 2, Univ. Islam Negeri Jakarta*, vol. 9, no. 2, pp. 120–127, 2016.

[2] T. Velmurugan, "Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points," *Int. J. Comput. …*, vol. 3, no. 5, pp. 1758–1764, 2012.

[3] R. Karthiyayini, "International Journal of Advanced Research in Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 6, no. 10, pp. 241–246, 2016, doi: 10.1021/jp0117009.

[4] W. Zhang, H. Liao, and N. Zhao, "Research on the FP growth algorithm about association rule mining," in *2008 International Seminar on Business and Information Management*, 2008, vol. 1, pp. 315–318.

[5] M. Ykhlef, "A Quantum Swarm Evolutionary Algorithm for mining association rules in large databases," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 23, no. 1, pp. 1–6, 2011, doi: 10.1016/j.jksuci.2010.03.001.

[6] Mustakim *et al.*, "Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau," *J. Phys. Conf. Ser.*, vol. 1114, no. 1, 2018, doi: 10.1088/1742-6596/1114/1/012131.

[7] M. M. Kavitha and S. T. Tamil Selvi, "Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons," *Int. J. Comput. Sci. Trends Technol. (IJCS T)*, vol. 4, no. 4, pp. 161–164, 2013.

[8] K. Wang, L. Tang, J. Han, and J. Liu, "Top down fp-growth for association rule mining," in *Pacific-Asia conference on knowledge discovery and data mining*, 2002, pp. 334–340.

[9] A. Shaikh, "Web Usage Mining Using Apriori and FP Growth Alogrithm," vol. 6, no. 1, pp. 354–357, 2015.

[10] K. Christidis, D. Apostolou, and G. Mentzas, "Exploring Customer Preferences with Probabilistic Topics Models," *Ke.Tu-Darmstadt.De*, pp. 1–13, 2010.

[11] M. Yin, W. Wang, Y. Liu, and D. Jiang, "An improvement of FP-Growth association rule mining algorithm based on adjacency table," in *MATEC Web of Conferences*, 2018, vol. 189, p. 10012.

[12] H.-Y. Chang, J.-C. Lin, M.-L. Cheng, and S.-C. Huang, "A novel incremental data mining algorithm based on fp-growth for big data," in *2016 International Conference on Networking and Network Applications (NaNA)*, 2016, pp. 375–378.

[13] M. Plasse, N. Niang, G. Saporta, A. Villeminot, and L. Leblond, "Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set," *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 596–613, 2007, doi: 10.1016/j.csda.2007.02.020.

[14] K. Govindasamy and T. Velmurugan, "Analysis of Student Academic Performance Using Clustering Techniques," *Int. J. Pure Appl. Math.*, vol. 119, no. 15, pp. 309–323, 2018.

[15] B. S. Kumar and K. V Rukmani, "Implementation of web usage mining using APRIORI and FP growth algorithms," *Int. J. Adv. Netw. Appl.*, vol. 1, no. 06, pp. 400–404, 2010.

[16] I. Syukra, A. Hidayat, and M. Z. Fauzi, "Implementation of K-Medoids and FP-Growth Algorithms for Grouping and Product Offering Recommendations," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 107, 2019, doi: 10.24014/ijaidm.v2i2.8326.

[17] A. Jaini, "Penerapan algoritma fuzzy c-means dan k-medoids untuk pengelompokan penjualan dan strategi pemasaran produk," 2019.

[18] R. Syarif, M. T. Furqon, and S. Adinugroho, "Perbandingan Algoritme K-Means Dengan Algoritme Fuzzy C Means (FCM) Dalam Clustering Moda Transportasi Berbasis GPS," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 4107–4115, 2018.

[19] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.

[20] Y. Li and H. Wu, "A Clustering Method Based on K-Means Algorithm," *Phys. Procedia*, vol. 25, pp. 1104–1109, 2012, doi: 10.1016/j.phpro.2012.03.206.

[21] P. Vogel, T. Greiser, and D. C. Mattfeld, "Understanding bike-sharing systems using Data Mining: Exploring activity patterns," *Procedia - Soc. Behav. Sci.*, vol. 20, pp. 514–523, 2011, doi: 10.1016/j.sbspro.2011.08.058.

[22] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.

[23] K. G. Soni and A. D. Patel, "Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data," 2017.

[24] Z. Cebeci and F. Yildiz, "Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures," *J. Agric. Informatics*, vol. 6, no. 3, pp. 13–23, 2015, doi: 10.17700/jai.2015.6.3.196.

[25] O. Ozdemir and A. Kaya, "K-medoids and fuzzy C-means algorithms for clustering CO2 emissions of turkey and other OECD countries," *Appl. Ecol. Environ. Res.*, vol. 16, no. 3, pp. 2513–2526, 2018, doi: 10.15666/aeer/1603_25132526.

[26] R. Novita, M. Mustakim, and F. N. Salisah, "Determination of the relationship pattern of association topic on Al-Qur'an using FP-Growth Algorithms," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1088, no. 1, p. 12020.

[27] A. Bala, "Performance Analysis of Apriori and FP-Growth Algorithms (Association Rule Mining)," *Int.J.Computer Technol. Appl.*, vol. 7, no. April, pp. 279–293, 2016.

[28] I. O. P. C. Series and M. Science, "Determination of the relationship pattern of association topic on Al- Qur ' an using FP-Growth Algorithms Determination of the relationship pattern of association topic on Al-Qur ' an using FP-Growth Algorithms," pp. 0–6, 2021, doi: 10.1088/1757-899X/1088/1/012020.

[29] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 7, pp. 881–892, 2002.

[30] T. Santhanam and T. Velmurugan, "Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points," *J. Comput. Sci.*, vol. 6, no. 3, pp. 363–368, 2010.

[31] Z. Mustofa and I. S. Suasana, "Algoritma Clustering K-Medoids pada E-Goverment Bidang Information and Communication," *J. Teknol. Inf. dan Komun.*, vol. 9, no. 1, pp. 1–10, 2018.

[32] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 012016, 2021, doi: 10.1088/1742-6596/1783/1/012016.

[33] T. Hadinata, Edrian, Sembiring, Rahmad W, kusumasari, Tien Febrianti, Herawan, "The Algorithm Expansion for Starting Point Determination Using Clustering Algorithm Method with Fuzzy C-Means," *Springer Int. Publ.*, vol. 549, 2017, doi: 10.1007/978-3-319-51281-5.

[34] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer US, 1981.

[35] L. Szilagyi, Z. Benyo, S. M. Szilágyi, and H. S. Adam, "MR brain image segmentation using an enhanced fuzzy c-means algorithm," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, 2003, vol. 1, pp. 724–726.

[36] K. M. Bataineh, M. Naji, and M. Saqer, "A comparison study between various fuzzy clustering algorithms," *Jordan J. Mech. Ind. Eng.*, vol. 5, no. 4, pp. 335–343, 2011.

[37] L. Chen, C. L. P. Chen, and M. Lu, "A multiple-kernel fuzzy c-means algorithm for image segmentation," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 41, no. 5, pp. 1263–1274, 2011.

[38] Mustakim, H. E. Saputa, Mutasir, Nuryanti, and Salmiah, "The Implementation of Fuzzy-C Means to Categorize Poverty Data in Riau Province," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 012013, 2021, doi: 10.1088/1742-6596/1783/1/012013.

[39] R. Rustiyan and M. Mustakim, "Penerapan Algoritma Fuzzy C Means untuk Analisis Permasalahan Simpanan Wajib Anggota Koperasi," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 5, no. 2, pp. 171–176, 2018.

[40] A. Starczewski and A. Krzyżak, "Performance evaluation of the silhouette index," in *International Conference on Artificial Intelligence and Soft Computing*, 2015, pp. 49–58.

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

[41] P. J. Kaur, "Cluster quality based performance evaluation of hierarchical clustering method," in *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, 2015, pp. 649–653.

[42] Mustakim, "Centroid k-means clustering optimization using eigenvector principal component analysis," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 15, 2017.