

Product Recommendations Using Market Basket Analysis with FP-Growth and Clustering Techniques

Muhammad Rizky AL
Master of Management Program
Prasetiya Mulya University
Jakarta, Indonesia
muhammadriezky@yahoo.com

Meilita Tryana Sembiring
Secretary of Master Management
Department of Industrial Engineering
University of Northern Sumatra
Medan, Indonesia
meilita@usu.ac.id

Richo Giwana Resdy Maulana
Department of Industrial Engineering
University of Northern Sumatra
Medan, Indonesia
richo.giwana18@students.usu.ac.id

Abstract

Global trade competition in using technology is getting tougher in meeting customer demands. The company stores sales data such as Sukku Coffee & Space. Is a family business in the food and drink industry that has problems in using data for marketing strategy. Promotion is a communication tool to introduce company products so that they can be known to the public and attract buyers so as to increase company sales. Market Basket Analysis (MBA) helps businesses make scientific decisions by conducting Association Rule Mining between items purchased simultaneously by customers. Analytics helps provide product recommendations and promotions, resulting in more targeted marketing strategies and attracting more customers. The results of designing an analytical data mining model using MBA with Association Rules method using FP-Growth and clustering techniques with K-Means. The initial data set of sales transactions is 34,745 data which is pre-processed so that the data results are 32,802 data with the attributes of time, total collected, items, total items, receipt number. Clustering was carried out for processing 32,802 datasets using RFM. Then normalized with Z-Transformation processed using RapidMiner with 10 iterations of cluster 3, resulting in Cluster 0 producing 17,038 items categorized as frequently purchased, Cluster 1 producing 11,459 items categorized as sometimes purchased, Cluster 2 producing 4,305 items categorized as rare. purchased. From the data validation test, the result of the smallest performance vector value is Avg. within centroid distance_cluster_0 with a value of -0.716. The results of the MBA processing in the recommendation of product promotion, namely sukkuaren with chocolate with a support value of 0.156 and a confidence value of 0.991.

Keywords

Promotion Recommendation, Association Rule, FP-Growth Algorithm, Clustering K-Means and Market Basket Analysis.

1. Introduction

Global trade competition in using technology is getting more challenging in meeting customer demands. As a result, it forces the perpetrators to constantly think about how to survive and share their business market share. Meeting consumer needs daily produces more transaction data over time, resulting in a set of databases. However, many

companies process transaction data for purposes only as transaction identity, and much of this data is not appropriately utilized. Sukku Coffee & Space is a family business in the food and beverage industry. Sukku Coffee & Space uses transaction data only to recapitulate sales results and calculate company finances. Data is information that has a broad meaning, data can be utilized and processed in more detail into helpful information, such as for increasing sales and promotion of company products.

Data mining is extracting and identifying useful information and related knowledge from large amounts of data (databases) using specific techniques such as statistical techniques, mathematics, artificial intelligence, and machine learning (Mustakim, 2012). Market basket analysis can help provide scientific decision support by conducting Association Rule Mining between items purchased simultaneously by customers. This can help provide product recommendations and promotions, resulting in a more targeted marketing strategy and attracting more customers because the goods being promoted are goods that customers need. Inaccurate determination of purchasing patterns can result in product recommendation and promotion policies that are not on target (Triyanto et al., 2014). One method that can be used in data mining is Association Rule and Clustering. Association rule is a data mining method that is used to find knowledge from a large amount of data contained in a database, which aims to find frequent itemsets in the database using minimum support and create an association rule from frequent itemsets with a certain confidence. With the provision that every time event A occurs, event B occurs (Abaya, 2012). Promoted products often do not meet customer needs, so many algorithms have been proposed to find association patterns. FP-Growth is a data mining method used to find a collection of items without using candidate generation. By using the concept of tree construction when searching for frequent item sets, candidate generation is omitted (Fitria, 2017)).

Clustering is an integral part of data analysis, which aims to partition a given dataset into several clusters where datasets in a cluster are similar and different from datasets in other clusters. Partitioning of data so that objects in a cluster are more dominantly similar than objects in other clusters is generated through partitioning-based clustering. K-means is a method of grouping non-hierarchical (border) data that attempts to partition existing data into two or more groups (Asroni et al., 2018).

This research was conducted in order to find a solution for the problem where Sukku Coffee & Space needs to utilize company datasets to help product promotion. This research will determine recommendations for the promotion of Sukku Coffee & Space products to make promotions based on data that can increase company profits. For this reason, the Market Basket Analysis method is used, namely the association rule with clustering techniques. The clustering method used is the K-Means algorithm to group data into 3 product clusters. The association rule method uses the FP-Growth algorithm to produce frequent item sets and rules that are used as product recommendations for product promotion.

1.1 Objectives

This paper aims to find information related to the sales pattern of Sukku Coffee and Space products by the K-Means algorithm for clustering on FP-Growth to generate rules to determine product promotion recommendations that can increase product sales profits. Several things must be done in this study to achieve this goal, there are:

1. Processing existing datasets in the Sukku Coffee and Space database warehouse.
2. Analyzing the results of dataset processing using clustering and association rules from the Sukku Coffee and Space dataset.
3. Recommend product promotions for Sukku Coffee and Space.

2. Literature Review

The company will not do whatever from its efforts without a product. Product more good made oriented market demand or appetite consumers (Riyono and Budiharja, 2016). The company should indeed desire to orient customer needs to have score competitive moment face to face with competitors. Customer satisfaction builds loyalty and creates close relationships between customers and companies (Rabiqy, 2017). The recommendation is communication and promotions that are carried out in a manner no directly by consumers or customers who have once bought and enjoyed the product or services offered. Then tell various experiences related to a product or service to others so that could interest other consumers in buying the products (Sumali, 2014).

Data mining is a process for finding useful hidden information in data with a large number of, as well as helping in making decisions for internal stakeholders make a decision, use develops the business. Current data mining this

already many applied in various fields like research, academia, and industry. Data mining technology combines artificial intelligence (Artificial Intelligence), machine learning (Machine Learning), introduction patterns (Pattern Recognition), techniques visualization, statistics (Liu et al., 2014), and systems database management (Diwate and Sahu, 2014). Data mining combines several disciplines of purposeful science for repair techniques that can be said to be old school or traditional so they could handle (Fajrin and Maulana, 2018):

1. Data with a substantial amount.
2. Data with high dimensions.
3. Heterogeneous and different data nature.

RapidMiner is a data science software platform developed by the company RapidMiner that provides a unified environment for machine learning, deep learning, text mining, and predictive analytics used for business and commercial applications, research, education, training, rapid prototyping, and application development. In addition, it supports all machine learning steps, including data preparation, result visualization, validation, and optimization (Kori, 2017). RapidMiner is a non-programming data mining platform in which data mining and analysis processes are designed from basic building blocks called operators. Each operator executes specific commands on data, such as loading and saving, changing, or inferring a model to data.

There are techniques in successful Data Mining developed and used in study research that has been done, like classification, clustering, regression, association, network nerves, etc., to find knowledge from the database (Ramageri, 2010). Clustering is a grouping process data set in groups so that existing objects in a group have high similarities and differences from an existing object in other groups (Ghuman, 2016). Clustering is part of data analysis, which aims to partition the given dataset into some clusters where the dataset resides in a cluster similar to and different from datasets in other clusters (Raval and Jani, 2016; Rani and Rohil, 2013).

The K-Means algorithm a method analysis group that leads to the partitioning of N objects observation in K groups where every object observation owned by a group to the nearest mean, similar to the expectation-maximization algorithm for Gaussian Mixture where both of them try for find center from a group in as much data iteration repairs made by both algorithms. K-Means is one method of grouping non-hierarchical data (blocks) that partition existing data into two or more groups. Method this partition data into a group, so the data is characteristic same entered into the one same group and characteristic data different grouped into another group. As for goals, this data grouping is for minimizing function set objective in the grouping process, which generally attempts to minimize inside variations in something group and maximize variation between groups (Prasetyo, 2014).

Market Basket Analysis identifies connections between group products, items, or categories. Analysis basket shopping first originated from field marketing. Analysis Market Basket makes it easy for researchers to understand customer behavior through the product which product to buy simultaneously (Aguinis, Forcum, and Joo, 2012). this method is used for determining purchased products in a manner with and for set layout supermarket products, as well as for designing a promotion strategy to increase customer interest in a shop (Annie and Kumar, 2012) and expand marketing strategy (Gupta and Mamtora, 2014).

Rule association (Association Rules) is one standard data mining technique used. The data mining process uses rule association to find connections associative between the data and conclude the rules formed from existing data. FP-Growth is a method of technique association in data mining for search frequent itemset without using candidate generations. The FP-Growth approach is designed to do mining to find frequent itemsets in the market basket (Market Basket Analysis). Building the FP-Tree is inexpensive computing. However, if the database involves thousands of items, the formed tree becomes large (Soni et al., 2013). Han (2001) proposed FP-Growth method that because of the Divide and Conquer method, algorithms are more efficient than popular methods in looking for frequent item set patterns, for example. Algorithm Apriori (Sidhu et al., 2014).

Application algorithm association in clustered data, especially formerly about analyzing Web Usage Mining. Fauzanu et al. (2017) use the K-Means algorithm for clustering and algorithms A priori for association in their web log file data research. Hermanto (2015) implements algorithm association that is A priori and the clustering algorithm, namely K-Means, with processing transaction data purchase and customer data. Which is testing on transaction data sale goods using five comparisons support value and confidence value of 5%, 6%, 7%, 8%, and 9%. Santosa and Jadi (2016) analyze sales data using the CLARANS cluster strategy and FP-Growth association.

3. Methods

This research is classified as quantitative research. The research data is taken from Sukku Caffee and Space, namely the transaction dataset. To help solve problems regarding the lack of utilization of large amounts of data in the database at Sukku Coffee and Space to determine the best product promotion recommendations so as to increase revenue turnover. To take advantage of multiple algorithms, an integrated methodology of RFM analysis, K-Means Cluster, FP-Growth is proposed. Figure 1 demonstrates the complete flow of the proposed methodology and can be explained with 4 steps.

Step 1: Data preprocessing, Data preprocessing stage is the cleaning and preparation of data to eliminate data inconsistencies, incomplete data and data redundancy contained in the initial data. The processes contained in the preprocessing stage are data cleaning, data selection, data transformation and data normalization.

Step 2: Clustering K-Means, Clustering data is done by dividing the existing data into 3 clusters (k) based on Recent, Frequency, and Monetary (RFM) then normalized with Z-Score. The purpose of the clustering analysis is to make the associated data smaller, so that the resulting pattern can be more accurate.

Step 3: FP-Growth Association; the association process is carried out using the FP-Growth algorithm to generate frequency itemsets and rules to determine customer product recommendations. The results of this association process are measured using Support and Confidence.

The steps are taken in conducting the research start from the introduction and literature review, data collection, data preprocessing, analysis and recommendation results (Figure 1).

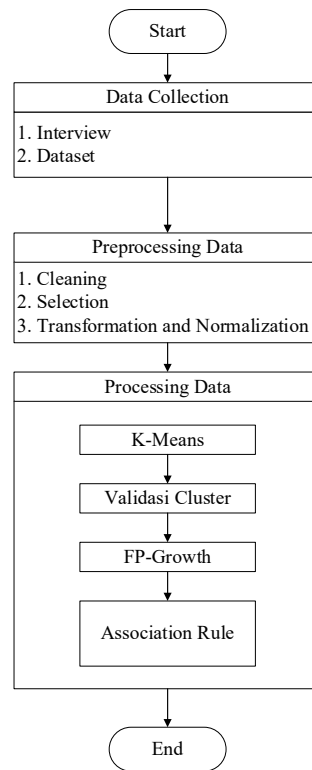


Figure 1. The Flowchart of the Proposed Methodology

4. Data Collection

The data used is secondary data obtained from Sukku Coffee and Space in the form of sales transaction data for 2020 - 2022, which is 34,745 data (Figure 2).

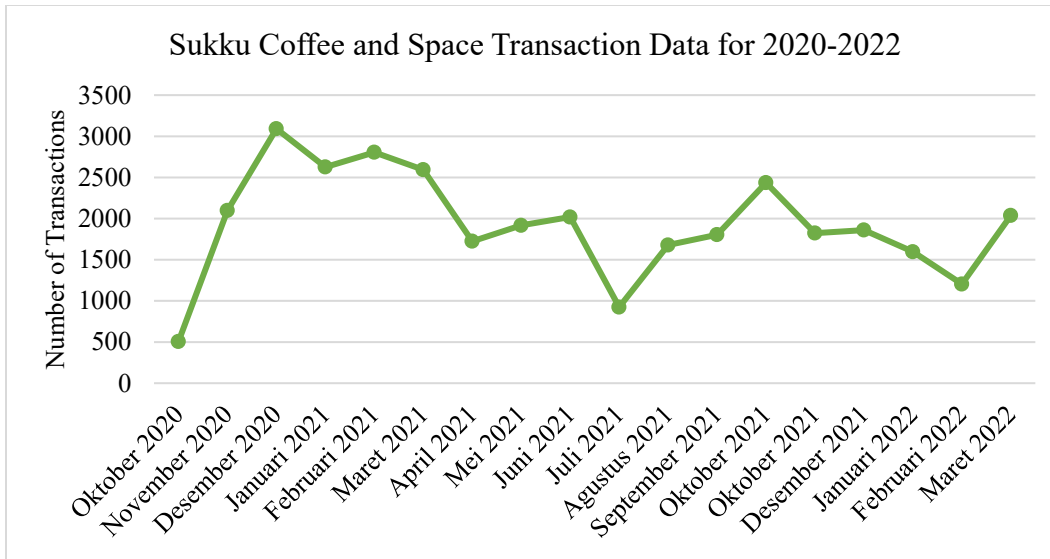


Figure 2. Sukku Coffee and Space Transaction Data for 2020-2022

Based on Figure 3 the highest number of transactions occurred in December 2020. In October 2020 Sukku Coffee and Space opened on October 28, 2020 so the number of transactions was quite low. In other months the average sales transaction ranges from 1000 to 3000 transactions per month

NO	Outlet	Date	Time	Gross Sales	Discounts	Refunds	Net Sales	Gratuity	Tax	Total Collected	Other Note (Optional)	Receipt Number	Collected By	Served By	Customer	Customer Phone	Items	Payment Method	Event Type	Reason of Refund
1	Sukku Coffee & Space	3/31/2022	20:46:59	58000	0	0	58000	0	0	58000		105J2F	Kasir Sukku				Red Velvet COLD, Choco Berry COLD	Cash	Payment	
2	Sukku Coffee & Space	3/31/2022	20:36:10	28000	0	0	28000	0	0	28000		105J2E	Kasir Sukku				Sumatera Reflection COLD	Cash	Payment	
3	Sukku Coffee & Space	3/31/2022	20:30:41	20000	0	0	20000	0	0	20000		105J2D	Kasir Sukku				French Fries Dine-In	Cash	Payment	
4	Sukku Coffee & Space	3/31/2022	20:26:42	56000	0	0	56000	0	0	56000		105J2C	Kasir Sukku				Vanilla Latte COLD, Matcha COLD	Cash	Payment	
5	Sukku Coffee & Space	3/31/2022	20:21:55	25000	25000	0	0	0	0	0		105J2B	Kasir Sukku				Sukku Aren COLD	Cash	Payment	
6	Sukku Coffee & Space	3/31/2022	20:20:10	68000	0	0	68000	0	0	68000		105J2A	Kasir Sukku				Cappuccino COLD, Lemon Tea COLD, French Fries Dine-In	Cash	Payment	
7	Sukku Coffee & Space	3/31/2022	20:04:41	28000	0	0	28000	0	0	28000		105J29	Kasir Sukku				Vanilla Latte COLD	Cash	Payment	
8	Sukku Coffee & Space	3/31/2022	19:59:30	20000	0	0	20000	0	0	20000		105J28	Kasir Sukku				Mineral Water HOT x 2	Cash	Payment	
9	Sukku Coffee & Space	3/31/2022	19:56:12	28000	0	0	28000	0	0	28000		105J27	Kasir Sukku				Sumatera Reflection COLD	Cash	Payment	
10	Sukku Coffee & Space	3/31/2022	19:45:17	80000	0	0	80000	0	0	80000		105J26	Kasir Sukku				Nasi Goreng Kampung, Sukku Katsu Dine-In, Lemon Tea COLD	Cash	Payment	
11	Sukku Coffee & Space	3/31/2022	19:40:39	50000	0	0	50000	0	0	50000		105J25	Kasir Sukku				Kaya Toast Dine-In, Chocolate COLD	BCA	Payment	
12	Sukku Coffee & Space	3/31/2022	19:35:05	50000	0	0	50000	0	0	50000		105J24	Kasir Sukku				Sukku Aren COLD x 2	Shopeepay	Payment	
13	Sukku Coffee & Space	3/31/2022	19:30:32	23000	0	0	23000	0	0	23000		105J23	Kasir Sukku				Lychee Tea COLD	Cash	Payment	
14	Sukku Coffee & Space	3/31/2022	19:25:38	78000	0	0	78000	0	0	78000		105J22	Kasir Sukku				Americano COLD, Chocolate COLD, Kanawa Twilight	Cash	Payment	
15	Sukku Coffee & Space	3/31/2022	19:13:16	55000	0	0	55000	0	0	55000		105J21	Kasir Sukku				Sukku Aren COLD, Mango Matcha	Cash	Payment	

Figure 3. Sales Transaction Dataset

As for the attributes used, such as, Outlet is the name or business identity owned by Sukku Caffe and Space . Date is data on the date of sales transactions for each receipt number starting from October 20, 2020 to March 31, 2022. Time is the time for sales transactions for each receipt number based on a 24-hour time system, which is the standard time where the day starts from 00.00 to 23.59 with West Indonesia Time (WIB) zone. . Total Collected is the total order price of the items per receipt number. Receipt Number is a label number / customer order identity consisting of 34,745 data including "105J2F, 105J2E, 105J2D, 105J2C, 105J2A, ..., 3B04LE, 3B04LD, 104S3O, 104S3M, 104S3L". Items are the number of menu items ordered based on the receipt number , namely " Vanilla Latte, Sukku Aren, Cappuccino, Americano, Dark Tribe, ..., Chicken Popcorn, Spring Roll, Banana Fritters, Banana Split, Garlic Bread". Payment method is the payment method used at the time of the transaction. Payment types available are cash and non-cash.

5. Results and Discussion

5.1. Data Preprocessing

Performed cleaning, data preparation, transformation and normalization of data to eliminate data inconsistency, incomplete data and data redundancy, changing the shape of the data so that it fits the processing format contained in the initial data.

5.1.1. Data Cleaning and Data Reduction

Data Cleaning aims to eliminate incomplete and inconsistent data in the dataset . Data reduction used for transaction data that has been collected is reduced or eliminated on the variables that are not needed in the study. The following Figure 4 is an example of the results of cleaning and reducing datasets at PT. XYZ.

No	Date	Bulan	Time	Jam	Total Collected	Receipt Number	Items	Total Item	Payment Method	Event Type	pebayaran
1	3/31/2022	3	20:46:59	20	58000	105J2F	Red Velvet, Choco Berry	2	Cash	Payment	1
2	3/31/2022	3	20:36:10	20	20000	105J3E	Sumatera Reflection	1	Cash	Payment	1
3	3/31/2022	3	20:30:41	20	20000	105J2D	French Fries	1	Cash	Payment	1
4	3/31/2022	3	20:26:42	20	56000	105J2C	Vanilla Latte, Matcha	2	Cash	Payment	1
5	3/31/2022	3	20:20:10	20	68000	105J2A	Cappuccino, Lemon Tea, French Fries	3	Cash	Payment	1
6	3/31/2022	3	20:04:41	20	20000	105J29	Vanilla Latte	1	Cash	Payment	1
7	3/31/2022	3	19:59:30	19	20000	105J28	Mineral Water x 2	2	Cash	Payment	1
8	3/31/2022	3	19:56:12	19	28000	105J27	Sumatera Reflection	1	Cash	Payment	1
9	3/31/2022	3	19:45:17	19	80000	105J26	Nasi Goreng Kampung, Saksu Katsu, Lemon Tea	3	Cash	Payment	1
10	3/31/2022	3	19:40:39	19	50000	105J25	Kava Toast, Chocolate	2	BCA	Payment	2
11	3/31/2022	3	19:35:05	19	50000	105J24	Saksu Aren x 2	2	Shopeepay	Payment	2
12	3/31/2022	3	19:30:32	19	25000	105J23	Lychee Tea	1	Cash	Payment	1
13	3/31/2022	3	19:25:38	19	78000	105J22	Americano, Chocolate, Kanawa Twilight	3	Cash	Payment	1
14	3/31/2022	3	19:13:16	19	55000	105J21	Saksu Aren, Mango Matcha	2	Cash	Payment	1
15	3/31/2022	3	19:10:50	19	84000	105J20	Matcha, Red Velvet, Chocolate	3	Cash	Payment	1
16	3/31/2022	3	19:08:56	19	20000	105J1Z	Mineral Water x 2	2	Cash	Payment	1
17	3/31/2022	3	19:08:08	19	25000	105J1Y	Lemon Tea	1	Cash	Payment	1
18	3/31/2022	3	19:06:07	19	60000	105J1X	Chocolate Vanilla, Choco Berry	2	Cash	Payment	1
19	3/31/2022	3	18:58:02	18	25000	105J1W	Lychee Tea	1	Cash	Payment	1
20	3/31/2022	3	18:38:39	18	110000	105J1V	Nasi Goreng Kampung x 2, Taro x 2	4	Cash	Payment	1
21	3/31/2022	3	18:28:18	18	25000	105J1U	Saksu Aren	1	Cash	Payment	1
22	3/31/2022	3	18:27:34	18	25000	105J1T	Cappuccino	1	Cash	Payment	1
23	3/31/2022	3	18:26:53	18	28000	105J1S	Red Ocean	1	Cash	Payment	1
24	3/31/2022	3	18:25:24	18	28000	105J1R	Chocolate	1	Cash	Payment	1
25	3/31/2022	3	18:23:53	18	25000	105J1Q	Dark Tribe	1	Shopeepay	Payment	2
26	3/31/2022	3	18:23:00	18	116000	105J1P	Chocolate, Kanawa Twilight, Creamy Spaghetti Chicken, Nasi Goreng Saksu	4	Cash	Payment	1
27	3/31/2022	3	18:12:42	18	10000	105J1O	Mineral Water	1	Cash	Payment	1
28	3/31/2022	3	18:12:00	18	55000	105J1N	Taro, Cappuccino	2	Cash	Payment	1
29	3/31/2022	3	18:00:30	18	25000	105J1M	Lemon Tea	1	Cash	Payment	1
30	3/31/2022	3	18:00:12	18	25000	105J1L	Cappuccino	1	Cash	Payment	1
31	3/31/2022	3	17:59:16	17	28000	105J1K	Chocolate	1	Cash	Payment	1
32	3/31/2022	3	17:22:20	17	60000	105J1J	Saksu Katsu x 2	2	GrabFood	Payment	2
33	3/31/2022	3	15:24:20	15	25000	105J1I	Dark Tribe	1	Cash	Payment	1
34	3/31/2022	3	15:12:07	15	28000	105J1H	Taro	1	Cash	Payment	1
35	3/31/2022	3	14:48:19	14	46000	105J1G	Lemon Tea, Lychee Tea	2	Shopeepay	Payment	2
36	3/31/2022	3	14:50:23	14	25000	105J1F	Saksu Aren	1	GrabFood	Payment	2

Figure 4. Data Cleaning and Data Reduction

In Figure 4 after data cleaning and data reduction has resulted in a new attribute dataset of 32,802, then proceed to the data attribute selection stage that will be used to process Market Basket Analysis.

5.1.2. Data Selection

Selection is done to simplify the clustering process so that the data processing is not too much and at this stage the data that has been collected will be selected variables that only have a relationship with the clustering process and analysis of recency, frequency, monetary (RFM)). Here are the attributes used can be seen as follows in Table 1.

Table 1. Attribute Data Selection

Attributes	Information
Receipt Number	Transaction identity information
Items	Transaction item type information
R	Transaction time information (Time)
F	Information on the number of transaction items (Total Items)
M	Information on total transaction amount (Total Collected)

5.1.3. Data Transformation

It is a further data transformation process so that the data is easier to process using the Association Rule Market Basket Analysis FP- Growth Method . Examples of the results of data transformation can be seen in Table 2.

Table 2. Data Items Separated Per Item

No	Receipt Number	Items_1	Items_2	Items_3	etc.
1	105J2F	Red Velvet	Choco Berry	-	...

2	1O5J2E	Sumatran Reflection	-	-	...
3	1O5J2D	French Fries	-	-	...
4	1O5J2C	Vanilla Latte	Matcha	-	...
5	1O5J2A	cappuccino	LemonTea	French Fries	...
6	1O5J29	Vanilla Latte	-	-	...
7	1O5J28	Mineral Water	-	-	...
8	1O5J27	Sumatran Reflection	-	-	...
9	1O5J26	Village Fried Rice	Katsu	Lemon Tea	...
...
32802	1O4S3L	Americano	-	-	...

5.1.4. Data Normalization

Normalization of the data used is the normalization method by determining the value of Z- score or also called zero-means normalization, where the value of an attribute A is normalized based on the average value and standard deviation of attribute A. This Z - score is useful when the number of items or forms the value between one attribute and another is not the same, even though theoretically these attributes have the same weight or relationship. For example, in this study, the value of the Frequency attribute is a number consisting of only one character, which describes the number of transactions, while the Monetary attribute is in the form of the nominal number of transactions, this will affect the cluster results. So comparison with raw data alone is not enough, therefore it is necessary to transform into a Z- score form . Processing of data transformation is carried out with the results of Z- score using tools RapidMiner . The variable that is changed is the attribute Recency, Frequency, Monetary (RFM). The following is an example of the results of data normalization as follows in Figure 5.

R Number	F Number	M Number	Receipt Number Category
0.990	-0.110	0.208	1O5J2F
0.990	-0.759	-0.599	1O5J2E
0.990	-0.759	-0.814	1O5J2D
0.990	-0.110	0.154	1O5J2C
0.990	0.540	0.477	1O5J2A
0.990	-0.759	-0.599	1O5J29
0.649	-0.110	-0.814	1O5J28
0.649	-0.759	-0.599	1O5J27
0.649	0.540	0.800	1O5J26
0.649	-0.110	-0.007	1O5J25
0.649	-0.110	-0.007	1O5J24
0.649	-0.759	-0.734	1O5J23

Figure 5. RFM Normalization Results with Z- Transformation

5.2. Clustering Algorithm

The application of the K- Means Method in this study is used for grouping transaction data which aims to make the dataset smaller so that it can facilitate the association process using associations. rule with the FP- Growth method . From the transaction dataset of PT. XYZ data that has been preprocessed is grouped using 3 clusters . The data entered in the process amounted to 32,802 rows of data with a maximum number of 10 iterations. The following is a clustering process using RapidMiner as follows in Figure 6.

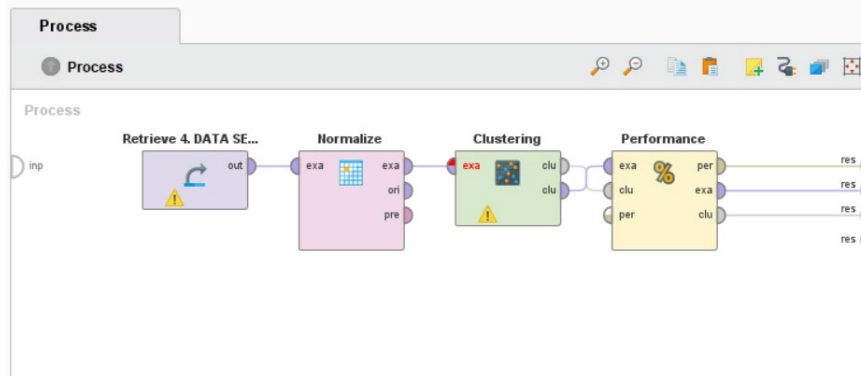


Figure 6. Clustering Process with RapidMiner

From Figure 6. The results of grouping 3 clusters using the K - Means algorithm with a Euclidean Distance measure are obtained, namely cluster 1 with 17,038 items, cluster 2 with 11,459 items, Cluster 3 with 4,305 items. The following is an image of the centroid data for C1, C2 and C3 based on the RFM dataset, which can be seen in the following Figure 7.

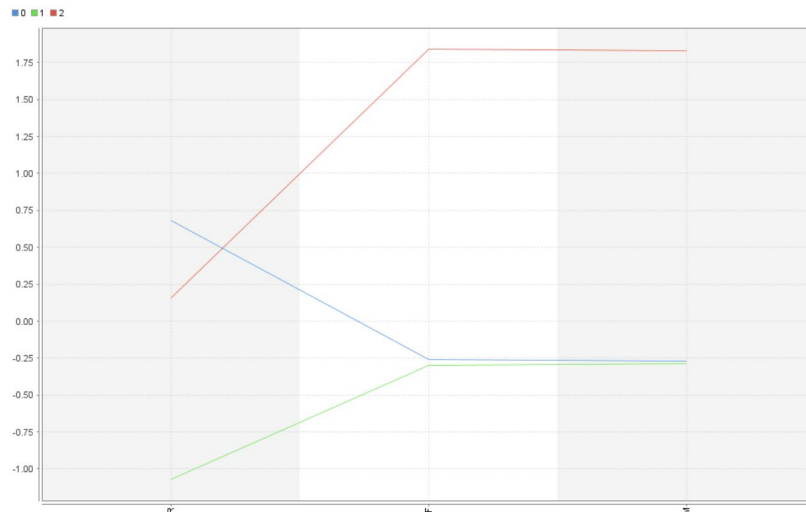


Figure 7. Centroid Data Results with RapidMiner

5.2.1 Validation test was performed using a performance vector, namely, Avg. within centroid distance and Davies Bouldin Index after the clustering process is complete. performance vector is used to evaluate the performance of a model that provides a list of performance criteria values according to the given task. The performance value for each cluster is as shown in Table 3.

Table 3. Performance Vector Results

Performance Vector	Score
Avg. within centroid distance	-1,339
Avg. within centroids distance cluster 0	-0.716
Avg. within centroids distance cluster 1	-0.871
Avg. within centroid distance cluster 2	-5,050
Davies Bouldin Index	-0.881

5.3. FP Growth Algorithm

The search for association rules was carried out in two stages, namely the search for frequent item sets and the compilation of rules. And for the frequent itemset search conducted in this study, the FP-Growth method was used. Where FP - Growth is an alternative algorithm that can be used to determine an item set frequency in a data set.

Then to measure the importance or not of an association rule, it can be seen with two parameters, namely the support value (supporting value) and the confidence value (certainty value). The support value indicates the level of dominance of the itemset from the whole data, and because in determining the itemset in this study using FP- Growth, the support value setting is regulated in the FP - Growth process. While the confidence value is a measure of how much valid or not an association rule is. The minimum support value used is 50%, while the minimum confidence value is 95%. The following is the association rule process using the RapidMiner application (Figure 8).

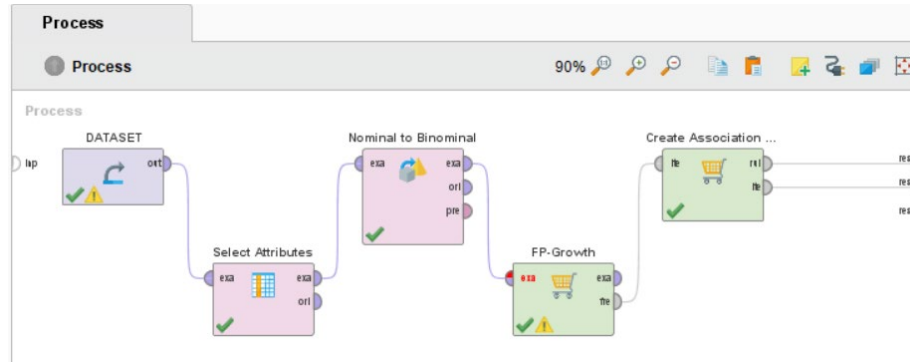


Figure 8. Association Rule Process with RapidMiner

Associate processing rule by using a minimum support t of 50% produces a set of all itemset. The itemset is known by seeing whether the itemset has met the minimum support limit. The following are the results of the frequency item set shown in Table 4.

Table 4. Frequency analysis item set FP-Growth

Size	Support	Item 1	Item 2	Item 3
1	0.983	Choco Berry	-	-
1	0.870	Red Velvet	-	-
1	0.157	Sukku Aren	-	-
1	0.124	Chocolate	-	-
2	0.854	Choco Berry	Red Velvet	-
2	0.156	Choco Berry	Sukku Aren	-
2	0.123	Choco Berry	Chocolate	-
2	0.145	Red Velvet	Sukku Aren	-
2	0.107	Red Velvet	Chocolate	-
3	0.144	Choco Berry	Red Velvet	Sukku Aren
3	0.106	Choco Berry	Red Velvet	Chocolate

From the results of the frequent items set, it produces a maximum size of 3 items with the highest support value of 0.983 item 1, namely Choco Berry and the lowest support value of 0.106 items ChocoBerry, Redvelvet, Chocolate. Furthermore, the results obtained from the association rules based on create association rule based on confidence benchmark (certainty value). The results of the association rule can be seen in Table 5 as follows.

Table 5. Association Rule Result Analysis

Premises	Conclusion	support	Confidence	elevators
SukkuAren	ChocoBerry	0.156	0991	1009

Red Velvet, SukkuAren	ChocoBerry	0.144	0991	1008
Chocolate	ChocoBerry	0.123	0989	1006
Red Velvet, Chocolate	ChocoBerry	0.106	0987	1,004

6. Conclusion

Based on the results of the processing and analysis that has been carried out, the conclusions of this study are obtained from the results of data processing and analysis that have been carried out. The following are some of the conclusions drawn:

1. In processing Market Basket Analysis using clustering technique on 34,745 rows of dataset with K-Means and FP-Growth algorithms. Starting from pre-processing to data normalization (Z-Transformation) against Recency, Frequent, and Monetary (RFM) to minimize data. Processing the K-Means algorithm with 3 data clusters, then testing for validation to obtain clusters that were processed with the FP-Growth algorithm with a min support of 50% and a min confidence of 95%.
2. Analysis of the results of Market Basket Analysis obtained 4 association rules between premises and conspiracy, first SukkuAren and ChocoBerry, have a support value of 0.156 and a confidence value of 0.991. Both Red Velvet, SukkuAren and ChocoBerry have a support value of 0.144 and a confidence value of 0.991. The three Chocolate and ChocoBerry, have a support value of 0.123. The four Redvelvets, chocolate with ChocoBerry, have a support value of 0.106 and a confidence value of 0.987.
3. Product promotion recommendations for Sukku Coffee and Space in determining product promotions based on the results of processing sales transaction datasets, namely the purchase promotion of Sukku Aren juxtaposed with Choco Berry.

References

- Abaya, SA, Association rule mining based on a priori algorithm in minimizing candidate generation, *International Journal of Scientific and Engineering Research*, vol. 3, no. 7, 2012.
- Asroni, Fitri, H., & Prasetyo, E, (2018), Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus : Fakultas Kedokteran dan Ilmu Kesehatan dan fakultas Ilmu Sosial dan Ilmu Politik), *Semesta Teknika*, 2018
- Diwate, R, B., dan Sahu, A, (2014), Data mining techniques in association rule : A review, *International Journal of Computer Science and Information Technologies*, 5(1), pp. 227–229, 2014
- Fajrin, A, A., dan Maulana, A., Penerapan data mining untuk analisis pola pembelian konsumen dengan algoritma fpgrowth pada data transaksi penjualan spare part motor, *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, vol. 5, no. 1, pp. 27–36, 2018.
- Fauzanu, A., Darwiyanto, E., dan Wisudiawan, G, A, A., Analisis web usage mining menggunakan teknik k-means clustering dan association rule (studi kasus: Www.owlexa.com), *E-proceeding of engineering*, pp. 3284–3291), 2017.
- Fitria, R., Nengsih, W., and Qudsi, DH, Implementation of the fpgrowth algorithm in determining the pattern of traffic accident relationships, *Journal of Information Systems*, vol. 13, no. 2, 2017.
- Ghuman, S, S., Clustering techniques- a review, *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 5, pp. 524–530, 2016.
- Gupta, S., dan Mamtora, R., A survey on association rule mining in market basket analysis, *International Journal of Information and Computation Technology*, vol. 4, no. 4, pp. 409–414, 2014
- Hermanto, T, I., Implementasi algoritma association rule dan k-means sebagai sistem rekomendasi produk pada website penjualan online, pp. 1-5, 2015.
- Kori, A., Comparative study of data classifiers using rapidminer, *International Journal of Engineering Development and Research*, vol. 5, no. 2, pp. 1041–1043, 2017.
- Liu, X., Zhang, X., Li, X., dan Sun, Z., Research on data mining clustering algorithm in cloud computing environments, *Bio-Technology An Indian Journal*, vol. 10, no. 17, pp. 9563–9566, 2014.
- Mustakim, Digital mapping and grouping of green areas in the province of Riau based on knowledge discovery in databases (KDD) with the k-means mining technique, *National seminar on information communication technology and industry (sntiki)* 4, 2012.

- Rabiqy, Y., Strategi peningkatan kepuasan pelanggan untuk meningkatkan loyalitas pelanggan pengguna internet telkomsel di banda aceh, *Prosiding forum keuangan dan bisnis indonesia*, pp. 253–264, 2017.
- Ramageri, B, M.,, Data mining techniques and applications, *Indian Journal of Computer Science and Engineering*, vol. 1, no. 4, pp. 301–305, 2010.
- Raval, UR, and Jani, C., Implementation and improvisation of k-means clustering algorithm, *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 5, 2016.
- Riyono, dan Budiharj, G, E, Pengaruh kualitas produk, harga, promosi dan brand image terhadap keputusan pembelian produk aqua di kota pati, *JURNAL STIE SEMARANG*, vol. 8, no. 2, pp. 92–121, 2016.
- Santosa, S., dan Jadi, Analisis keranjang pasar untuk rekomendasi produk (consumer good) menggunakan fp-growth dengan klasterisasi clarans, *Jurnal Teknologi Informasi*, vol. 12, no. 2, pp. 103–115, 2016.
- Shih, S, P., Yu, S., dan Tseng, H, C., The study of consumers' buying behavior and consumer satisfaction in beverages industry in tainan, taiwan, *Journal of Economics, Business and Management*, vol. 3, no. 3, pp. 391–394, 2015.
- Sidhu, S., Meena, U, K., Nawani, A., Gupta, H., dan Thakur, N, Fp growth algorithm implementation, *International Journal of Computer Applications*, vol. 93, no. 8, pp. 6–10, 2014.
- Soni, R, K., Gupta, N., dan Sinhal, A., An fp-growth approach to mining association rules, *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 2, pp. 1–5, 2013.
- Sumali, E, A., Pengaruh perceived quality dan push strategy terhadap rekomendasi minyak gosok sanhong cap pida, *Jurnal Manajemen Pemasaran Petra*, vol. 2, no. 1, pp. 1–6, 2014.
- Triyanto, WA, Suhartono, V., and Himawan, H., Market basket analysis using k-medoids and fp-growth, *Journal of Pseudocode*, vol. 2, no. 1, 2014.

Biographies

Muhammad Riezky A. L. is graduated from Universitas Sumatera Utara with Bachelor of Engineering degree in Industrial Engineering. He is currently running his own coffee shop business and taking a Master's Degree at Magister Management in Universitas Prasetiya Mulya. His research interest include feasibility study, digital marketing business intelligence and data mining.

Meilita Tryana Sembiring received her Doctoral from Institut Pertanian Bogor (IPB) in 2015 and Masters Degree from Institut Teknologi Bandung (ITB) in 2004. Received her bachelor degree in Industrial Engineering and is currently an active lecturer since 1997 at the Department of Industrial Engineering, Engineering Faculty, Universitas Sumatra Utara. She is the Chairman of the Industrial Engineering Cooperation Agency (BKSTI) for North Sumatra - Aceh Region and is a member of the ISLI (Indonesian Supply Chain Association). Her main research interest is in Supply Chain Management (SCM). In one of her studies, Palm Oil Based Biodiesel Supply Chain Process Business Model, He has written dozens of Scopus-indexed research articles on processing in the palm oil sector and other fields. As a Principal Researcher at USU, she has received a total of several million grants from internal and government grants from her participation in community service activities. Her administrative experience shines even more when she was appointed as Chair of the Industrial Engineering Department for a period of five years from 2017 to 2022 and has successfully obtained coaching certification in 2020. Currently there are several PhD graduates and quite a number of Masters or other postgraduate who are currently working under her watch. She has also participated in the counterpart training of the Higher Education Development Support (HEDS) Project in Indonesia at Toyohashi University of Technology, Nagoya, Japan. She was also selected as a Program Examiner for Candidates for Civil Servants (CPNS) at several local universities.

Richo Giwana Resdy Maulana is an assistant professor in Industrial Engineering and Magister Management at Universitas Sumatera Utara. He graduated from Universitas Sumatera Utara with Bachelor of Engineering degree in Industrial Engineering. He is an assistant laboratory Computation at Industrial Engineering Universitas Sumatera Utara. His research interest includes data mining, operations management, optimization and simulation.