Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Customer Profilling for Precision Marketing using RFM Method, K-MEANS algorithm and Decision Tree

Sularso Budilaksono[1)*], Jupriyanto[2)], M.Anno Suwarno[3)], I Gede Agus Suwartane[4)], Lukman Azhari[5)] , Achmad Fauzi[6)] , Mahpud[7)] , Novita Mariana[8], Maya Syafriana Effendi[9))]

[1,3,4,9)]Persada Indonesia YAI University, [2]STMIK Nusa Mandiri, [5,6,7)] Muhammadiyah Tangerang University, [8)]Stikubank University

[1)]sularso@upi-yai.ac.id, [2)] jupriyanto.kahar@gmail.com, [3)] suwarno@upi-yai.ac.id, [4)] agus.suwartane@gmail.com, [5)] lukmanazhari85@gmail.com , [6)] ac.fauzi25@yahoo.com, [7)] mahpud18@gmail.com, [8)]novita_mariana@edu.unisbank.ac.id, [9)]mayasyafriana@gmail.com

Submitted : Oct 1, 2021 | **Accepted** : Oct 15, 2021 | **Published** : Oct 25, 2021

Abstract

Precision marketing is the companys ability to offer products specifically made to customers. This decision can give the company the ability to attract customers to always buy continuously. This study presents a trend model for accurately predicting monthly supply quantities / The method used in the first stage is the RFM (Recency, Frequency, Monetary) method for selecting attributes to group customers into different groups. The output of the first stage is clustered using the K-Means Algorithm. The output of clustering is then classified using the Decision Tree and compared with the K Nearest Neighbor method. The dataset that is processed is sales data from Syifamart As-Syifa Boarding School in Subang with 351,158 rows of data. The clustering process produces 4 optimal clusters. The four clusters are then classified using the Decision Tree algorithm to determine the potential and non-potential characteristics of each customer.

Keywords: Precision Marketing; Customer Profiling; RFM (Recency, Frequency, Monetary); K-Means Algorithm; Decision Tree Algorithm, K Nearest Neighbor Algorithm.

INTRODUCTION

In doing business, companies need marketing strategies that are suitable and effective with the business they are running. It aims to attract more customers and increase sales so that they can bring profits and success. The accelerating pace of economic globalization and increasing market competition, economic pressures and business competition have caused companies to face the problem of choosing strategic decision-making policies. This strategy decision is in the form of the right way to sell the right product to the right customer at the right time so that the company can increase their profit.

Customers are assets that are very important for the sustainability of the company. Companies must plan and use clear strategies to enforce customers. For this purpose, companies can group customers. Customer grouping known as customer segmentation or customer profile is very useful for the company.

There are large data sources in the company that can be used for customer grouping and profiles. Historical data in the past can be referred to as data mining. Data mining is part of knowledge discovery data which is a data information extraction process to develop models used to understand the phenomenon of data analysis and prediction. Based on the background of the above problems, the formulation of the problem in this study is how to find the algorithm method, to produce a framework for precision marketing decision making using the RFM Method, K-Means Algorithm and Decision Tree.

Syifamart is one of the Business Units that supports programs, activities and activities at the As-Syifa Al-Khoeriyyah Foundation. At present Syifamart already has a Sales System that is integrated with Smartcards for its payment methods. Syifamart's location is in the vicinity of the Junior and High School Educational Boarding School complex. The problem faced by the manager of Syifamart is that in the daily lives of many dormitory

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

complexes there are still many who shop outside Syifamart, preferring to shop at Indomaret and Alfamart outlets around the campus complex location. From this problem a marketing model is needed that can attract potential SyifaMart customers.

LITERATURE REVIEW

In precision marketing, the principle is the same as utilizing the right technological foundation and data analysis capabilities to design and implement marketing campaigns. Its strength is the ability to deliver accurate and precise marketing messages to people at a narrow customer segment

Clustering is the process of grouping a set of physical or abstract items into classes of similar items in which groups are meaningful or useful, or both. The well-known grouping algorithm is K-means, which was first proposed by (Breiman, 2017) The accuracy of this algorithm depends on the initialization and number of clusters (Mesforoush, A., & Tarokh, M. J., 2013).

The basic idea of K-means is to find the k cluster, so that the records in each cluster are similar to each other and different from the notes in other clusters. K-means is an iterative algorithm: the initial collection of clusters is defined and clusters are repeatedly updated until no further improvements are possible (or the number of iterations exceeds the specified threshold).

In this learning algorithm, computers classify the data themselves as input without knowing the target class first. This learning is included in unsupervised learning. The input received is the data or object and the desired k group (cluster). This algorithm will group data or objects into k groups of groups. In each cluster there is a center point (centroid) that represents the cluster.

Decision trees are efficient data mining algorithms with strong explanatory capabilities (Zhang, D., Zhou, X., Leung, S. C., & Zheng, J., 2010). The J48 algorithm is the development of the most common conventional induction tree decision algorithm, ID3. This algorithm, which is a development of ID3, can classify data using a decision tree method that has the advantage of being able to process numerical (continuous) and discrete data, can handle missing attribute values, generate easily interpreted rules, and fastest among algorithms that use main memory on the computer. In applying several cases of classification techniques, this algorithm is able to produce good performance. With this advantage, it is expected that this algorithm can handle case studies optimally and it is also expected that this algorithm will produce good accuracy and performance.

The RFM model was proposed by (Iriana, 2007). This model is very popular in customer value analysis and has been widely used in measuring customer age values (Cheng, C. H., & Chen, Y. S., 2009) and in customer segmentation and behavior analysis (Chen, D., Sain, S. L., & Guo, K., 2012). The RFM model has also been used in some cases, especially in selecting grouping indices. Recently, researchers have used RFM attributes and K-means methods to improve customer relationship management (CRM) for companies (Cheng, C. H., & Chen, Y. S., 2009) (Wei, J. T., Lee, M. C., Chen, H. K., & Wu, H. H., 2013). In this study, we use the RFM model to select variables for grouping so that grouping standards are objectively determined.

The RFM segmentation model is a model that distinguishes important customers according to three variables: R represents "recency", which is defined as the interval between the current and current consumption behavior; the shorter the interval, the greater the R. F represents "frequency", which is defined as the frequency of consuming behavior over a period of time. M represents "monetary", which is defined as the value of consumption money over a period of time.

METHOD

The research methodology used in this study using the Cross-Industry Standard Process for Data Mining (CRISP-DM) method consists of six stages namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment. Figure 1 explains the stages in the CRISP-DM method.

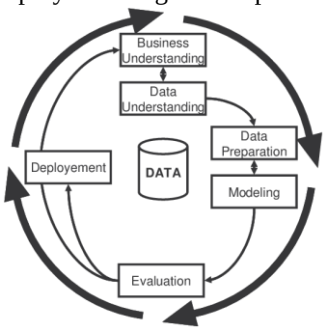

Fig 1. Research Methodology.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The first stage of CRISP-DM is Business Understanding. At this stage the researcher makes one question. The question is about the purpose of this study. The answer is the creation of a precision marketing decision-making framework by conducting customer grouping and clustering using the Recency Frequency and Monetary (RFM) method and the K-means Algorithm and using trend prediction methods to maintain product stock availability.
Syifamart is one of the Business Units that supports programs, activities and activities at the As-Syifa Al-Khoeriyyah Foundation. Currently Syifamart has a system that is integrated with Smartcards for its payment methods. Syifamart's location is around the Junior and Senior High School Boarding School complex, with the hope that the needs of the dormitory are fulfilled in Syifamart starting daily snacks, school needs, office needs to monthly household needs, no need to go out of the institutional complex to meet the internal needs of the institution.

This *study uses data taken from the Syifamart sales system for 6 months, period 1 July 2018 until 31 December 2018. The dataset for 6 months is 351.158 records based on existing sales transactions. The sample sales dataset can be seen in Table 1.*

Table 1. Sample History Sales Data

id	transacID	Product_code	Product_name	qty	Transact_ date	price	Cust_ID
1	J-2018-220429	8996001304983	slay olay blubery	20	7/1/2018	1,000	PLG001
2	J-2018-221862	kp3	keripik pusaka 3000	1	7/6/2018	3,000	PLG001
3	J-2018-221862	8886013221203	Twistko jagung bakar 80gr	1	7/6/2018	6,000	PLG001
4	J-2018-221862	8992761166229	frestea markisa 350ml	1	7/6/2018	4,000	PLG001
5	J-2018-222284	8992696428287	millo combo pack 32g	3	7/7/2018	7,500	PLG001
6	J-2018-220398	8995103200049	axo 600ml	1	7/1/2018	2,500	UMUM
7	J-2018-220400	8999999045265	rinso molto 900g	1	7/1/2018	19,000	UMUM
8	J-2018-220400	8999999041892	molto pink all in one pouch 300ml	1	7/1/2018	9,500	UMUM
9	J-2018-220402	8998866501026	sedaap minyak greng 1liter	1	7/1/2018	13,500	UMUM
10	J-2018-220402	8999999514006	Kecap Bango 60 ml	1	7/1/2018	3,000	UMUM
11	J-2018-220402	8996001301142	Roma Klapa	1	7/1/2018	8,500	UMUM
12	J-2018-220402	g37	garam 37 200gr	1	7/1/2018	2,000	UMUM
13	J-2018-220402	8998009700392	campina sponge bob cup80ml	1	7/1/2018	4,500	UMUM
14	J-2018-220402	8991002103238	Goodday Mocacino	2	7/1/2018	1,250	UMUM
15	J-2018-220404	8992761110017	Frestea Melati cup	1	7/1/2018	2,000	UMUM
16	J-2018-220406	8998866610087	teh javana 350ml	7	7/1/2018	3,000	UMUM
17	J-2018-220408	kuramas	buku kuramas	1	7/1/2018	5,500	UMUM
18	J-2018-220408	Alm	al matsurat	1	7/1/2018	3,000	UMUM
19	J-2018-220410	Batk	Batre Abc Kecil	4	7/1/2018	2,500	UMUM
20	J-2018-220412	8995103200049	axo 600ml	1	7/1/2018	2,500	UMUM

RESULT

In this section, this study presents a trend model for predicting monthly supply quantities. This trend is indicated by changes in the average value of sales of "3000 heirloom chips" during the period of 1 July 2018 to 31 December 2018. Following are the sales trends for 6 months without seeing who is shopping, both members and non members:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 2. The 6-month trend of selling products "3000 heirloom chips"

Period	Amount of Transaction
2018-07	1,773
2018-08	3,231
2018-09	3,483
2018-10	1,749
2018-11	1,597
2018-12	1,239

From the table above, a description of the data will be made in the form of a polygon so that it can easily analyze the data. The following is a polygon of data from the results of the sale of the product "3000 heirloom chips"

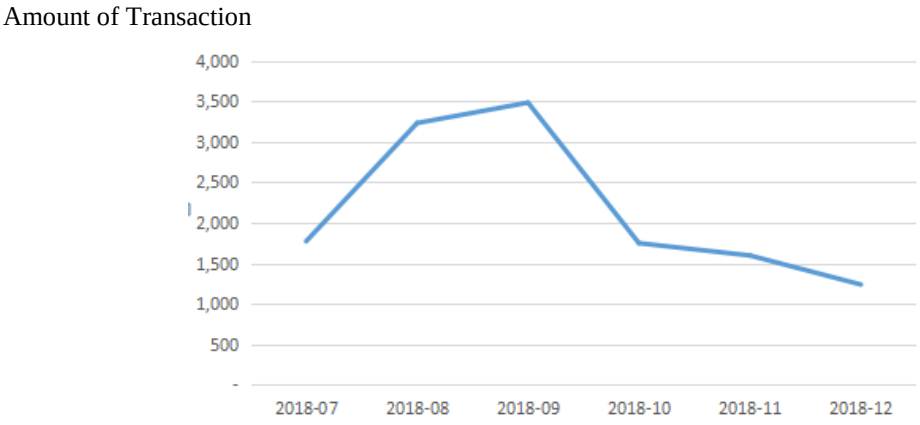

Fig 2. Graph of sales transaction polygons

From the polygon graph above, it can be seen that every month the sale of "3000 heirloom chips" products fluctuates in terms of the number of transactions, the peak of the most transactions is in September 2018.

To classify this research customer focuses on customers who have previously shop for products at Syifamart. Therefore, we pre-process the data before building a decision-making model, namely:

1. Dispose of Transaction Data that is not from a member, this research focuses on transactions carried out by members who have an RFID Smartcard.
2. Determining the products to be researched, this study will focus on researching "heirloom 3000 products" with total purchase transactions by members totaling 6,744 transactions and total transaction nominal amounting to Rp. 22,424,400.
3. Next, we choose the attributes that will be used in the data mining procedure of this study, namely the RFM attribute, so that data is obtained according to table 9 below:

Table 3. RFM Attribute Data

Number	Customer			RFM		
	Name	Class	ID	R (days to)	F(Frequency)	M (Rp)
1	HILMY KUMARA SANTOSO	8 ABU BAKAR (WANAREJA)	2017300006	42	14	42,000
2	RAID LENDRA PRINGGANDANI	8 ABU BAKAR (WANAREJA)	2017300008	31	18	60,000

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

3	MOHAMMAD LANTIPTRENGGINAS	8 ALI BIN ABI THALIB (WANAREJA)	2017300010	13	20	60,000
4	IRFAN RANGGA MIFTAHURRIZQI	8 ALI BIN ABI THALIB (WANAREJA)	2017300018	13	46	138,000
5	RAKHA RAIHANURRAHMAN	8 UTSMAN BIN AFFAN (WANAREJA)	2017300022	53	10	33,000
6	MUHAMMAD SALMAN GHIFFARY	8 ABU BAKAR (WANAREJA)	2017300027	13	10	30,000
7	MUHAMMAD ATHALLAH GHOZY SOEMANTRI	8 UMAR BIN KHATTAB (WANAREJA)	2017300029	17	33	99,000
8	MUHAMMAD NAFI`AN RASYID SATRIO	8 UMAR BIN KHATTAB (WANAREJA)	2017300036	11	32	96,000
9	RANTISI MUHAMMAD YASIN	8 UMAR BIN KHATTAB (WANAREJA)	2017300043	28	26	93,000
10	MAULANA HAMID HALIM	8 ABU BAKAR (WANAREJA)	2017300045	11	34	105,000
11	FAARIS RIDLO IQBAL	8 ABU BAKAR (WANAREJA)	2017300046	125	3	12,000
12	FITRA NUR RAMADHAN	8 ABU BAKAR (WANAREJA)	2017300061	11	30	93,000
13	RASHIF AHMAD NURFUADI	8 UTSMAN BIN AFFAN (WANAREJA)	2017300068	49	26	81,000
14	MUHAMMAD AZMI NASHIRUL HAQ	8 UTSMAN BIN AFFAN (WANAREJA)	2017300074	27	1	3,000
15	MUHAMMAD FAATIH DZULQARNAIN	8 UMAR BIN KHATTAB (WANAREJA)	2017300075	103	20	60,000
16	MUHAMMAD THORIQ RIZQI	8 UMAR BIN KHATTAB (WANAREJA)	2017300084	46	23	72,000
17	DAFFA HAFIZH MUSYAFA	8 UMAR BIN KHATTAB (WANAREJA)	2017300087	28	6	18,000
18	NAUFAL NADHIF RASYID	8 UMAR BIN KHATTAB (WANAREJA)	2017300102	132	1	3,000
19	MUHAMAD RIZA AZMI SAFRUDIN	8 ALI BIN ABI THALIB (WANAREJA)	2017300105	19	58	174,000
20	MALIK HAFIDZH AL RIZALI	8 UTSMAN BIN AFFAN (WANAREJA)	2017300112	39	9	27,000

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In this phase directly involves data mining techniques, namely by selecting data mining techniques and determining the algorithm to be used. After all transaction data is converted into numbers, the data can be grouped using the K-means method algorithm. To be able to group data into several clusters it is necessary to do the following steps:

a. Determine the desired number of clusters. In this study the existing data will be grouped into four groups.

b. Determine the initial center point of each cluster. In this study the initial center point is determined randomly and the center point of each cluster can be seen in table 4

Table 4. The initial center point of each cluster

Initial of cluster centroid	Customer_ID	R	F	M
CLUSTER 1	2017300029	5	3	3
CLUSTER 2	2017300312	5	5	5
CLUSTER 3	2017300375	5	2	2
CLUSTER 4	2017300668	4	2	2

Place each data in the cluster. In this study, the k-means method is used to allocate each data into a cluster that is closest to the center point of each cluster. To find out which cluster is closest to the data, it is necessary to calculate the distance of each data with the center point of each cluster. For example, calculate the distance from customer transaction data = 2017300027 to the first cluster center:

$$D(i,j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \cdots + (XKi - XKj)^2}$$

diperoleh jarak ke masing-masing cluster adalah :

1. D(C1) = 2.828427125
2. D(C2) = 5.656854249
3. D(C3) = 1.414213562
4. D(C4) = 1.732050808

From the calculation above, the closest distance of the 2017300027 customer is to cluster 3 (1.414213562), so that 2017300027 members are entered into cluster 3. The calculation results complete with the initial cluster center can be seen in table 13.

Table 5. The results of the calculation of each data to each step initial cluster

Number	CustomerID	R	F	M	Distance to C1	Distance to C2`	Distance to C3	Distance to C4	Close to
1	2017300006	5	2	2	1.414213562	4.242640687	0	1	C3
2	2017300008	5	2	2	1.414213562	4.242640687	0	1	C3
3	2017300010	5	2	2	1.414213562	4.242640687	0	1	C3
4	2017300018	5	4	4	1.414213562	1.414213562	2.828427125	3	C1
5	2017300022	4	1	1	3	5.744562647	1.732050808	1.414213562	C4
6	2017300027	5	1	1	2.828427125	5.656854249	1.414213562	1.732050808	C3
7	2017300029	5	3	3	0	2.828427125	1.414213562	1.732050808	C1
8	2017300036	5	3	3	0	2.828427125	1.414213562	1.732050808	C1
9	2017300043	5	3	3	0	2.828427125	1.414213562	1.732050808	C1
10	2017300045	5	3	3	0	2.828427125	1.414213562	1.732050808	C1
11	2017300046	3	1	1	3.464101615	6	2.449489743	1.732050808	C4

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

12	2017300061	5	3	3	0	2.828427125	1.414213562	1.732050808	C1
13	2017300068	5	3	3	0	2.828427125	1.414213562	1.732050808	C1
14	2017300074	5	1	1	2.828427125	5.656854249	1.414213562	1.732050808	C3
15	2017300075	4	2	2	1.732050808	4.358898944	1	0	C4
16	2017300084	5	2	2	1.414213562	4.242640687	0	1	C3
17	2017300087	5	1	1	2.828427125	5.656854249	1.414213562	1.732050808	C3

After all data is placed into the nearest cluster, then recalculate the new cluster center based on the average member in the cluster. After getting a new center point from each cluster, returning from step c to the center point of each cluster does not change again and no data moves from one cluster to another cluster.
From the results of data processing based on customer transaction datasets using K-Means through 4 iterations, a cluster is formed as shown in Figure 3 which shows that clustering results obtained 65 members of cluster C1, 21 members of cluster C2, 294 members of C3 and 232 clusters C4 cluster member.
From the results of data processing based on customer transaction datasets using K-Means through 4 iterations, a cluster is formed as shown in Figure 5 which shows that clustering results obtained 65 members of cluster C1, 21 members of cluster C2, 294 members of C3 and 232 clusters C4 cluster member.

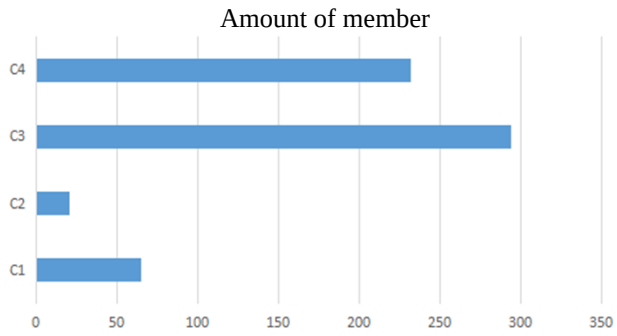

Fig 3. Results of Clustring using K-Means

DISCUSSIONS

From the results of grouping, it is necessary to test cluster validity by calculating the Davies Bouldin Index to determine whether the number of clusters that are formed is optimal or not. Clusters are convergent if there are no changes or movements of members from one cluster to another cluster. In addition, the convergent cluster is also marked with no changes in the DB index values. Of the five test scenarios that have been carried out, the ranking will be made based on the best index-DB value. The optimal level of a cluster can be measured by the DB Index Value. Davies and Bouldien said that the best number of clusters is that which has the index closest to 0 among other groups. The following are the results of calculating the DB Index value using Rapidminer Studio 9.1

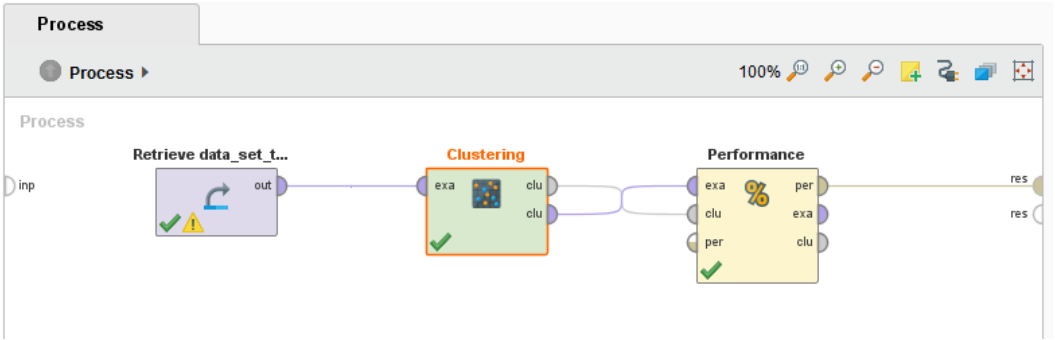

Fig 4. Design checking performance clusters

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

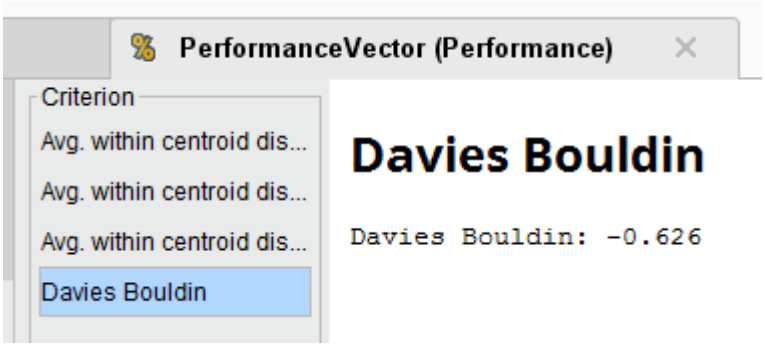

Figure 5. Davies Bouldin 2 Cluster

The formation of the classification model is carried out by implementing data into Rapidminer Studio 9.1 with the dataset being the results of clustering using K-Means, which amount to 612 data resulting in a model tree and rules for determining customer characteristics.
The results of the classification model can be seen as a tree in Figure 4 and this classification model involves 4 attributes namely Recency, Frequency, Monetary (RFM) and Cluster.

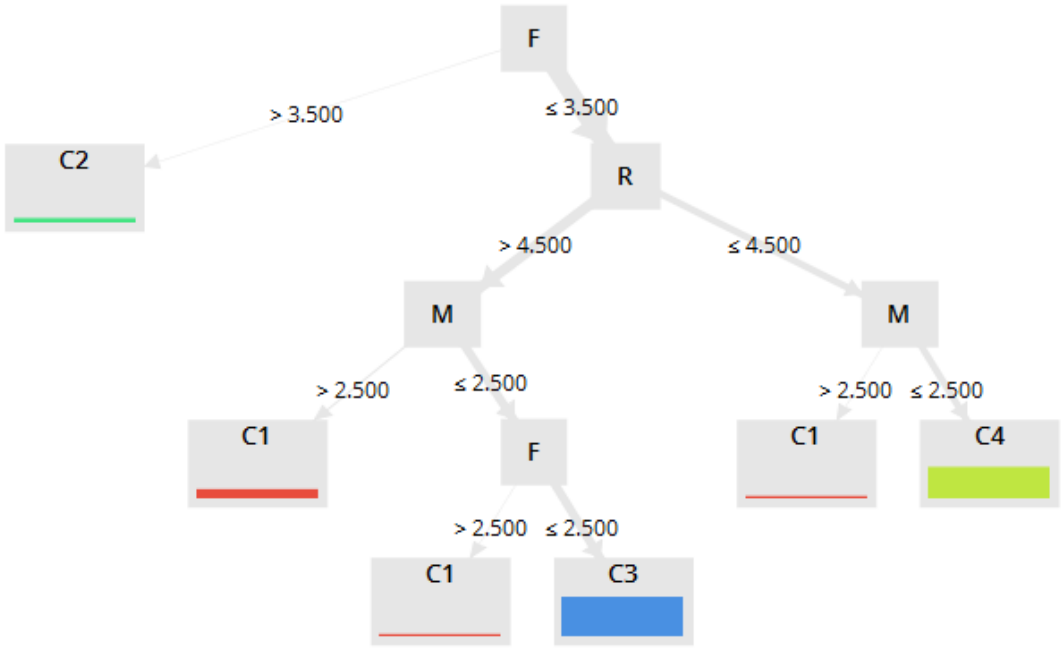

Fig 6. Model Decision Tree.

accuracy: 98.86% +/- 1.05% (micro average: 98.86%)

	true C3	true C2	true C4	true C1	class precision
pred. C3	294	0	0	3	98.99%
pred. C2	0	21	0	0	100.00%
pred. C4	0	0	232	4	98.31%
pred. C1	0	0	0	58	100.00%
class recall	100.00%	100.00%	100.00%	89.23%	

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, October 2021
DOI : https://doi.org/10.33395/sinkron.v6i1.11225
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig 7. Accuracy and Description of the Model Decision Tree

Customer profiles are carried out using rules generated from the previous process. The customer characteristics used in the customer profile are characteristics according to Larose, D. T. Table 6 is the customer characteristics generated in this study.

Table 6. Characteristics of Customers Each Cluster

Customer type	Cluster	Customer Characteristic
Repetitive Customer	1	Customers in Cluster C1 are referred to as permanent customers, namely customers who have purchased a product several times and irregularly purchase it.
Client	2	Customers in Cluster C2 are referred to as Client customer types. They buy regularly, and relationships with these types of customers have been strong and will last a long time which makes them unaffected by competitors' attraction from other products.
Advocate	3	Customers in the C3 Cluster make regular purchases, with a limited amount. This is very profitable for the company. This type of customer can also be used as a marecting company because this type of customer is marketing by encouraging their friends to buy goods / services in the company.
First Time Customer	4	Customers in the C4 Cluster have characteristics of low intensity levels. They are new customers whose arrival rates are very low. Customers of this type are customers who buy for the first time, they are still new customers.

The rules in table 6 will facilitate the Syifamart marketing team or management in classifying customers for future marketing needs.

CONCLUSION

The process of data mining consists of sales transaction history of 351.158 rows. Then aggregation is based on the customer using the RFM method. After obtaining RFM data the process was continued with extraction using the K-Means clustering algorithm so that 4 (four) optimal clusters were formed. The four clusters are then classified using a decision tree algorithm to know each customer characteristic so that the company can know which potential customers and potential customers. To maintain the availability of stock supply, Syifamart's management can predict the inventory needs of the product. This study uses a trend method where sales / stock in the following month can be predicted by using the sales history in the previous month. To facilitate the company / management in making decisions the results of this study are implemented by making an application tool. This application is expected to be used by marketing and management to see the results of clustering and can easily make recommendations for the best marketing model based on the results of clustering and customer profiles.

REFERENCES

Breiman, L. (2017). *Classification and Regression Trees.* New York: Routledge.

Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: Acase study of RFM model-based customer segmentation using data mining. .*Journal of Database Marketing & Customer Strategy Management, 19(3)*, , 197–208.

*name of corresponding author

Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications, 36(3), 4176–4184*.

Iriana, R. (2007). Strategic, Operational, and Analytical Customer Relationship Management. *ournal of Relationship Marketing*.

Mesforoush, A., & Tarokh, M. J. (2013). Customer profitability segmentation forSMEs case study: network equipment company. .*International Journal of Researchin Industrial Engineering*, 30–44.

Wei, J. T., Lee, M. C., Chen, H. K., & Wu, H. H. (2013). Customer relationshipmanagement in the hairdressing industry: An application of data miningtechniques. .*Expert Systems with Applications, 40(18), 7513–7518*.

Zhang, D., Zhou, X., Leung, S. C., & Zheng, J. (2010). Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications 37(12),7838–7843*.

*name of corresponding author