

K-Medoids Customer Segmentation Algorithm by Utilizing Customer Relationship Management

Imam Shodako Afari ¹

¹Putra Indonesia University YPTK Padang
imamshodakoafari@gmail.com

Abstract

Sales is an activity that needs to be considered in the company, because the end of production is sales. The more sales that occur, the need for a system that can make it easier to process sales. At this time, motorbikes are one of the profitable sales to market. Central Motor Yamaha is present as a dealer that sells various motorcycle products, with the cheapest and original product claims. Customer segmentation is used at the Lubuk Begalung Yamaha Motor Center to group customers, which are processed from sales transaction data. By utilizing Customer Relationship Management (CRM) provides interaction between dealers and customers to increase sales. CRM comes in the form of a system that makes it easier for customers to make purchases, which has various features in it, of course it makes it very easy for customers to use. One method that can be used in CRM is data mining. With the k-medoids method clustering technique produces priority and ordinary customer clusters. Problems at Sentral Motor Yamaha Lubuk Begalung can be resolved using this method, of course it is useful for the company, namely attracting customers with the various features it contains. By implementing the K-Medoids grouping technique, it is hoped that it will become a customer relationship management strategy with the aim of attracting new customers.

Keywords: Customer Segmentation, Customer Relationship Management, K-Medoids

JCSITech is licensed under a Creative Commons 4.0 International License.

1. Introduction

With advances in information and communication technology, it has made big changes to life, including companies or organizations [1]. Information technology makes it easier to work on various documents, reports and other correspondence so that with information technology everything can be completed as effectively and efficiently as possible [2].

The need for quality information is the main factor why the use of information technology needs to be planned. Therefore, there is a need for governance that regulates how information technology is used in the organization. The existence of Customer Relationship Management can reduce the error rate when collecting data and followed by increased business output [3].

Customer segmentation is a key for companies in improving relationships with customers. Without this information, the company will still apply the same treatment to all customers. One of the strategies implemented is Customer Relationship Management (CRM). The CRM concept allows companies to identify customers by segmenting customers. The purpose of the customer segmentation process is to find out customer behavior and implement the right marketing strategy so that it brings benefits to the company [4].

CRM is a method of maintaining customer attraction and how about customer choice, which is an activity in

recruiting customers to use the company's products in the long term [5]. Customer Relationship Management (CRM) is a process for obtaining, maintaining and enhancing profitable relationships with customers in order to create value and customer satisfaction and maximize company profits related to competitive advantage and pay attention to product quality so that customers can achieve excellent satisfaction [6].

The data mining technique used to find customer segmentation is the K-Medoids Clustering technique. The K Medoids method is used because it can analyze sales history data and group customer types well. Researchers will produce a software that is made using methods in managing customer data, so that it will make it easier for the Lubuk Begalung Yamaha Motor Center to segment customers. With this CRM system, it is hoped that it can help and improve service to customers, making it easier and attracting customers to place orders [7].

The K Medoids method is a clustering algorithm that is almost the same as the K-Means algorithm. The difference between these two algorithms is that the K-Medoids algorithm uses a representative object (medoid) as the cluster center for each cluster, while K-Means uses the average value (mean) as the cluster center [8]. K-medoid has the same shortcomings as k-means. Where initializing the initial cluster greatly influences the cluster results. Because this method immediately takes randomly as many clusters as

desired by the user [9]. The K-Medoids algorithm has the advantage of overcoming the weakness of the K-Means algorithm which is sensitive to noise and outliers, where objects with large values may deviate from the data distribution [10].

Sentral Motor Yamaha Lubuk Begalung is one of the Yamaha motorbike dealer branches in the city of Padang. The goods sold include various brands of Yamaha motorbikes such as Nmax , Vixion, Mio and others. This research needs to be able to produce an example program that can cluster or group customer data. Customer segmentation results can be used to provide customer assessments and determine customer profiles more precisely.

2. Research methodology

To assist in the preparation of this research so that the steps in solving the problems to be discussed can be clearly structured, it is necessary to have a framework arrangement. The research framework is in Figure 1 below:

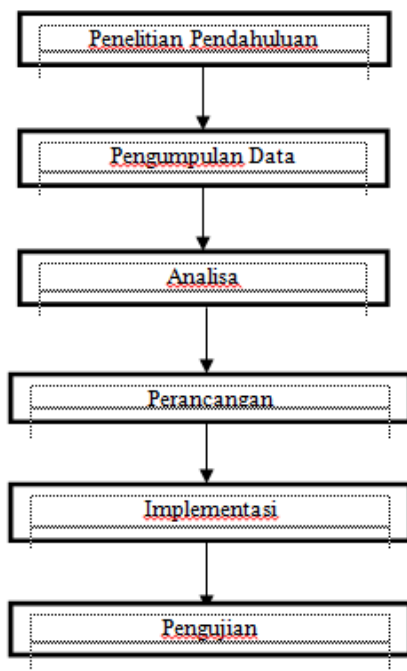


Figure 1. Research Framework

Research stages are a sequence of processes or steps that will be carried out in completing this research. The stages of this research are as follows:

2.1 Preliminary Research

This research was conducted by interviewing the central manager of Yamaha Lubuk Begalung regarding problems with customer segmentation, in order to analyze the problem and obtain the information needed by the researcher. then carry out observation activities related to the research object directly at the Yamaha Lubuk Begalung central location, such as collecting data, observing research objects, etc.

2.2 Data collection

In collecting data sources it is very important in research to obtain data source data in this study the authors obtain data from various sources:

1. Interview

Interviews, namely by asking questions , by conducting questions and answers directly to the Wali Nagari office to obtain the data and information needed.

2. Observation

Observation is by conducting a direct observation at the Wali Nagari Singgulung Padang Pariaman Office to find out the performance problems of employees at the Wali Nagari Singgulung Padang Pariaman Office.

2.3 System analysis

This system analysis stage is carried out to design the system that will be created. The design includes designing the user interface, designing the database for the system so that the existing file management is more organized and designing the coding program of the information.

2.4 Planning

In this design stage, the researcher uses an explanation of the program analysis flow, where UML is used.

2.5 Implementation

System implementation is the stage of putting the system in place so that it is ready to operate. This stage of design uses the PHP programming language and MySQL database.

2.6 Testing

After the application has been designed, a testing process will be carried out on the resulting application to find out whether the application that has been designed is running correctly and in accordance with the formulation of the problem that has been previously defined.

3. Results and Discussion

3.1 K-Medoids Algorithm Calculation

Before calculating the K-Medoids algorithm, we need to look at the data that will be used for research, which can be seen in Table 1 below:

Table 1. Customer Data

No	Name	Purchase Amount	Total
1	Raki Anggara	2	51,500,000
2	Gendi Metra	2	39,500,000
3	Fariuswan	2	47,800,000
4	Imelda	1	21,900,000
5	Ryan Athira	1	31,000,000
6	Yusril Umam	1	17,000,000

7	Fahrul Zaki	1	31,000,000
8	Reby Safwan	1	30,800,000
9	Bayu Ramadhan	1	37,500,000
10	Aafrison Saleh	3	57,500,000
11	Fitriani	1	21,900,000
12	Asep Naldi	1	32,000,000
13	Ahmad Ramadhani	1	30,800,000
14	Rizaldi	1	17,000,000
15	Noviatul Ukhwah	1	22,500,000
16	Andre Gunawan	1	32,000,000
17	Divya Amabel	1	31,000,000
18	Willy Afdilla	1	22,500,000
19	Angga Saputra	1	37,500,000
20	Dafriman	1	18,000,000

At this stage there are several calculation processes that can be seen as follows:

1. Determine *Min* and *Max* Values

Min and *Max* values for the sales and stock variables that need to be looked for and can be seen in the following table:

Table 2 . Min and Max

set	Purchase Amount	Total
Min	1	17,000,000
Max	3	57,500,000

2. Data Normalization

This normalization is used as a reference for calculations in the *K-Medoids algorithm* where the following formula can be seen :

$$v1 = \frac{V - MINa}{MAXa - MINa}$$

Where:

V = Variable

Min = Lowest value

Max = Highest value

Calculation of quantities

$$v1 = \frac{2 - 1}{3 - 1} = 0.5$$

$$v2 = \frac{2 - 1}{3 - 1} = 0.5$$

$$v2 = \frac{2 - 1}{3 - 1} = 0.5$$

Total calculation

$$V1 = \frac{51500000 - 17000000}{57500000 - 17000000} = 0.8522$$

$$V2 = \frac{39500000 - 17000000}{57500000 - 17000000} = 0.5555$$

$$V3 = \frac{47800000 - 17000000}{57500000 - 17000000} = 0.7604$$

Table 3. Normalized Data

No	Buyer Name	Amount	Total
1	Raki Anggara	0.5	0.85224691358025
2	Gendi Metra	0.5	0.55555555555556
3	Fariruswan	0.5	0.76049382716049
4	Imelda	0	0.12098765432099
5	Ryan Athira	0	0.34567901234568
6	Yusril Umam	0	0
7	Fahrul Zaki	0	0.34567901234568
8	Reby Safwan	0	0.34074074074074
9	Bayu Ramadhan	0	0.50617283950617
10	Aafrison Saleh	1	1
11	Fitriani	0	0.12098765432099
12	Asep Naldi	0	0.37037037037037
13	AhmadRamadhani	0	0.34074074074074
14	Rizaldi	0	0
15	Noviatul Ukhwah	0	0.1358024691358
16	André Gunawan	0	0.37037037037037
17	Divya Amabel	0	0.34567901234568
18	Willy Afdilla	0	0.1358024691358
19	Angga Saputra	0	0.50617283950617
20	Dafriman	0	0.02469135802469

3. Define *k* (number of clusters) desired

At this stage it is required to randomly select clusters which can be seen in the following table:

Table 4. Cluster Values

C1	C2
0.5	0.85224691358
0.5	0.55555555556

4. Calculating Distance

At this stage you are required to process the 4. Cluster table with table 3. Cluster normalization of data where there are formulas and tables that can be seen as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

calculation :

$$\sqrt{(0.5 - 0.5)^2 + (0.8522 - 0.8522)^2} = 0$$

$$\sqrt{(0.5 - 0.5)^2 + (0.5555 - 0.8522)^2} = 0.2966$$

$$\sqrt{(0.5 - 0.5)^2 + (0.7604 - 0.8522)^2} = 0.0917$$

calculation :

$$\sqrt{(0.5 - 0.5)^2 + (0.8522 - 0.5555)^2} = 0.2966$$

$$\sqrt{(0.5 - 0.5)^2 + (0.5555 - 0.5555)^2} = 0$$

$$\sqrt{(0.5 - 0.5)^2 + (0.7604 - 0.5555)^2} = 0.2049$$

Whereas for the *Cluster column* the value is obtained from the value of the description column contained in the *COST* column. If the information value is at *COST* 1 then the *Clustering* value is 1. For more details, see table 5.

Table 5. Clustering Values

Buyer Name	C1	C2	Proximit y	Cluster s	Information
Raki Angara	0	0.296691	0	1	Priority

Gendy Metra	0.29 6691	0	0	2	Normal
Fariruswan	0.09 1753	0.20 4938	0.091753	1	Priority
Imelda	0.88 5855	0.66 2457	0.662457	2	Normal
Ryan Athira	0.71 1766	0.54 2262	0.542262	2	Normal
Yusril Umam	0.98 8092	0.74 7424	0.747424	2	Normal
Fahrul Zaki	0.71 1766	0.54 2262	0.542262	2	Normal
Reby Safwan	0.71 5289	0.54 4192	0.542262	2	Normal
Bayu Ramadhan	0.60 8085	0.50 2433	0.542262	2	Normal
Aafrison Saleh	0.52 1374	0.66 8977	0.521374	1	Priority
Fitriani	0.88 5855	0.66 2457	0.662457	2	Normal
Asep Naldi	0.69 441	0.53 3192	0.533192	2	Normal
AhmadR amadhani	0.71 5289	0.54 4192	0.544192	2	Normal
Rizaldi	0.98 8092	0.74 7424	0.747424	2	Normal
Noviatul Ukhwah	0.87 3667	0.65 2835	0.652835	2	Normal
Andre Gunawan	0.69 441	0.53 3192	0.533192	2	Normal
Diva Amabel	0.71 1766	0.54 2262	0.542262	2	Normal
Willy Afdilla	0.87 3667	0.65 2835	0.652835	2	Normal
Angga Saputra	0.60 8085	0.50 2433	0.502433	2	Normal
Dafriman	0.96 6875	0.72 9258	0.729258	2	Normal

From the table above it can be seen that there are 3 that have priority values, namely the names of buyers Raki Anggara, Fariruswan and Aafrison Saleh, and the rest have ordinary status.

3.1 System Testing

of the decision support information system will be explained . An explanation of the system created includes the appearance of the system, control functions in the system, and how to use it. In this sub-chapter, we will explain the use of the per system menu system, starting from the login menu display , its functions and how to use it until it's finished.

1. Data Mining Process Page

The data mining process page after calculations have been carried out by the system that has been created, can be seen in Figure 2 below:

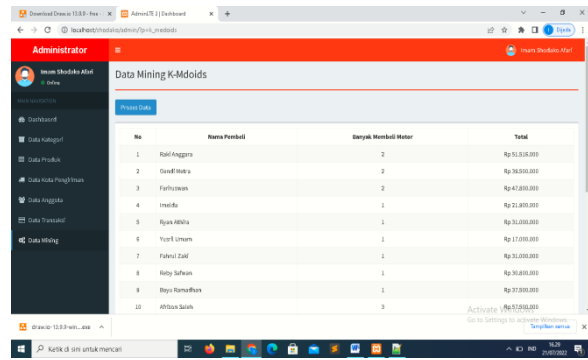


Figure 2 . Page n data mining process

2. Normalization result page

On this page will display the results of data normalization in Figure 3 below:

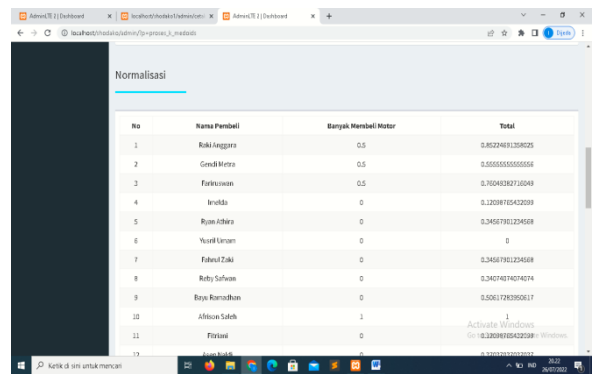


Figure 2 . Normalization result page

3. Results Page Euclidean Distance

On this page will display the results of the Euclidean Distance data.

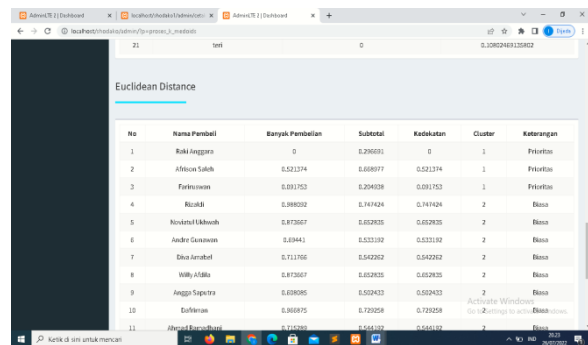


Figure 3 . Results Page Euclidean Distance

4. Conclusion

With the existence of the K-Medoids algorithm system, it can help the Lubuk Begalung Yamaha Motor Center in preparing sales strategies by grouping them into 2 criteria, namely many priorities, and ordinary. Customer segmentation is a group of entities that have similarities and also differences between other entities. This cluster technique works by grouping objects in the data (entities, patterns, events, units, observations) or performing a separation, segmentation, and splitting of data into a certain number of clusters.

References

- [1] Abdillah, LA, Alwi, M., Simarmata, J., Bisyri, M., Nasrullah, N., Asmeati, A., ... & Bachtiar, E. (2020). Information Technology Applications: Concepts and Applications.
- [2] Purba, RA, Sudarso, A., Silitonga, HP, Sisca, S., Supitriyani, S., Yusmanizar, Y., ... & Novita, AD (2020). Information technology applications: theory and implementation. We Write Foundation.
- [3] Rumiarti, CD, & Budi, I. (2017). Customer segmentation in customer relationship management in retail companies: a case study of PT Gramedia Asri Media. *Journal of Information Systems*, 13(1), 1-10. <https://doi.org/10.21609/jsi.v13i1.525>
- [4] Puteri, ET, Kusnanto, G., & Thomas, CJ (2020). Application of K-Means Clustering for Customer Segmentation in the Customer Relationship Management System at Pt. Unichem Candi Indonesia.
- [5] Sobarnas, MA (2020). Comparison of the Accuracy of the K-Nearest Neighbor and C4 Methods. 5 For Customer Relationship Management in Financing Companies. *INFOTECH: Journal of Informatics & Technology*, 1(1), 1-14. <https://doi.org/10.37373/infotech.v1i1.33>
- [6] Siregar, DNP, Cahyani, W., & Chaniago, AU (2020). The Effect of Customer Relationship Management (Crm) on Irian Card (I-Card) User Loyalty at the Irian Dept Store & Medan Marelan Supermarket. *SAINTIKOM Journal (Informatics and Computer Management Science Journal)*, 19(1), 17-23. <https://doi.org/10.53513/jis.v19i1.221>
- [7] Tamaela, J., Sedyono, E., & Setiawan, A. (2017). Cluster analysis uses the Fuzzy C-means and K-means algorithms for clustering and mapping agricultural land in Southeast Minahasa. *Journal of Buana Informatics*, 8(3). <https://doi.org/10.24002/jbi.v8i3.1317>
- [8] Wakhidah, N., & Putri, AN (2020). Application of the K-Medoids Method for Grouping Road Conditions in the City of Semarang. *JATISI (Journal of Informatics Engineering and Information Systems)*, 6(2), 171-180. <https://doi.org/10.35957/jatisi.v6i2.193>
- [9] Rofiqi, AY (2017). Clustering Sports News in Indonesian Using the Conditional K-Medoid Method. *Simantec Journal*, 6(1). <https://doi.org/10.21107/simantec.v6i1.3376>
- [10] Silitonga, DA, Windarto, AP, Hartama, D., & Sumarno, S. (2019, August). Application of the K-Medoid Method in Grouping Households in the Treatment of Sorting Waste According to Province. In *National Seminar on Information Science and Technology (SENSASI) (Vol. 2, No. 1)*.