Application of Data Mining Using the K-Means Clustering Method in Analysis of Consumer Shopping Patterns in Increasing Sales (Case Study: Abie JM Store, Jaya Mukti Morning Market, Dumai City)

Afrido
Student At Industrial Engineering
Universitas Islam Negeri Sultan Syarif Kasim Riau
Indonesia
afridoedo2000@gmail.com

Muhammad Rizki, Ismu Kusumanto, Nazaruddin, Misra Hartati
Assistant Professor Industrial Engineering
Universitas Islam Negeri Sultan Syarif Kasim Riau
Indonesia
muhammad.rizki@uin-suska.ac.id

Fitra Lestari Nohirza
Associated Professor Industrial Engineering
Universitas Islam Negeri Sultan Syarif Kasim Riau
Indonesia
fitra.lestari@uin-suska.ac.id

Abstract

This study applied Data Mining method to cluster sales transactions at Abie JM Stores experienced a decline in sales. Therefore, a strategy was needed to increase sales again. One way that can be done to determine customer needs is to analyze sales transaction data. The sales transaction data can be further processed to obtain more helpful information to increase income, sales, and purchase turnover. Data mining by using k-means grouping or clustering. Data mining can be used to find. solutions in making sales decisions to increase revenue. Sales data storage stores a large number of sales transaction records, where each record provides products purchased by consumers in each sales transaction. From the calculation results, it can be concluded that the k-means clustering method can support the system well. Therefore we need a data processing process using a data mining technique. This study's data collection process uses the interview process and shopping transaction data collection

Keywords
Clustering K-Means, Data Mining, Customer Analysis, Consumers, Transaction Data.

1. Introduction

Based on Table 1 and the Figure 1, it can be seen that the sales of goods at the Abie JM store are less than optimal, where for the most products sold with the number 25 products, further research is needed to increase sales of goods at the Abie JM store.

Tabel 1 Sales Data of Abie JM Store

Name Of Goods	Number Of Transactions	Date
Sovia 1 Liter	1	27/01/2022
Sovia 2 Liter	1	27/01/2022
Sovia 1 Liter	2	27/01/2022
Sovia 2 Liter	2	27/01/2022
Sovia 2 Liter	3	27/01/2022
Sovia 1 Liter	3	27/01/2022
Sovia 2 Liter	4	27/01/2022
Sovia 1 Liter	4	27/01/2022
Gula Merah Besar	5	27/01/2022
***	***	***
Nuvo Nature Protect	100	08/02/2022
Sovia 2 Liter	100	08/02/2022

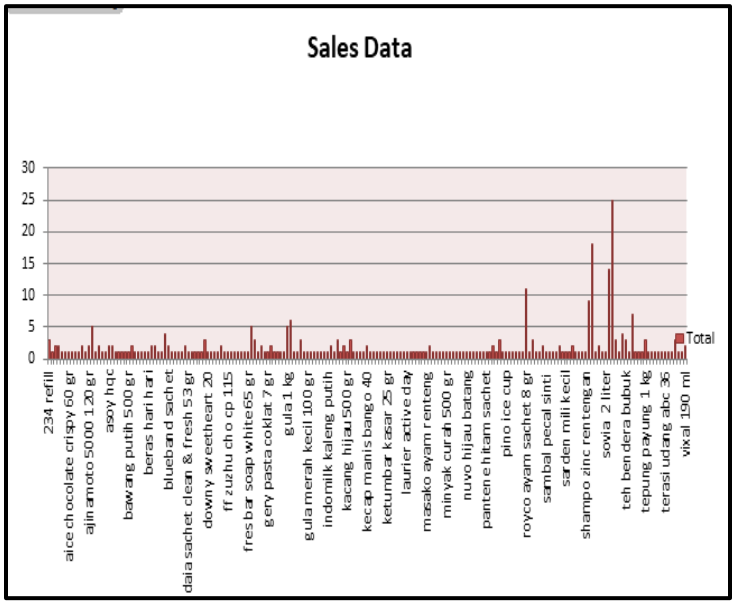

Figure 1. Sales Chart of Abie JM Store Business

1.1.Objectives

The purpose of this final project is to create a cluster system for grouping goods transactions using the K-means algorithm at Abie JM Stores in determining the strategy for determining the stock of goods. This system is expected to be able to assist in providing a stock of goods by way of clustering reports by applying data mining clustering techniques with the K-Means algorithm by grouping the goods sold into 3 criteria, namely: excellent, selling well,

and not selling well, and can solve problems Abie JM shop in determining the strategy of determining the stock of goods.

2. Literature Review

The literature in this study is as follows :

Putra R.R., and Wadisman Cendra (2018), in their article entitled Implementation Of Data Mining For Potential Customer Selection Using K-Means Algorithm, explain how to keep potential MC Laundry customers by providing excellent service and giving gifts. Gifts given by the company to customers are in the form of souvenirs and are carried out consistently every year. Research on the application of Data Mining using the K Means Algorithm has been carried out by many people who are in the field of Information Technology or in other fields. Based on the research objectives of Putra R.R., and Wadisman Cendra, they stated that the purpose of the research was to prevent the displacement of potential customers. The research method used by them is K-means Clustering. The results of their research stated that potential customer data was obtained by using Tanagra tools, data serving customer name data, transaction number data and total transaction data. Potential customer data, obtained after inputting customer data that has been stored in a notepad, then the data is imported into Tanagra. Furthermore, the conclusions and suggestions from their research is to group the data with the K-Means algorithm by determining the number of clusters, calculating the closest distance to the center of the cluster. The data with the closest distance states the members of the cluster, recalculates the data not moving to another cluster to see the objective function. customer data import data includes customer name data, transaction data and total transaction data, and potential customer data, obtained after inputting customer data stored in notepad data then the data into Tanagra.

2.1. Data Mining

According to Widodo, data mining is an analysis of data to find transparent relationships and conclude things that were not previously known in a way that is currently understood and valuable for the owner of the data.(Figure 2)
Data mining can be grouped into two main categories, namely (Yulia Darmi, 2016):

a. Descriptive mining is a process of finding essential data characteristics in one database. Data mining techniques that include descriptive mining are clustering, association, and sequential mining.
b. Predictive is the process of finding patterns from the data by using some other variable in the future. One of the techniques contained in predictive mining is classification. In simple terms, data mining is usually a process of filtering or "mining" knowledge from large amounts of data.

The data mining can be divided into several stages as follows:

1. Get rid of inconsistent data
2. Data integration to combine data from multiple sources
3. Transforming data into other forms.
4. For example, the clustering method can only accept categorical data input, so numerical data needs to be divided into several intervals.
5. Application of data mining techniques m valuable knowledge of the data.
6. Knowledge evaluation and presentation

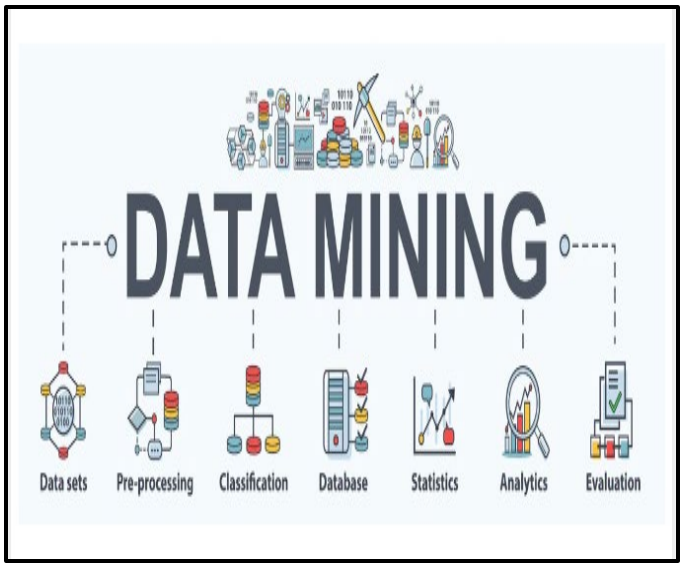

Figure 2. Data Mining

2.2.Knowledge Discovery In Database (KDD)

Knowledge Discovery in Database (KDD) is a structured analysis process for obtaining correct, new, useful information and finding patterns from big and complex data. Data mining (DM) is at the core of the KDD process, namely by using specific algorithms to explore data, build models and find patterns that are not yet known. Models are used to understand data phenomena, analysis, and predictions. The KDD process generally includes nine steps described in Figure 3 (Maimon and). Lior, 2010)

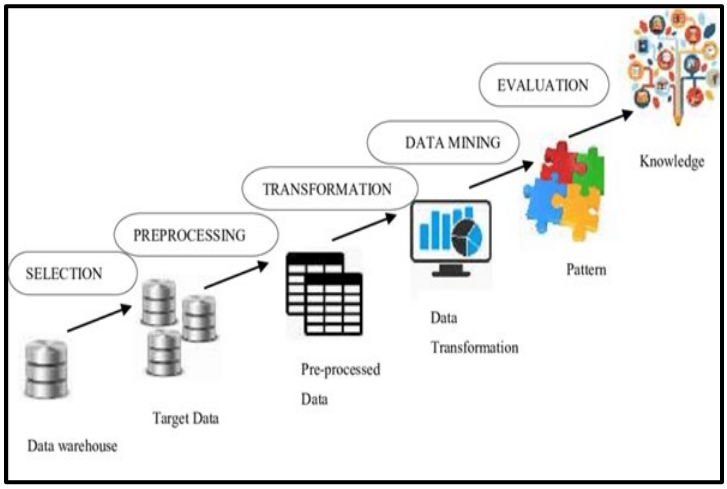

Figure 3. Knowledge Discovery In Database Process

2.3.Classification

Classification is a data mining technique that extracts the model by describing the critical data classes. The model is called classification by predicting class labels based on categories. For example, create a classification model to categorize bank loan applications with a safe or risky class label. Researchers propose many classification methods in machine learning, pattern recognition, and statistics. There have been many applications of classification data mining methods, such as fraud detection, marketing target determination, performance prediction, manufacturing, and medical diagnosis (Narulita, et al, 2021).

2.4.Clustering

According to Widodo, Clustering is a method used to divide a data series into several groups based on previously determined similarities. A cluster is a group or set of data objects that are similar to each other in the same cluster and dissimilar to objects of different clusters. Objects will be grouped into one or more clusters so that objects in one cluster will have high similarity. Objects are grouped based on the principle of maximizing the similarity of objects in the same cluster and maximizing inequality in different clusters. Object similarity is usually obtained from attribute values that describe data objects, so data objects are usually represented as a point in a multidimensional space (Yulia Darmi, 2016).s

This data clustering aims to minimize the objective function set in the clustering process, which generally tries to minimize variations within a cluster. Furthermore, minimize variations between clusters (Yulia Darmi, 2016).

2.5.K-Means

K-Means is a non-hierarchical data clustering method that seeks to partition existing data into one or more clusters or groups so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into groups. Other. K-Means is a distance-based clustering method that divides data into several clusters, and this algorithm only works on numeric attributes. The K-Means algorithm includes partitioning clustering, separating data into k subregions. The K-Means algorithm is well known for its ease and ability to cluster extensively quickly and outlier data. In the K-Means algorithm, each data must belong to a specific cluster, and it is possible for any data that belongs to a specific cluster at one stage of the process, at the next stage to move to another cluster (Yulia Darmi, 2016).

The alternative application of Kmeans uses the development of related computational theories that have been proposed, including the following selections (Seimahuira, 2021):

1. Distance Space to calculate the distance between a data and the centroid.

   Distance Space has been implemented in calculating the distance between the data and the centroid. The distance measurement formula used is Euclidean distance or what is commonly called a straight line distance which is calculated using the formula :

$$d_((ip)) = (⟦(X_i1-X_i1)⟧^2 + ⟦(X_i2-X_i2)⟧^2+⋯.+ ⟦(X_ip-X_ip)⟧^2) ... (1)$$

   where i and j are two data objects that have p attribute with a numeric value, which is expressed as i=(xi1,xi2,…..,xip) (Suyanto, 2019)
2. Method of allocating data back into each cluster.
3. Objective functions used.

3. Methods

Research methodology is the systematic steps or stages that will be taken in the research process to collect the desired data or information with the object being studied. The following research flowchart is as follows (Figure 4):

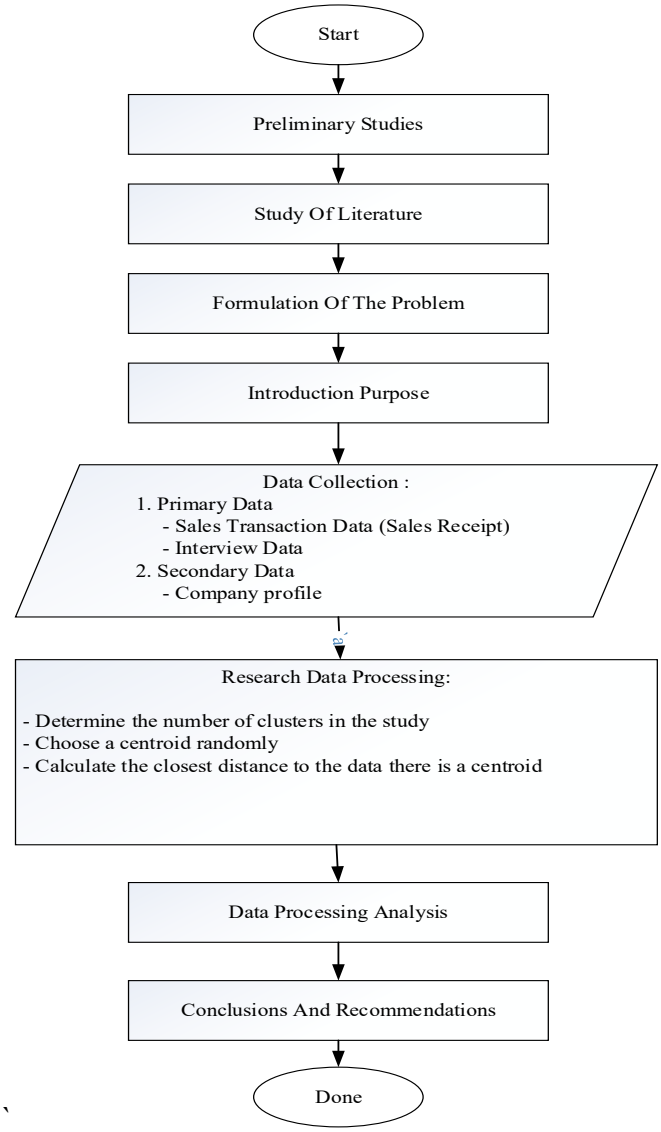

Figure 4. Flowchart

4. Data Collection

The data collection consists of primary and secondary data, where primary data is obtained from observations and interviews. In contrast, secondary data consists of the business profile of Abie JM Stores and sales data for 2022 from January to February.

Sales data used is sales data for 2022 from January to February. The following is a recapitulation of sales data for Abie JM's store business: (Table 2)

Tabel 2 Sales Data of Abie JM Store

Item Name	Sold Number	Transaction Number	Average Transaction
234 Refill	6	3	2
Abc Manis Botol 135	2	3	0,666666667
Abc Manis Botol 275	8	2	4

Abc Manis Refill 225	4	2	2
Abc Mocca 27 Gr	6	1	6
Agar Agar Powder Swallow Sun 9	3	1	3
Aice Chocolate Crispy 60 Gr	1	1	1
Aice Freezy Choco 49 Ml	1	1	1
Aice Milk Melon 50 Gr	1	1	1
***	***	***	***
Vape	2	1	2
Vixal 190 Ml	2	2	1
Grand Total	**651**	**17404**	**0,037405194**

4. Results And Discussion

The K-Means clustering method is used for clustering at the Abie JM store, where in this research, the transaction data will be clustered into three criteria, namely excellent, very good and not good by using the following steps :

1. Determine the number of clusters (k) randomly. Based on the research, the value of the cluster used is 3 clusters (k=3), namely cluster 1 (C1) showing the results of not selling well, cluster 2 (C2) showing the results of selling well, cluster 3 (C3) showing the results being very good.

Selecting a centroid at random in this calculation, the centroid is selected as follows: (Table 3)

Tabel 3. Initial Center Determination Cluster

CLUSTER INITIAL CENTER DETERMINATION		
CLUSTERINS KE -1	**1**	**1**
CLUSTERINS KE -2	**7**	**1**
CLUSTERINS KE -3	**26**	**33**

2. Calculate the closest distance from the data to the selected centroid (cluster center) using the Euclidean distance formula in equation (1), namely:

$$D(X_2 - X_1) = \sqrt{(X_2\,a - X_1 a)^2 + (X_2\,b - X_1 b)^2 + (X_2\,c - X_1 c)^2}$$

• Data Distance to -1 Center Cluster 1

$$d_{1,1} = \sqrt{(3-1)^2 + (6-1)^2} = 27$$

• Data Distance to -1 Center Cluster 2

$$d_{1,2} = \sqrt{(3-7)^2 + (6-1)^2} = 6{,}40312$$

• Data Distance to -1 Cluster Center 3

$$d_{1,3} = \sqrt{(3-26)^2 + (6-33)^2} = 5{,}385164807$$

• Distance to the 2nd Data Center Cluster 1

$$d_{1,1} = \sqrt{(3-1)^2 + (2-1)^2} = 31$$

• Distance to 2nd Data Center Cluster 2

$$d_{1,2} = \sqrt{(3-7)^2 + (2-1)^2} = 4{,}12311$$

• Distance to the 2nd Data Center Cluster 3

$$d_{1,3}= \sqrt{(3-26)^2+(2-33)^2} \quad = 2{,}236067977$$

3. Calculating the shortest distance and grouping the data into clusters, after getting the results of the data distance from the cluster center to 182, then determining the shortest distance using equation (2), so that the following results are obtained: (Table 4)

Table 4. The Shortest Distance In Iteration 1

NO	Name of goods	Number of Transactions	Amount Sold	Average Sales	C 1	C 2	C 3	Shortest Distance
1	234 Refill	3	6	2	6,708203932	6,708203932	6,708203932	6,708203932
2	Abc Manis Botol 135	3	2	0,666666667	3,605551275	3,605551275	3,605551275	3,605551275
3	Abc Manis Botol 275	2	8	4	8,246211251	8,246211251	8,246211251	8,246211251
4	Abc Manis Refill 225	2	4	2	4,472135955	4,472135955	4,472135955	4,472135955
5	Abc Mocca 27 Gr	1	6	6	6,08276253	6,08276253	6,08276253	6,08276253
6	Agar Agar Powder Swallow Sun 9	1	3	3	3,16227766	3,16227766	3,16227766	3,16227766
...	...	...	...	...	...	...	...	...
181	Vape	1	2	2	2,236067977	2,236067977	2,236067977	2,236067977
182	Vixal 190 Ml	2	2	1	2,828427125	2,828427125	2,828427125	2,828427125

Table 5. Clustering Results In Iteration 1

NO	Name of goods	C 1	C 2	C 3
1	234 Refill			1
2	Abc Manis Botol 135			1
3	Abc Manis Botol 275			1
4	Abc Manis Refill 225			1

5	Abc Mocca 27 Gr			1
6	Agar Agar Powder Swallow Sun 9			1
...	...	...	...	...
181	Vape			1
182	Vixal 190 Ml			1

The iteration results obtained C1 as many as four products, C2 as many as 11 products, C3 as many as 167 products, and a new centroid to calculate iteration 2. (Tables 5, 6,7 & 8)

4. Calculate the new centroid using the results of each member in each cluster using equation (3), namely:
   - For cluster 1 there are 4 products, namely the 137,161,166, and 167 th products, so that:

     $C1,1 = (\frac{1+26+1+8}{4}) = 9$

     $C1,2 = (\frac{31+33+2+79}{4}) = 36,25$

   - For cluster 2, there are 11 products, namely the 14,31,34,59,70,71,138,155,156,160, and 164 th products so that:

     $C2,1 = (\frac{1+4+4+5+7+1+15+9+18+14+4}{11}) = 7,81818$

     $C2,2 = (\frac{1+10+6+4+1+3+4+11+18+18+9}{11}) = 8,09091$

   - For cluster 3 there are 167 products, namely:
     1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,32,33,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,60,61,62,63,64,65,66,67,68,69,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,157,158,159,162,163,165,168,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181, and 182 product to so:

     $C3,1 = (\frac{3+3+2+2+1+1+1+1+1+1+2+3+5+2+1+1+\cdots+3}{167}) = 1,353293413$

     $C3,2 = (\frac{6+2+8+4+6+3+1+1+1+1+1+3+4+5+3+2+2+\cdots+2}{167}) = 2,562874251$

Table 6. New Centroids In Iteration 2

Determination Of Awa Cluster Center		
CLUSTERINS KE -1	**9**	**36,25**
CLUSTERINS KE -2	**7,81818**	**8,09091**
CLUSTERINS KE -3	**1,35329**	**2,56287**

Next, calculate the Euclidean distance again and the shortest distance from the cluster so that the grouping results in iteration 2.

• Data Distance to -1 Center Cluster 1

$d_{1,1} = \sqrt{(3-9)^2 + (6-36,25)^2} = 30,83930122$

• Data Distance to -1 Center Cluster 2

$$d_{1,2}=\sqrt{(3-7{,}81818)^2+(6-8{,}09091)^2} \quad = 5{,}25231$$

• Data Distance to -1 Cluster Center 3

$$d_{1,3}=\sqrt{(3-1{,}35329)^2+(6-2{,}56287)^2} \quad = 3{,}811235556$$

• Distance to the 2nd Data Center Cluster 1

$$d_{1,1}=\sqrt{(3-9)^2+(2-36{,}25)^2} \quad = 34{,}77157604$$

• Distance to 2nd Data Center Cluster 2

$$d_{1,2}=\sqrt{(3-7{,}81818)^2+(2-8{,}09091)^2} \quad = 7{,}76621$$

• Distance to the 2nd Data Center Cluster 3

$$d_{1,3}=\sqrt{(3-1{,}35329)^2+(2-2{,}56287)^2} \quad = 1{,}740251838$$

Table 7. Shortest Distance In Iteration 2

NO	Name Of Goods	Number Of Transactions	Amount Sold	Average Sales	C 1	C 2	C 3	Shortest Distance
1	234 Refill	3	6	2	6,708203932	6,708203932	6,708203932	6,708203932
2	Abc Manis Botol 135	3	2	0,666666667	3,605551275	3,605551275	3,605551275	3,605551275
3	Abc Manis Botol 275	2	8	4	8,246211251	8,246211251	8,246211251	8,246211251
4	Abc Manis Refill 225	2	4	2	4,472135955	4,472135955	4,472135955	4,472135955
5	Abc Mocca 27 Gr	1	6	6	6,08276253	6,08276253	6,08276253	6,08276253
6	Agar Agar Powder Swallow Sun 9	1	3	3	3,16227766	3,16227766	3,16227766	3,16227766
7	Aice Chocolate Crispy 60 Gr	1	1	1	1,414213562	1,414213562	1,414213562	1,414213562
8	Aice Freezy Choco 49 Ml	1	1	1	1,414213562	1,414213562	1,414213562	1,414213562
9	Aice Milk Melon 50 Gr	1	1	1	1,414213562	1,414213562	1,414213562	1,414213562
10	Aice Strawberry Cone	1	1	1	1,414213562	1,414213562	1,414213562	1,414213562
***	***	***	***	***	***	***	***	***
181	Vape	1	2	2	2,236067977	2,236067977	2,236067977	2,236067977
182	Vixal 190 Ml	2	2	1	2,828427125	2,828427125	2,828427125	2,828427125

Table 8. Clustering Results In Iteration 2

No	Name Of Goods	C1	C2	C3
1	**234 Refill**			1
2	**Abc Manis Botol 135**			1
3	**Abc Manis Botol 275**			1
4	**Abc Manis Refill 225**			1
5	**Abc Mocca 27 Gr**			1
6	**Agar Agar Powder Swallow Sun 9**			1
7	**Aice Chocolate Crispy 60 Gr**			1
8	**Aice Freezy Choco 49 Ml**			1
9	**Aice Milk Melon 50 Gr**			1
10	**Aice Strawberry Cone**			1
***	***	***	***	***
181	**Vape**			1
182	**Vixal 190 Ml**			1

The results of iteration 2 obtained C1 as many as 3 products, C2 as many as 11 products, and C3 as many as 168 products and continued to the next iteration until there was no change in results. The results of clustering in iterations 1 and 2 are not the same, then the grouping is carried out again with Fuzzy K-Means with the same steps for the next iteration until the results obtained are the same, and in iteration 5 the same clustering results have been obtained with the previous iteration, namely : (Table 9)

Table 9. Clustering Results In Iteration 3

NO	Name Of Goods	C1	C2	C3
1	234 Refill			1
2	Abc Manis Botol 135			1
3	Abc Manis Botol 275		1	
4	Abc Manis Refill 225			1
5	Abc Mocca 27 Gr			1
6	Agar Agar Powder Swallow Sun 9			1
7	Aice Chocolate Crispy 60 Gr			1
8	Aice Freezy Choco 49 Ml			1
9	Aice Milk Melon 50 Gr			1
10	Aice Strawberry Cone			1
***	***	***	***	***
181	Vape			1
182	Vixal 190 Ml			1

Based on the results obtained, C1 is categorized as a product that has the lowest level of sales, there are 3 products, C2 is categorized as a product that has a moderate level of sales, there are 12 products, and C3 is categorized as a product with a high level of sales, there are 167 products. So clusters with low and moderate sales levels need to be considered by business managers to evaluate so that income increases at the Abie JM store.

5. Conclusion

From the results of processing and discussion using fuzzy K-Means, it can be concluded that the sales level clustering based on low sales levels, medium sales levels, and high sales levels at the Abie JM store in 2022, namely:

1. The first cluster is categorized as products with low sales levels. There are 3 products.
2. The second cluster is categorized as products with a moderate level of sales. There are 12 products.
3. The third cluster is categorized as products with a high level of sales. There are 167 products.

References

Algarni, A. Data mining in education. *International Journal of Advanced Computer Science and Applications*, *7*(6), 456-461. (2016).

Annur, H. Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering. *Jurnal Informatika Upgris*, *5*(1). (2019).

Darmi, Y. D., & Setiawan, A. Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama*, *12*(2). (2016).

Gustientiedina, G., Adiya, M. H., & Desnelita, Y. Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi dan Sistem Informasi*, *5*(1), 17-24. (2019).

Hamzah, M. L., Purwati, A. A., Jamal, A., & Rizki, M. An Analysis of Customer Satisfaction and Loyalty of Online Transportation System in Pekanbaru, Indonesia. In IOP Conference Series: Earth and Environmental Science (Vol. 704, No. 1, p. 012029). IOP Publishing. (2021).

Handoko, K. Penerapan Data Mining dalam Meningkatkan Mutu Pembelajaran Menggunakan Metode K-MEANS Clustering. *Jurnal Nasional Teknologi dan Sistem Informasi*, *2*(3), 31-40. (2016).

Irawan, Y. Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive . *JTIULM*, 13 – 20. (2019).

Kameshwaran, K., & Malarvizhi, K. Survey on clustering techniques in data mining. *International Journal of Computer Science and Information Technologies*, *5*(2), 2272-2276. (2014).

Mirmozaffari, M., Alinezhad, A., & Gilanpour, A. Heart disease prediction with data mining clustering algorithms. *Int'l Journal of Computing, Communications & Instrumentation Engg*, *4*(1), 16-19. (2017).

Muningsih, E., & Kiswati, S. Penerapan metode K-means untuk clustering produk online shop dalam penentuan stok barang. *Bianglala Informatika*, *3*(1). (2015).

Muni, A. Analisis Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Penjualan Sepeda Motor Studi Kasus PT. Alfa Scorpii. *JUTI UNISI*, *4*(1), 1-8. (2020).

Narulita, S. Pengujian Akurasi Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Algoritma C4. 5 untuk Penentuan Peminatan Peserta Didik. *Media Aplikom*, *13*(2), 15-29. (2021).

Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier. *Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)*, *7*(1), 8-17. (2021).

Putra, R. R., & Wadisman, C. Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, *1*(1), 72-77. (2018).

Rizki, M., Wenda, A., Pahlevi, F. D., Umam, M. I. H., Hamzah, M. L., & Sutoyo, S. Comparison of Four Time Series Forecasting Methods for Coal Material Supplies: Case Study of a Power Plant in Indonesia. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)* (pp. 1-5). IEEE. (2021, July).

Rizki, M., Devrika, D., & Umam, I. H. Aplikasi Data Mining dalam penentuan layout swalayan dengan menggunakan metode MBA. *Jurnal Teknik Industri: Jurnal Hasil Penelitian dan Karya Ilmiah dalam Bidang Teknik Industri*, *5*(2), 130-138. (2020).

Rizki, M., Umam, M. I. H., & Hamzah, M. L. Aplikasi Data Mining Dengan Metode CHAID Dalam Menentukan Status Kredit. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, *18*(1), 29-33. (2020).

Seimahuira, S. Implementasi datamining dalam menentukan destinasi unggulan berdasarkan online reviews tripadvisor menggunakan algoritma K-Means. *Technologia: Jurnal Ilmiah*, *12*(1), 53-58. (2021).

Shah, M., & Nair, S. A survey of data mining clustering algorithms. *International Journal of Computer Applications*, *128* (1), 1-5. (2015).

Sumadikarta, I., & Abeiza, E. Penerapan Algoritma K-Means Pada Data Mining Untuk Memilih Produk Dan Pelanggan Potensial. *Jurnal Satya Informatika*, *1*, 12-22. (2014).

Windarto, A. P. Implementation of data mining on rice imports by major country of origin using algorithm using k-means clustering method. *International Journal of artificial intelligence research*, *1* (2), 26-33. (2017).

Wiyanto, W., Sulistyohati, A., & Umilhuda, U. (2022). Penerapan Data Mining Untuk Menganalisa Pola Pembelian Sayuran Hidroponik Menggunakan Metode Algoritma Apriori. *Journal of Practical Computer Science*, *1*(2), 38-49.

Zulyanti, T., & Noeryanti, N. Perbandingan Pengelompokan Usaha Mikro Kecil dan Menengah di Kabupaten Klaten Tahun 2019 dengan Metode k-Means dan Clustering Large Application. *Jurnal Statistika Industri dan Komputasi*, *7*(01), 46-59. (2022).

Biography

Afrido is student in Industrial Engineering Departement at Sultan Syarif Kasim State Islamic University, Indonesia

Muhammad Rizki is an Assistant Professor in Sultan Syarif Kasim State Islamic University at Industrial Engineering Department. His master's degree was in Industrial Engineering Department from University of Indonesia, and he got a dual degree in Master Business and Administration from National Taiwan University of Science and Technology in Taiwan. He is currently a member of IEOM as a faculty advisor and published several articles in the international conference about Simulation Modeling, Healthcare Management, and industrial engineering area

Ismu Kusumanto is assistant professor Industrial Engineering Departement at Sultan Syarif Kasim State Islamic University, Indonesia

Nazaruddin is an Assistant Professor in Industrial Engineering Department - Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia. He has published some articles about Supply Chain Management, Operational Research and Metaheuristic

Misra Hartati is assistant professor Industrial Engineering Departement at Sultan Syarif Kasim State Islamic University, Indonesia