Jurnal Mantik, 6 (3) (2022), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Implementation of data mining to determine stock inventory at kenza grocery stores using the k-means clustering method

Kasini[1], Nani Hidayati[2]
[1]Universitas Pahlawan Tuanku Tambusai, Indonesia
[2]STIKOM Tunas Bangsa Pematang siantar, Indonesia

A R T I C L E I N F O

Article history:

Received Nov 02, 2022
Revised Nov 16, 2022
Accepted Nov 30, 2022

Keywords:

Clustering;
Data Mining;
Groceries.

ABSTRACT

Toko Kenza is a store engaged in the sale of basic necessities, however, of the various kinds of groceries that are sold, not all of them are in demand, some are best-selling, best-selling and not best-selling. The data on sales and purchases of goods as well as unexpected expenses at this Kenza store are not well structured, so that the data only functions as an archive for the store and cannot be used for developing marketing strategies. Therefore it is necessary to apply data mining using the K-Means method at the Kenza store. The K-means method can be applied to Kenza stores to determine which basic food items are selling the best, selling and not selling. The application of the K-Means method to the Kenza store, namely by grouping the basic food stock data, then randomly selecting 3 clusters as the initial centroid. After the data in each cluster does not change, it can be seen that the end result is that there are 2 best-selling data, 15 data that are best-selling, 23 data that are not selling well. Then applying the K-Means method to Rapidminer is done by entering product stock data, namely initial stock, sold stock and final stock which will become a Database on MS. Check, the data is then connected to the RapidMiner tools and will be processed and formed K-Means. After that, RapidMiner will generate which products are the best selling, selling and not selling.

Corresponding Author:

Kasini,
Universitas Pahlawan Tuanku Tambusai, Indonesia,
Jl. Tuanku Tambusai No.23, Bangkinang, Kec. Bangkinang, Kabupaten Kampar, Riau 28412
Email: kasiniaqm@gmail.com

1. INTRODUCTION

The high number of market demands for basic food needs has made businesses increase [1]. The role of Small and Medium Enterprises (SMEs) in the community has a very important role in meeting daily basic needs[2]. Grocery store is one of the businesses engaged in the economy, which currently still uses conventional and offline systems [3]. So that an offline to online sales system is needed as a marketing strategy so that sales can increase [4].

At the Kenza grocery store at this time not yet has a standard method applied, the supply of basic necessities is carried out only by check the supplies of basic necessities that are almost depleted then update the stocks of basic necessities the. So this is less efficient if

at one time you need groceries in quantity large and apparently out of stock [5]. Data mining is a method for finding certain patterns from large data sets [6].

K-Means is one of the clustering algorithms. clustering is an algorithm for grouping a number of data into certain data sets (clusters) [7]from one of the data mining functions.

Clustering method is a data analysis method that can be used insolve problems in a data grouping[8]. One of the methods in in the clustering method is the K-mens method. The K-means method is a method which can group large amounts of data with relatively fast and efficient computation time. The K-Means method has been used in grouping systems in various business fields, such as in the field of Marketing and Sales [9][10], healthcare [11][12], education [13][14], and business others [15].

The purpose of this research is to make it easy for kenza grocery stores to find out the inventory management strategy. By knowing which items are the best-selling, best-selling and not selling well. So that the stock is always updated and there is no product purchase.

2. RESEARCH METHOD

The framework for research at the Kenza grocery store is as follows in Figure 1 :

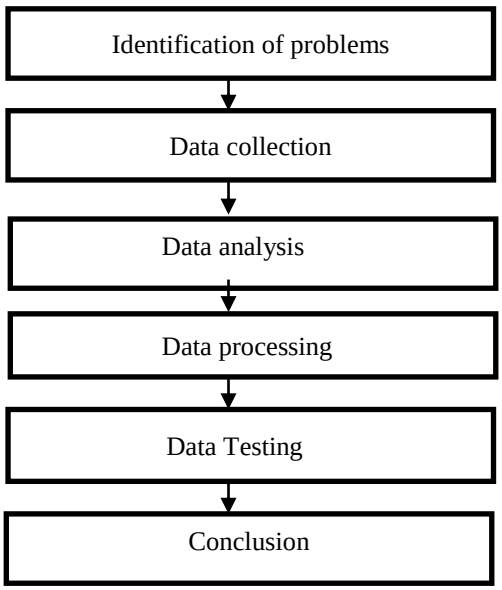

Figure 1. Research Stages

1. Identification of problems
   Namely identifying the problem it's in the store that is to find out which products at kenza grocery stores are the best-selling, best-selling, and not selling.
2. Data collection
   That is collecting data by asking stock data for all groceries in the kenza store
3. Data analysis
   That is analyzing the problem, namely by analyzing basic food stock data at kenza stores
4. Data processing
   Namely by processing the data provided enza shop, then the data is processed with calculated K-means clustering method manually to classify groceries the best-selling, best-selling and not-selling.

5. Data Testing
   Namely by testing the data already processed with Rapid Miner.
6. Conclusion
   Namely by drawing conclusions from the results calculation and processing of calculated data manually with Rapid Miner results data.

3. RESULTS AND DISCUSSIONS

The data used in this grouping of basic food sales is data on kenza groceries in Sukajadi village, Rokandownstream, Riau. Basic food data consists of several types, including variable data and primary data from kenza basic food stores. The data will be processed to create knowledge of inventory management strategies. In the application of data mining grouping food sales data will be collected into 3 clusters, namely the best-selling, best-selling and not-selling, it must apply the variables used as a guide for testing K-means clustering. These variables are as follows:

Table. Variabel

No	Variable	Description
1	Name Of Goods	The Name Of The Groceries
2	Unit	Groceries Unit
3	Stock	Stock Items In Store
4	Sales Amount	Number Of Items Sold

So that the above data can be processed using the K-means clustering method, the data must be arranged according to the variables. The following is the data that will be used in this study, namely the latest sales data, namely October 2022 data:

Table 2 Sample Data

No	Types of goods	Unit	Stock	Amount Sold
1	Rice 5 Kg	Kg	70	40
2	Bulk Cooking Oil	Kg	60	35
3	1 Kg Packaged Oil	Kg	25	7
4	2 Kg Packaged Oil	Kg	15	5
5	Powder Detergent	Wrap	100	60
6	Toothpaste	Pcs	80	50
7	Bath Soap	Pcs	60	25
8	Sweetened Condensed Milk	Can	50	15
9	Sugar 1 Kg	Kg	50	35
10	Dishwashing Bar Soap	Pack	50	30
11	Sunlight	Packaging	30	20
12	Prawn Crackers	Packaging	30	10
13	Brown Sugar	Kg	10	5
14	Indomie	Cardboard Box	4	2
15	Delicious Noodles	Cardboard Box	3	2
16	Noodle Eco	Cardboard Box	10	7

17	Rice Flour	Kg	25	10
18	Flour	Kg	30	20
19	Floor Deodorizer	Packaging	25	15
20	Eggs	Board	40	35
21	Sauce	Packaging	30	15
22	Sweet Soy Sauce	Packaging	35	20
23	Royco	Packaging	25	18
24	Ajinomoto	Packaging	30	10
25	Moderate Mineral Water	Cardboard Box	28	10
26	Great Mineral Water	Cardboard Box	33	10
27	Glass Mineral Water	Cardboard Box	10	15
28	Coarse Salt	Pack	50	30
29	Fine Salt	Pack	10	5
30	Rice 10 Kg	Bag	30	10
31	Rice 30 Kg	Bag	32	7
32	Stick Noodles	Packaging	20	10
33	Vermicelli Noodles	Kg	20	5
34	Sugar 1/2 Kg	Kg	25	5
35	Lpg	Tube	35	30
36	Packaged Bread	Packaging	40	20
37	Large Can Of Sardines	Can Packaging	25	3
38	SMALL Canned Sardines	Can Packaging	15	8
39	Wasp Nest Syrup	Bottle/Dozen	36	15
40	Abc Syrup	Bottle/Dozen	24	10

The K-means clustering algorithm in grouping sales data at kenza grocery stores can be described as follows:

1. Iteration 1

   In determining the initial centroid in iteration 1, it is determined randomly from existing data. In this study, the data taken were the 5th data, the 20th data, and the 29th data.

Table 3 Iteration 1

Cluster 1	Data to 5	100	60
Cluster 2	Data to 20	40	35
Cluster 3	Data to 29	10	5

Calculation of the distance from data 1 to the cluster center point is as follows:

$D(1,1) = \sqrt{(70-100)^2 + (40-60)^2} = 36$

$D(1,2) = \sqrt{(70-40)^2 + (40-35)^2} = 30$

$D(1,3) = \sqrt{(70-10)^2 + (40-5)^2} = 69$

And so on, calculating the distance of the second data to the cluster center data, then determining to compare between the 3 clusters with the smallest value being an option, if the smallest value is found then it can be grouped into that cluster. Following are the cluster results in iteration 1:

C1 = 2 data C2 = 15 data and C3 = 23 data

Table 3 Results in Iteration 1

Number	Type Of Goods	Unit	Stock	Amount Sold	C1	C2	C3	Shortest Distance	Clustering
1	Rice 5 Kg	Kg	70	40	36	30	69	30	2
2	Bulk Cooking Oil	Kg	60	35	47	20	58	20	2
3	1 Kg Packaged Oil	Kg	25	7	92	32	15	15	3
4	2 Kg Packaged Oil	Kg	15	5	101	39	5	5	3
5	Powder Detergent	Wrap	100	60	0	65	105	0	1
6	Toothpaste	Pcs	80	50	22	43	83	22	1
7	Bath Soap	Pcs	60	25	53	22	54	22	2
8	Sweetened Condesed Milk	Can	50	15	67	22	41	22	2
9	Sugar 1 Kg	Kg	50	35	56	10	50	10	2
10	Diswashing Bar Soap	Pack	50	30	58	11	47	11	2
11	Sunlight	Packaging	30	20	81	18	25	18	2
12	Prawn Crackers	Packaging	30	10	86	27	21	21	3
13	Brown Sugar	Kg	10	5	105	42	0	0	3
14	Indomie	Cardboard Box	4	2	112	49	7	7	3
15	Delicious Noodles	Cardboard Box	3	2	113	50	8	8	3
16	Eco Noodles	Cardboard Box	10	7	104	41	2	2	3
17	Rice Flour	Kg	25	10	90	29	16	16	3
18	Flour	Kg	30	20	81	18	25	18	2
19	Floor Deodorizer	Packaging	25	15	87	25	18	18	3
20	Egg	Board	40	35	65	0	42	0	2
21	Packaged Sauce	Packaging	30	15	83	22	22	22	2
22	Sweet Soy Sauce	Packaging	35	20	76	16	29	16	2
23	Royco	Packaging	25	18	86	23	20	20	3
24	Ajinomoto	Packaging	30	10	86	27	21	21	3
25	Moderate Mineral Water	Cardboard Box	28	10	88	28	19	19	3
26	Great Mineral Water	Cardboard Box	33	10	84	26	24	24	3
27	Glass Mineral	Cardboard Box	10	15	101	36	10	10	3

	Water								
28	Coarse Salt	Pack	50	30	58	11	47	11	2
29	Fine Salt	Pack	10	5	105	42	0	0	3
30	Rice 10 Kg	Bag	30	10	86	27	21	21	3
31	Rice 30 Kg	Bag	32	7	86	29	22	22	3
32	Stick Noodles	Packaging	20	10	94	32	11	11	3
33	Vermicelli Noodles	Kg	20	5	97	36	10	10	3
34	Sugar 1/2 Kg	Kg	25	5	93	34	15	15	3
35	LPG Gas	Tube	35	30	72	7	35	7	2
36	Packaged Bread	Wrap	40	20	72	15	34	15	2
37	Big Sardines	Can	25	3	94	35	15	15	3
38	Small Sardines	Can	15	8	100	37	6	6	3
39	Sarang Tawon Syrup	Bottle	36	15	78	20	28	20	2
40	ABC Syrup	Bottle	24	10	91	30	15	15	3

1. Then determine the new centroid value, this value is determined by the data that enters the cluster, based on the table above (data 1-40) the following values are obtained:

   Cluster 1 has 2 data
   Cluster 2 contains 15 data
   Cluster 3 contains 23 data
   To determine the new centroid value, the formula used is:

   $$Ck = \frac{the\ sum\ of\ the\ values\ entered\ into\ the\ cluster}{the\ amount\ of\ data\ entered}$$

   So that the new centroid value is obtained in iteration 2 as follows:

Table 4 Iteration 2

Cluster 1	90	55
cluster 2	44	26
Cluster 3	21	8

To find the next centroid value, repeat step 1 above. After the new centroid value is found, repeat the distance calculation step in the previous step to enter data into the cluster.

If the above steps are repeated with the same steps to produce data in a cluster exactly the same as the previous data with the data in the next step, or in other words the data does not change its position in the cluster. The position in each of the last clusters in this study is in the 4th iteration the cluster position did not change with the 3rd iteration cluster position, so that it can be seen in the following table with the 4th iteration calculations:

Table 5 4th iteration Calculation

No	Type Of Goods	Unit	Stock	Amount	C1	C2	C3	Shortest	Clustering

				Sold				Distance	
1	Rice 5 Kg	Kg	70	40	17	28	57	17	1
2	Bulk Cooking Oil	Kg	60	35	28	16	46	16	2
3	1 Kg Packaged Oil	Kg	25	7	72	29	4	4	3
4	2 Kg Packaged Oil	Kg	15	5	82	38	8	8	3
5	Powder Detergent	Wrap	100	60	19	64	93	19	1
6	Toothpaste	Pcs	80	50	3	41	71	3	1
7	Bath Soap	Pcs	60	25	34	14	41	14	2
8	Sweetened Condesed Milk	Can	50	15	48	12	29	12	2
9	Sugar 1 Kg	Kg	50	35	37	10	38	10	2
10	Diswashing Bar Soap	Pack	50	30	39	5	35	5	2
11	Sunlight	Packaging	30	20	61	17	13	13	3
12	Prawn Crackers	Packaging	30	10	67	23	8	8	3
13	Brown Sugar	Kg	10	5	86	42	12	12	3
14	Indomie	Cardboard Box	4	2	93	49	19	19	3
15	Delicious Noodles	Cardboard Box	3	2	94	49	20	20	3
16	Eco Noodles	Cardboard Box	10	7	85	41	12	12	3
17	Rice Flour	Kg	25	10	71	27	3	3	3
18	Flour	Kg	30	20	61	17	13	13	3
19	Floor Deodorizer	Packaging	25	15	68	24	7	7	3
20	Egg	Board	40	35	46	11	31	11	2
21	Packaged Sauce	Packaging	30	15	64	20	10	10	3
22	Sweet Soy Sauce	Packaging	35	20	57	13	17	13	2
23	Royco	Packaging	25	18	67	23	9	9	3
24	Ajinomoto	Packaging	30	10	67	23	8	8	3
25	Moderate Mineral Water	Cardboard Box	28	10	68	24	6	6	3
26	Great Mineral Water	Cardboard Box	33	10	64	21	11	11	3
27	Glass Mineral Water	Cardboard Box	10	15	81	38	13	13	3
28	Coarse Salt	Pack	50	30	39	5	35	5	2
29	Fine Salt	Pack	10	5	86	42	12	12	3
30	Rice 10 Kg	Bag	30	10	67	23	8	8	3
31	Rice 30 Kg	Bag	32	7	67	24	11	11	3
32	Stick Noodles	Packaging	20	10	75	31	2	2	3
33	Vermicelli Noodles	Kg	20	5	78	34	5	5	3
34	Sugar 1/2 Kg	Kg	25	5	74	30	5	5	3
35	Lpg Gas	Tube	35	30	52	12	25	12	2
36	Packaged Bread	Wrp	40	20	53	9	21	9	2
37	Big Sardines	Can	25	3	75	31	7	7	3

38	Small Sardines	Can	15	8	80	36	7	7	3
39	Sarang Tawon Syrup	Bottle	36	15	59	15	15	15	3
40	ABC Syrup	Bottle	24	10	72	27	2	2	3

The final results of the K-Means clustering calculation of kenza grocery store sales can be concluded with the results of Cluster 1 with the 3 best-selling types of goods, namely 5kg rice, toothpaste and powder detergent, while in cluster 2 with 10 best-selling types of goods, namely bulk cooking oil, soap bath, sweetened condensed milk, 1 kg sugar, dishwashing soap, eggs, sweet soy sauce, coarse salt, LPG gas, packaged bread. And lastly cluster 3 with 27 types of goods.

While the results based on the implementation of rapid miner 7.5 with Kenza basic food store data are as follows:

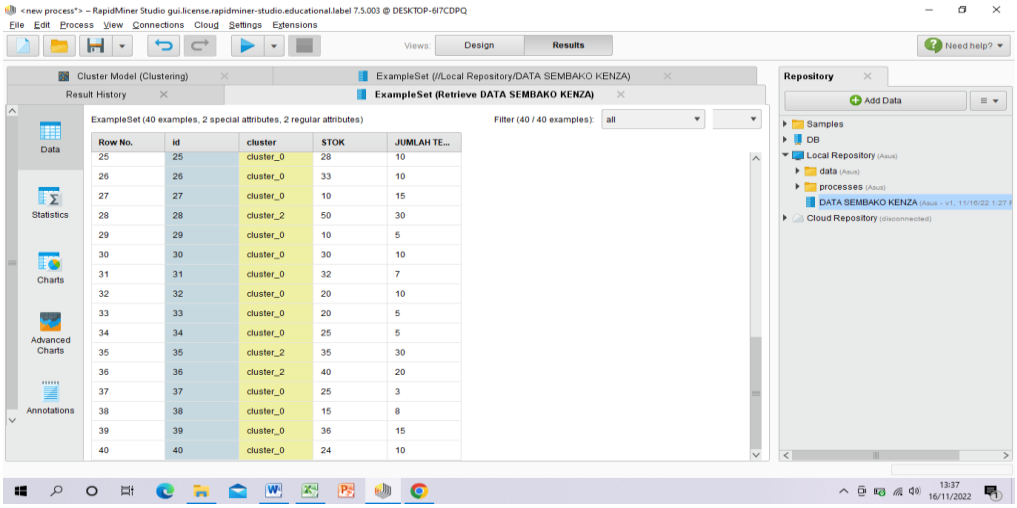

Figure 2. Rapid Results 1

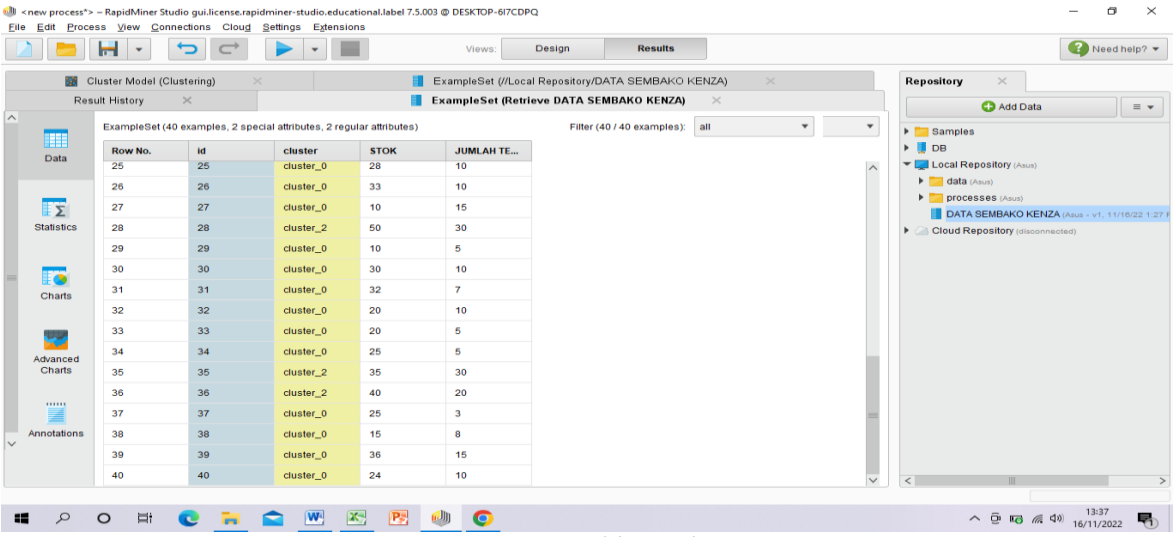

Figure 3. Rapid Results 1

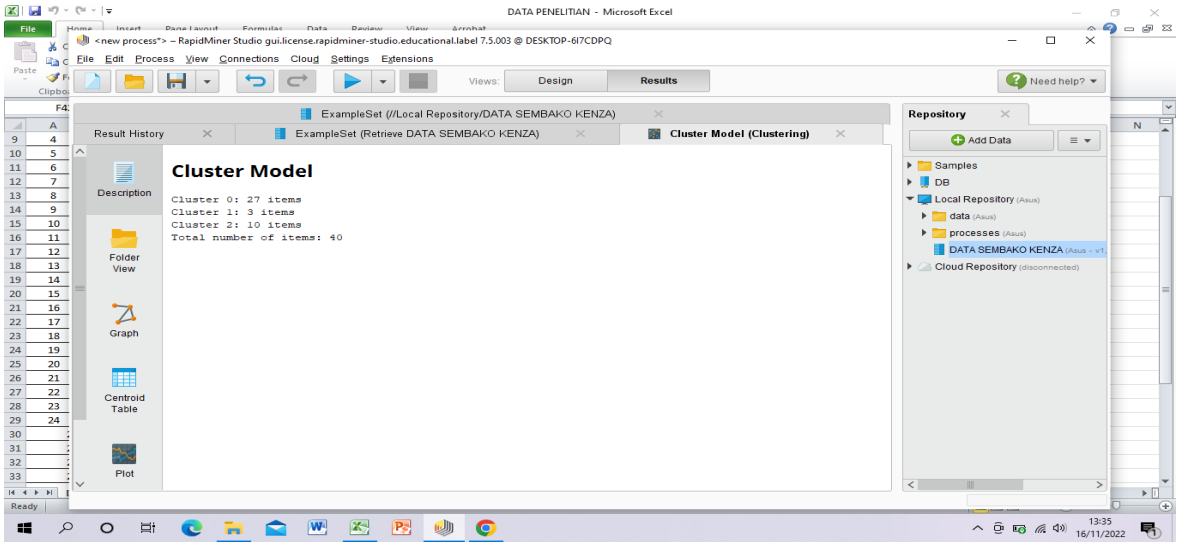

Figure 4. Cluster and Result

CONCLUSION

The conclusion that can be drawn from the K-means Clustering calculation of the Kenza grocery store is that the results of cluster 1 have 3 types of goods that are the best-selling, namely 5 kg of rice, toothpaste and powder detergent with a large amount sold per day, while in cluster 2 with 10 types of goods sold best selling, namely bulk oil, bath soap, coarse salt, LPG gas, packaged bread which are types of goods that are often purchased by consumers and must have sufficient supply for the next supply and finally cluster 3 with 27 types of goods with few buyers. So that with the clustering of Kenza grocery stores, we can provide stock according to the best-selling goods.

REFERENCES

[1] T. F. Syahputri and S. A. Suryaningsih, "Penerapan Etika Bisnis Islam Dalam Transaksi Jual Beli Sembako Di Pasar Kedurus Surabaya," *J. Ekon. dan Bisnis Islam*, vol. 5, no. 1, pp. 146–159, 2022, doi: 10.26740/jekobi.v5n1.p146-159.

[2] Junaini and R. Mulyadi, "Analisis Sistem Distribusi Penjualan Sembako Usaha Kecil Dan Menengah Dengan Menggunakan Algoritma Dijkstra Berbasis Android Di Kota Bontang," *J. Tek. Juara Aktif Glob. Optimis*, vol. 1, no. 2, pp. 55–69, 2021, doi: 10.53620/jtg.v1i2.40.

[3] И. В. Томилова, Д. С. Красавин, and Т. В. Бордовицына, "Динамическая Структура Околоземного Орбитального Пространства В Области Резонанса 1 : 2 Со Скоростью Вращения Земли," *Астрономический Вестник*, vol. 54, no. 4, pp. 337–348, 2020, doi: 10.31857/s0320930x20040088.

[4] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

[5] L. ' Izzah and A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif J. Ilm. Komput.*, vol. 18, no. 1, p. 69, 2022, doi: 10.35889/progresif.v18i1.769.

[6] H. Prastiwi, J. Pricilia, and E. Raswir, "Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)," *J. Inform. Dan Rekayasa Komput.*, vol. 1, no. April, pp. 141–148, 2022.

[7] M. Tri, I. Rahmayani, and N. Hidayati, "Implemention K-Means Algorithm Determine the Recovery Rate Of Covid- 19 Patients In Indonesia," vol. 6, no. 36, pp. 127–135, 2022.

[8] M. F. Julianto, S. W. Hadi, S. Setiaji, W. Gata, and R. Pebrianto, "Clustering Pencapaian Target Penjualan Rumah Para Karyawan Marketing Menggunakan Rapid Miner Dan Algoritma K-Means," *Bianglala Inform.*, vol. 8, no. 2, pp. 79–85, 2020, doi:

10.31294/bi.v8i2.8189.
[9] N. Hidayati and S. A. Rahmah, "Clasterization Of Zeeida Product Sales Using K - Means Method In Medan Distributors," vol. 6, no. 36, pp. 1685–1692, 2022.
[10] R. R. Putra and C. Wadisman, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means Implementation of Data Mining for Potential Customer Selection Using K-Means Algorithm," *J. Inf. Technol. Comput. Sci.*, vol. 1, no. 1, pp. 72–77, 2018.
[11] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.
[12] D. T. Saputro and W. P. Sucihermayanti, "Penerapan Klasterisasi Menggunakan K-Means untuk Menentukan Tingkat Kesehatan Bayi dan Balita di Kabupaten Bengkulu Utara," *J. Buana Inform.*, vol. 12, no. 2, p. 146, 2021, doi: 10.24002/jbi.v12i2.4861.
[13] M. A. Amri, A. P. Windarto, A. Wanto, and I. S. Damanik, "Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi Menurut Provinsi Berdasarkan Fasilitas Yang Dimiliki Desa," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 674–679, 2019, doi: 10.30865/komik.v3i1.1677.
[14] R. Rosmini, A. Fadlil, and S. Sunardi, "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, vol. 3, no. 1, pp. 22–31, 2018, doi: 10.25299/itjrd.2018.vol3(1).1773.
[15] Y. Prayoga, H. S. Tambunan, and I. Parlina, "Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means," *BRAHMANA J. Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 24–30, 2019, doi: 10.30645/brahmana.v1i1.4.