Jurnal Mantik, 6 (2) (2022) 1685-1692

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

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

Clasterization Of Zeeida Product Sales Using K-Means Method In Medan Distributors

Nani Hidayati[1], Kasini[2], Sabrina Aulia Rahmah[3]

[1] STIKOM Tunas Bangsa Pematangsiantar, Indonesia
[2]Universitas Pahlawan Tuanku Tambusai, Indonesia
[3]Universitas Dharmawangsa, Indonesia

E-mail: nanihidayati3@gmail.com[1], kasiniaqm@gmail.com[2], sabrinaaulia@dharmawangsa,ac,id[3]

ARTICLE INFO

Article history:
Received: Jun 20, 2022
Revised: Jun 30 , 2022
Accepted: Jul 12, 2022

Keywords:
Zeeida,
Clustering,
Data Mining.

ABSTRACT

Product clustering is one of the determinants of product development in sales activities. Zeeida Herbal products are engaged in health and beauty, of all the products sold, not all of them are sold, some are less well sold. Sales at the distributor of zeeida Medan products are still not computerized, namely by using manual recording. Every buyer who purchases either an agent, sub-agent, reseller or general customer who makes purchases through social media such as WhatsApp, Facebook, marketplace, and other E-Commerce is recorded in the manual bookkeeping, so there is often stock accumulation and even stock shortages at distributors. In this study, the authors apply the k-means clustering algorithm to classify products that do not sell (C0), sell very well (C1) and sell (C2). Clustering is a technique of one of the data mining functionality, the Clustering Algorithm is an algorithm for grouping a number of data into a certain data group (cluster). From this study, the output generated from the last 4 months, namely January-April 2022, shows that from 47 Zeeida products, sales of Zeeida products did not sell well in cluster 0, there were 39 products, while sales were very good in cluster (C1), there were 4 products and sales were sold in cluster. (c2) there are 4 products.

1. Introduction

Herbal products have long been known and used by the wider community, ranging from traditional herbs, health supplements, honey, to various types of medicines to treat various diseases. The emergence of various kinds of diseases until the emergence of covid-19 in the last 3 years, makes herbs very needed by the community to increase body immunity, maintain stamina in fighting viruses [1]. So that there are many herbal products on the market, especially in E-Commerce. In meeting the needs of herbs, zeeida is here to meet the needs of the community, not only herbs, zeeida also has beauty products. The products that Zeeida sells for now have 47 items and continue to grow every year following market needs. Maintaining an empty stock of goods is one way to maintain customer satisfaction [2] So it is necessary to implement information systems in the business world to produce valid [3]. The important role of data mining in determining the best-selling products so that the stock is never empty. Data mining is the process of searching for patterns or information on selected data through certain methods [4].

K-Means is one of the clustering algorithms. Clustering is one technique, the clustering algorithm is an algorithm for grouping a number of data into certain data groups (clusters) [5][6] from one of the data mining functionality, where the K-Means algorithm is very well known for its ease and ability to clarify large data and outliers very quickly [7] [8]. The K-Means method has been widely used in various cases other than sales, for example illiteracy grouping [9], cure rate of COVID-19 [10], and book data Grouping in the library [11].

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

The purpose of this research is to provide convenience for Medan distributors in knowing which goods are not selling, very selling and selling well. So that the stock is always updated and no product is close to expiration.

2. Method

This research uses data mining, in the industrial era 4.0 the development of data into big data requires a method in order to be able to convert big data to be useful in everyday life. In data mining is a way to automate the process of finding patterns from large-scale data sets [12] This method can identify objects that have certain characteristics in common, then use these characteristics as "centroids" [13]. The grouping of salable products will be calculated using the K-Means method, which is able to calculate accurately and consistently using computer-based technology [14].

The K-Means algorithm performs point-based clustering (centroid) by determining three parameters, namely the number of clusters and the distance of the system [15] which is adjusted to the object of research, both elderly and with the help of rapid miner software, the following are the steps (Figure 1).

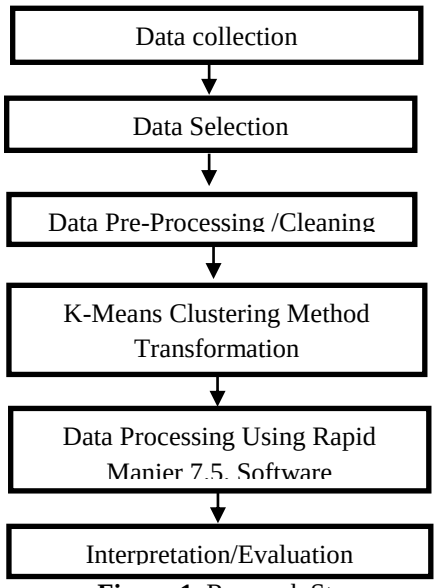

Figure 1. Research Stages

a. Data Collection
   From the Zeeida product data as many as 47 products will be processed using the K-Means algorithm and literature studies taken from several sources, namely in the form of scientific journals about the K-means method, articles and other supporting materials [16].
b. Data Selection
   Selecting zeeida sales data every day. The selected data is then entered into Excel and saved in .xls format.
c. Pre-processing/Cleaning
   The data cleaning process includes removing duplicate data, checking for inconsistent data, and correcting errors in data.
d. Transformation
   In this stage the data obtained will be tested using RapidMiner 7.5 software. For the testing phase, the data format used is in the form of read Excel .xls.
e. Data Processing Using Rapid Manier
   The process of looking for patterns will be done manually and testing using RapidMiner to generate calculation equations. Manual processes and software must produce clusters of sales that are not selling, very selling and selling.
f. Interpretation/Evaluation

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

This stage includes patterns or information that is found to contradict facts or pre-existing hypotheses.

3. Result and Discussion

This research data is taken from the Zeeida Distributor Medan, where the data used is sales data for the last 4 months, namely January-March which has 47 products. The expected output is to produce 3 clusters, namely C0 not selling, C1 very selling, C2 selling well. The variables or attributes used in grouping this sales data consist of the sales stock of Zeeida products as shown in table 1.

TABLE 1
Sample Data

NO	Produk Name	Zeeida Product Sales				Total Sales
		January	February	March	April	
1	Clove Aromatherapy	30	20	40	10	100
2	Greentea Aromatherapy	12	24	10	15	61
3	Lime Aromatherapy	24	12	6	12	54
4	Bidara leaf	2	1	3	4	10
5	Gurah Miss V	50	30	10	40	130
6	Habbat Capsules 100	30	20	10	20	80
7	Habbat Capsules 200	2	2	1	5	10
8	Lemongrass oil 30 ML	300	250	200	150	900
9	Lemongrass oil 60 ML	250	200	150	100	700
10	Lemongrass oil 100 ML	200	150	100	100	550
11	Lemongrass oil 250 ML	10	5	12	6	33
12	lavender telon oil 30 ML	200	150	200	150	700
13	lavender telon oil 60 ML	150	200	150	100	600
14	lavender telon oil 100 ML	200	150	100	100	550
15	lavender telon oil 250 ML	5	10	10	5	30
16	Lemongrass oil 30 ML BPOM	10	2	15	30	57
17	Lemongrass oil 60 ML BPOM	15	10	20	50	95
18	Lemongrass oil 100 ML BPOM	25	40	70	100	235
19	lavender telon oil 30 ML BPOM	5	2	16	30	53
20	lavender telon oil 60 ML BPOM	20	15	40	50	125
21	lavender telon oil 100 ML BPOM	7	2	10	20	39
22	Baked Candlenut Oil	210	250	200	150	810
23	Nutriclay	30	10	6	12	58
24	RGH Dewasa 250 ML	5	2	0	5	12
25	RGH Kids 250 ML	2	5	8	10	25
26	Bidara VCO Soap	30	15	20	10	75
27	Moringa Bidara Soap	20	20	30	40	110
28	Bidara Collagen Soap	70	30	50	30	180
29	Healthy uterus	24	15	20	15	74
30	Bidara Powder	15	6	8	10	39
31	Zuriat Powder	5	2	5	3	15
32	VICO 60 ML	10	15	20	10	55
33	VICO 100 ML	5	10	5	5	25
34	VICO 250 ML	2	1	3	5	11
35	VICO Caosules	15	10	20	10	55
36	Premium RGH 100 ML	15	10	9	20	54
37	RGH Kids 100 ML	24	15	20	30	89
38	Calliandra Honey FLIP 300 ML	10	3	15	10	38
39	Calliandra Honey 330 ML	5	10	5	10	30
40	Calliandra Honey 650 ML	3	5	5	4	17
41	Sari Dates Angkak 470 grams	50	30	40	50	170
42	Sari Dates Angkak 280 grams	150	100	80	50	380
43	Qusthul Hindi+Sambiloto	5	7	16	10	38
44	Hindi Qusthul Powder	1	1	2	1	5
45	Hindi Qusthul Teabags	1	1	1	1	4
46	Ath-Thiflun Madu Honey	10	12	5	12	39
47	Mozzarella oil	5	10	5	10	30

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

a. Cluster start center or centroid
The initial center or centroid is obtained randomly (randomly), for the initial determination of this cluster it can be assumed as follows:
Is known:
Number of clusters: 3 (C0 Not selling, C1 Very selling, C2 Selling)
Number of Data: 47
Number of attributes :5
Determination of the initial center of the cluster:

Taken from the -8th data as cluster 1	300	250	200	150	900
Taken from the -14th data as cluster 2	200	150	100	100	550
Taken from the -46th data as cluster 3	1	1	1	1	4

b. The next step is to calculate the distance between the data and the initial center of the cluster using the Euclidean distance equation with the formula:

$$= {}^{D(i,j)}\sqrt{\left(X_{1i} - X_{1j}\right)^2 + \left(X_{2i} - X_{2j}\right)^2 + \ldots + \left(X_{ki} - X_{kj}\right)^2}$$

Dimana:
D (*i*,*j*) = Distance of data to *i* to cluster center *j*
X *ki* = Data to *i* on attribute data to *k*
X *kj* = Center point to *j* on attribute *k*

c. The next step is to calculate the distance between the data and the initial center of the cluster using the Euclidean distance equation. After the cluster center value is obtained, then the value with the shortest distance is seen. The iteration process continues until the centroid value and the position or location of the cluster do not move anymore so that the results obtained are as follows:

TABLE 2
Shortest Distance Value

No	C0	C1	C2
1	1		
2	1		
3	1		
4	1		
5	1		
6	1		
7	1		
8		1	
9		1	
10			1
11	1		
12		1	
13			1
14			1
15	1		
16	1		
17	1		
18	1		
19	1		
20	1		
21	1		
22		1	
23	1		
24	1		
25	1		
26	1		
27	1		
28	1		
29	1		
30	1		
31	1		

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

32	1		
33	1		
34	1		
35	1		
36	1		
37	1		
38	1		
39	1		
40	1		
41	1		
42	1		
43			1
44	1		
45	1		
46	1		
47	1		

The results of manual data processing on the sale of zeida products can be seen in Figure 2:

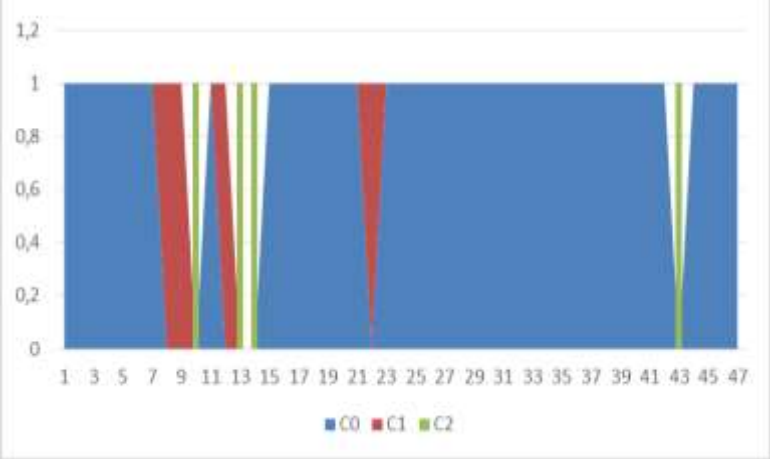

Figure 2 Manual data processing

From the calculation results, the sales table that does not sell in cluster 0 is a product with data 1,2,3,4,5,6,7,11,15,16,17,18,29,20,21,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,,41,42,44,45,46,47, while sales were very good in the cluster 1 is data 8,9, 12,22 and data cluster behavior in cluster 3 is 10,13,14 and 43. Furthermore, the implementation uses Rapid Manir software with sample data of 47 Zeeida products. The following is data processing using K-Means for sample data, which can be seen in Figure 3:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

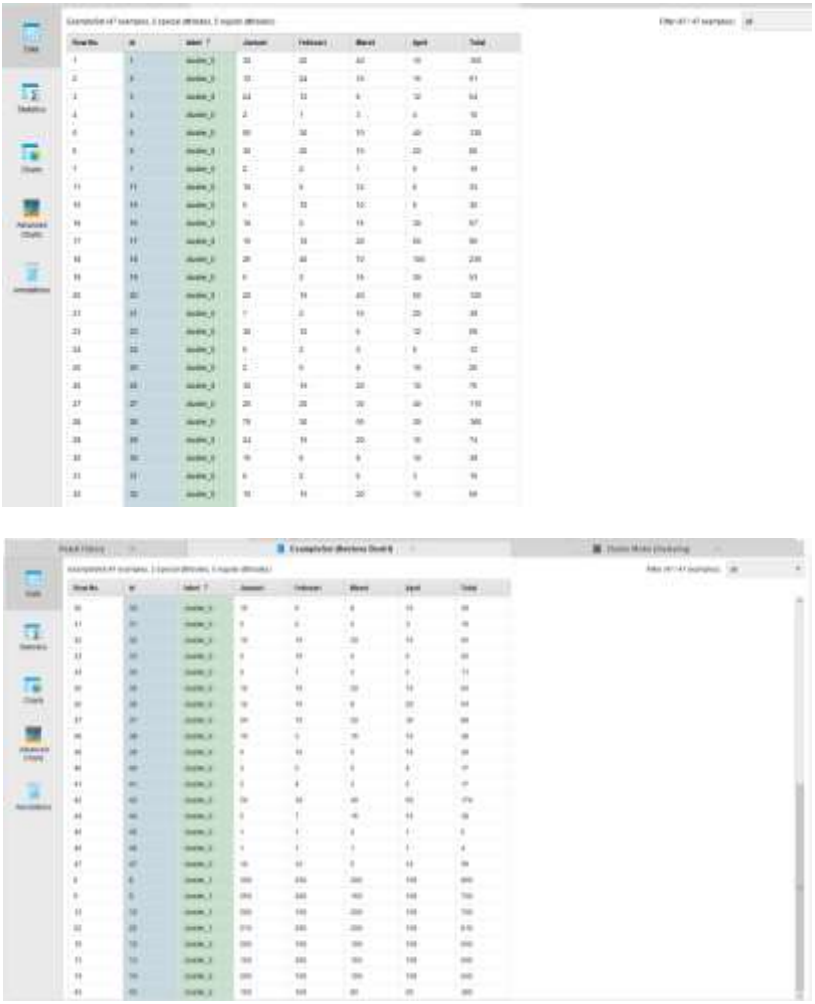

Figure 3 Results Using Rapid Miner

Cluster Model

```
Cluster 0: 39 items
Cluster 1: 4 items
Cluster 2: 4 items
Total number of items: 47
```

Figure 4 Model Cluster Display with Rapid Miner 7.5 . Software

Based on the results of the Data Mining test using the K-means Clustering algorithm that has been carried out manually and testing with the RapidMiner 7.5 software application, the following conclusions can be drawn:

TABLE 3
Perhitungan Manual Dan Aplikasi

Atribut	C0	C1	C2
Januari	609	960	700
Februari	434	850	600
Maret	589	750	430
April	715	550	350
Total Penjualan	2347	3110	2080

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

From the table above, it can be seen that the total sales attribute has the highest value in cluster 1.

4. Conclustion

The conclusions that can be drawn from the data mining technique of zeeida herbal product clustering using K-Means clustering to determine products that do not sell, sell well, and sell very well are as follows With this data grouping, Zeeida distributors can find out which products are not selling, very selling and selling well. So that the goods that the distributor has are always updated and the distributor can determine the priority of purchasing the best-selling stock of goods. From this research, the outputs produced are 39 products that are not selling well in cluster C0, 4 products that sell very well in cluster C1 and 4 products that sell well in cluster C2.

References

[1] R. Dewati and W. A. Saputro, "Persepsi Konsumen Terhadap Pembelian Produk Herbal Di Kabupaten Sukoharjo," *AGRISAINTIFIKA J. Ilmu-Ilmu Pertan.*, vol. 4, no. 2, p. 144, 2020, doi: 10.32585/ags.v4i2.889.

[2] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Clustering Sales Data at Outdoor Equipment Stores Using K-," vol. 7, no. November, pp. 109–113, 2019.

[3] Feryanto, F. T. Kesuma, and S. P. Tamba, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.

[4] W. Purba, S. Tamba, and J. Saragih, "The effect of mining data k-means clustering toward students profile model drop out potential," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, 2018, doi: 10.1088/1742-6596/1007/1/012049.

[5] M. H. Fakhriza and K. Umam, "Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Pada "Pt.Sukanda Djaya," *JIKA (Jurnal Inform.*, vol. 5, no. 1, p. 8, 2021, doi: 10.31000/jika.v5i1.3236.

[6] S. A. Rahmah and J. Antares, "Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering," *I N F O R M a T I K a*, vol. 13, no. 2, p. 25, 2022, doi: 10.36723/juri.v13i2.282.

[7] W. J. Mawaddah, I. Gunawan, and I. P. Sari, "Implementation of Data Mining Algorithm for Clustering of Palm Oil Harvested Data," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 1, pp. 43–54, 2022, doi: 10.55123/jomlai.v1i1.163.

[8] M. Mardalius, "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means," *Jurteksi*, vol. 4, no. 2, pp. 123–132, 2018, doi: 10.33330/jurteksi.v4i2.36.

[9] S. Saifullah and N. Hidayati, "Pengelompokan Persentase Buta Huruf Umur 15-44 Menurut Provinsi Menggunakan Algoritma K-Means," *Klik - Kumpul. J. Ilmu Komput.*, vol. 7, no. 3, p. 230, 2020, doi: 10.20527/klik.v7i3.329.

[10] 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.

[11] M. H. Siregar, "Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.

[12] H. Sugara, M. A. Hanafiah, P. Sekretari, P. B. Indonesia, P. K. Akuntasi, and P. B. Indonesia, "Book Data Grouping in the Library Using the K-Means Clustering Method," vol. 4, no. 3, pp. 2225–2231, 2020.

[13] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, "Book data grouping in libraries using the k-means clustering method," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, 2019, doi: 10.1088/1742-6596/1230/1/012074.

[14] A. Ardian, "Perancangan Aplikasi Pengolah Data Siswa Berbasis Android (Studi Kasus : Mis Nurul Huda Labuhan Batu Selatan)," *Pap. Knowl. . Towar. a Media Hist. Doc.*, vol. 7, no. 2, pp. 107–15, 2014.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

[15] S. Susliansyah, H. Sumarno, H. Priyono, and N. Hikmah, "Pengelompokkan Data Pembelian Tinta Dengan Menggunakan Metode K-Means," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 3, no. 2, p. 381, 2019, doi: 10.30645/j-sakti.v3i2.156.

[16] Y. Aswan, S. Defit, and G. W. Nurcahyo, "Algoritma K-Means Clustering dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai)," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 245–250, 2021, doi: 10.37034/jsisfotek.v3i4.73.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.