Jurnal Teknologi, Vol. 10 , No.02 , Oktober 2020, ISSN : 2301-4474, e-ISSN : 2541-1535

Goods Stock Management using the K-Means Algorithm Method

Dhio Saputra

[1] Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Indonesia
E-mail:dhio_saputra@upiyptk.ac.id

Abstract

The grouping of Mazaya products at PT. Bougenville Anugrah can still do manuals in calculating purchases, sales and product inventories. Requires time and data. For this reason, a research is needed to optimize the inventory of Mazaya goods by computerization. The method used in this research is K-Means Clustering on sales data of Mazaya products. The data processed is the purchase, sales and remaining inventory of Mazaya products in March to July 2019 totaling 40 pieces. Data is grouped into 3 clusters, namely cluster 0 for non-selling criteria, cluster 1 for best-selling criteria and cluster 2 for very best-selling criteria. The test results obtained are cluster 0 with 13 data, cluster 1 with 25 data and cluster 2 with 2 data. So to optimize inventory is to multiply goods in cluster 2, so as to save costs for management of Mazaya products that are not available. K-Means clustering method can be used for data processing using data mining in grouping data according to criteria.

Keyword: *K-Means, Data Mining, Clustering, Cluster, Pengelompokan*

1. INTRODUCTION

Data Mining can be implemented to group data in various cases, such as grouping the Human Immunodeficiency Virus / Acquired Immunodeficiency Syndrome (HIV / AIDS) indicator in Indonesia based on 34 provinces in Indonesia [1]. Data Mining with the K-Means Clustering algorithm has also been used to determine which employees are eligible to take part in the Assessment Center for the SDP program [2].

PT. Bougenville Anugrah is a company as a distributor of PT. Pesona Amaranthine Cosmetiques which is engaged in Mazaya Cosmetics. Due to the large number of Mazaya product data and stock items that must be analyzed, a management must work hard in analyzing and determining the next item to be purchased from the existing stock.

Based on the problems that have been described, this research will produce information about the grouping of mazaya product sales from "NOT LARIS", "LARIS" and "VERY LARIS", then it will be calculated from the existing sales using the K-Means Clustering method with the application. RapidMinner Studio, then the results can be used as a guide or knowledge base in restoring goods, especially for mazaya products at PT. Bougenville Grace.

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

2. LITERATURE

Knowledge Discovery in Database (KDD) is a method used to obtain knowledge from existing databases. The results of the knowledge obtained can be used as a knowledge base that is used for decision making purposes [3].

Data mining is the process of finding interesting patterns and knowledge of large amounts of data. Data sources can include databases, data warehouses, web, other information repositories, or data that is streamed into a dynamic system. Data mining is commonly used for classification, clustering, prediction, estimation, association and data description [4].

One of the grouping techniques of data mining is clustering. Clustering is the process of dividing data into classes or clusters based on the level of similarity. In clustering, data that have similarities are entered into the same cluster, while data that have no similarities are entered into different clusters [5].

The K-Means algorithm is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters / groups. This method partitions the data into clusters / groups so that data that has the same characteristics are grouped into the same cluster and data that has different characteristics are grouped into other groups [6].

The K-Means algorithm will group objects into K clusters. This method will search for cluster centers and cluster boundaries through an iterative process. The proximity or similarity of an object to another object or the center of the cluster is calculated using the distance function. In general, K-Means use the Euclidean distance to calculate these similarities [7].

1. The calculation steps using the K-Means algorithm are:
2. 1. Determine the value of k as the number of clusters to be formed;
3. 2. Determine the initial k centroid (cluster center point) randomly;
4. 3. Calculate the distance of each data to each cendroid using the correlation formula between two objects such as the Euclidean Distance formula;

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad \text{.....(1)}$$

Information:
dij = distance between i and j
xi = the x coordinate of the object
xj = center x coordinate
yi = the object's y coordinate
yj = center y coordinate

5. 5. Classify each data based on the closest distance between the data and the centroid;
6. 6. Determine the position of the new centroid by calculating the average value of existing data in the same cluster.

$$Ci = \frac{x1+x2+x3+\cdots+xn}{x} \quad \text{.....(2)}$$

Information:
x1 = the value of the 1st data record
x2 = 2nd record data value
$\sum x$ = number of data records

7. Return to step 3 if the position of the new centroid with the old centroid is not the same [8].

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

3. METHOD

In the preparation of this study, the framework used can be seen in the following figure:

Figure 1: Research Framework

Information :

1. Identify the Problem

The problem identified in this study is to classify Mazaya products to optimize the stocks of Mazaya products at PT.Bougenville Anugrah.

2. Analyzing the Problem

At this stage, a problem analysis is carried out to be able to understand the problems that have been determined in scope or boundaries and identify problems. By analyzing the predetermined problems, it is hoped that the problem can be understood well.

3. Determining Research Objectives

After the problem can be understood, at this stage the objectives to be achieved through the research process are stated. In this study, the goal to be achieved is to determine the clusters of mazaya products, there are 3 clusters, namely: (C0) Not Selling, (C1) Selling, (C2) Very Selling.

4. Studying Literature Studies

At this stage the researcher studies the literature study related to the research title. Literature sources are obtained from journals that discuss the K-Means Clustering Algorithm, Data Mining and other reading materials that support this research.

5. Collecting Data and Information

The purpose of this stage is to obtain information about products that are not selling, selling, and are selling well based on test results. The collection of data and information is carried out in several ways, such as searching for and studying literature and journal studies, making observations by making direct observations to PT. Bougenville Anugrah to obtain the required data and then conduct interviews with parties directly related to the problems being discussed in this study to get a more detailed description and explanation, so that the desired data can be achieved, namely complete and accurate data.

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

6. Processing Data with the K-Means Clustering Algorithm

At this stage, the data that has been obtained at the data collection stage will be processed and processed with the K-Means Clustering Algorithm, the steps are as follows:

a. Determine the number of clusters K

b. Randomly determine the center of the centroid

c. Calculates the closest distance to the centroid

d. Group data into each cluster

7. Testing Research Results

The author conducts training and testing after collecting data, processing the K-Means clustering algorithm, and installing the software that will be used, namely Rapid Miner, for further training and testing of the data. Testing is data obtained from initial data processing using the K-Means clustering method. The results obtained from the test will be analyzed so that the conditions are more accurate.

8. Concluding the test results

At this stage the writer will get the results of the K-Means Clustering Algorithm calculation using Excel and the Rapid Miner application. From the results of this grouping, several groups of mazaya products will be obtained according to their respective clusters, and the author can recommend the results to the management of PT. Bougenville Anugrah as a consideration for decision making to optimize Mazaya stock.

4. RESULT

From the Knowledge Discovery in Database (KDD), the K-Means Clustering algorithm can be understood that the steps for solving it can be explained as follows:

1. Data Selection Process

In the Knowledge Discovery in Database (KDD) process, data selection is the process of data collection and creating target data on the data sample, where the discovery will be made and the selection results are stored in a separate file from the operational database data. After the data is obtained from PT. Bougenville Anugrah, like the initial data above, a data selection process is carried out, this process is carried out to facilitate the calculation process of the K-Means Clustering algorithm. The data used are data from March to July 2019.

2. Data Input

The data entered is the data that has been obtained from PT. Bougenville Anugrah from March to July 2019 which can be seen in table 1 above. Before inputting the data into the system, it must be analyzed and carried out the calculation process correctly, while the process of the K-means clustering algorithm is as follows:

a. Determine the number of clusters (K)

K here is the number of clusters to be formed. In this study, 3 types of clusters were formed, namely purchasing, selling and remaining stock obtained from the monthly data recapitulation of Mazaya PT. Bougenville Anugrah which was taken from March to July 2019, the cluster is divided into non-selling, best-selling and very in-demand.

b. Randomly Determining the Center Point of the Centroid

After determining the number of clusters as many as 3 clusters, then the center point is determined randomly here taken for the center cluster 0 (C0): taken from number 5 for the

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

center cluster 1 (C1): taken from number 18 and the center of cluster 2 (C2) : taken from number 31 can be seen from the random value table as follows :

Table 1: Random Centroid Values

Centroid	Pembelian	Penjualan	Sisa Stok
C_0	181	173	8
C_1	96	81	15
C_2	302	257	45

a. Calculates the Nearest Centroid Distance

The euclidean distance formula is used to calculate the distances of the center object to each centroid so that the closest distance from each data to the centroid is found. The following is the equation formula Classifying the data for each cluster Classify each data based on its proximity to the shortest distance centroid. The calculation process is as follows:

1) Iteration Process I (First)

To calculate the center of cluster 0, cluster 1 and cluster 2, at this stage the distance of each data will be calculated using the Euclidean Distance (D) formula in the first iteration. Calculate the distance of each data to the center point as follows:

Euclidean Distance formula:

$$D_{L_i}(x_2,x_1) = \|x_2 - x_1\| \sum_{j=1}^{p} |x_{2j} - x_{ij}| \quad D_{L_2} = (x_2,x_1) = \|x_2 - x_1\| = \sqrt{\sum_{j=1}^{p} (x_{2j} - x_{1j})^2}$$

1. Proses Hitung cluster (C_0)

$X_1 = (66; 21; 45)$ $C_{1.1} = (181; 173; 8)$

$D_{11} = \sqrt{(66-181)^2 + (21-173)^2 + (45-8)^2}$

$= 194{,}1597$

2. Proses Hitung cluster (C_1)

$X_2 = (66; 21; 45)$ $C_{1.2} = (96; 81; 15)$

$D_{12} = \sqrt{(66-96)^2 + (21-81)^2 + (45-15)^2}$

$= 73{,}48469$

3. Proses Hitung cluster (C_2)

$X_3 = (66; 21; 45)$ $C_{1.3} = (302; 257; 45)$

$D_{13} = \sqrt{(66-302)^2 + (21-257)^2 + (45-45)^2}$

$= 333{,}7544$

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

Perform the calculation process for the above data up to 40 times according to the amount of data, the calculation results can be seen in table 2 obtained from the calculation of the Euclidean Distance in iteration I above, the comparison of the distance from each cluster C0, C1, C2 is obtained.

a. Grouping Data Into Each Cluster

After the data has been processed and calculated in the steps above, the results of the above process are grouped as follows:

Table 3: Cluster Data Clusters with the Closest Distance

No	C_0	C_1	C_2	Jarak Terdekat C_0	Jarak Terdekat C_1	Jarak Terdekat C_2
1	194,15973	73,48469	333,7544		1	
2	199,76486	82,74056	336,1458		1	
3	190,51509	65,28399	335,44895		1	
4	150,73818	25,57342	298,15768		1	
5	0	125,4512	151,87495	1		
6	135,32184	71,37226	257,39852		1	
7	93,08061	40,47221	234,77223		1	
8	77,239886	78,43469	203,70567	1		
9	71,386273	54,461	218,57264		1	
10	195,36632	71,93052	337,47		1	
11	207,64874	82,42572	352,20165		1	
12	136,565	18,38478	280,08213		1	
13	98,173316	36,76955	239,61636		1	
14	154,61565	33,25658	297,47941		1	
15	59,866518	69,46942	203,16004	1		
16	93,155784	37,17526	236,61361		1	
17	172,72522	49,31531	315,8354		1	
18	125,45119	0	272,60227		1	
19	194,05669	70,02857	336,90355		1	
20	215,30908	89,87769	361,43326		1	
21	191,64551	68,94926	333,15762		1	
22	216,43013	91,78235	359,51078		1	
23	188,88621	65,55913	331,15253		1	
24	192,81079	68,27884	336,37776		1	
25	191,7707	67,63135	334,84026		1	
26	202,86449	77,76889	347,18295		1	
27	186,01613	61,18823	330,24839		1	
28	172,6731	59,71599	309,12457		1	
29	124,42669	43,1509	258,84745		1	
30	224,83327	100,04	367,98913		1	
31	151,87495	272,6023	0			1
32	301,84764	424,0118	151,45296			1
33	136,03676	35,69314	273,65307		1	
34	167,82729	204,7242	187,08287	1		

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

35	164,54179	169,1094	221,68897	1		
36	128,15615	183,0792	154,4215	1		
37	205,96602	81,03086	349,85997		1	
38	181,3946	64,94613	318,96395		1	
39	205,77172	80,94443	353,4855		1	
40	87,988636	44,45222	229,93912		1	
Jumlah				**6**	**32**	**2**

Based on the calculation process obtained in iteration I according to table 3 above, a cluster of each has been obtained where the results of C0 are 6 data, the results of C1 are 32 data and the results of C2 are 2 data.

1) Iteration Process II (Second)

From table 3 above, it is found that the group of members C0, C1 and C2 is the result of this group to get new centroid values as follows :

a. Anggota cluster 0 (C0) terdiri dari 6 anggota
   C1(X1) = (181+160+149+252+213+243)/6
   C1(X1) = **199,6667**
   C1(Y2) = (173+111+125+107+76+131)/6
   C1(Y2) = **120,5**
   C1(Z3) = (8+49+24+145+137+112)/6
   C1(Z3) = **79,1667**
b. Anggota cluster 1 (C1) terdiri dari 32 anggota
   C2(X1) = (66+70+55+77+135+129+135+59+44+96+126+85+126 +72+96+58+34+63+42+63+57+59+48+60+87+121+36 +108+47+79+36+132)/32
   C2(X1) = **78,1563**
   C2(Y2) = (21+14+31+64+63+99+119+23+18+68+95+54+100+41 +81+25+16+25+9+28+27+27+21+33+35+72+3+64+18 +29+27+103)/32
   C2(Y2) = **45,4063**
   C2(Z3) = (45+56+24+13+72+30+16+36+26+28+31+31+26+31 +15+33+18+38+33+35+30+32+27+27+52+49+33+44+29 +50+9+29)/32
   C2(Z3) = **32,75**
c. Anggota cluster 2 (C2) terdiri dari 2 anggota
   C3(X1) = (302+415)/2
   C3(X1) = **358,5**
   C3(Y2) = (257+357)/2
   C3(Y2) = **307**
   C3(Z3) = (45+58)/2
   C3(Z3) = **51,5**

After the above calculation process is complete, a new centroid will be obtained which can be seen in table 4 with the following values:

Table 4: Second Iteration New Centroid Value

Centroid	Pembelian	Penjualan	Sisa Stok
C_0	199,6667	120,5	79,1667
C_1	78,1563	45,41	32,75
C_2	358,5	307	51,5

After obtaining the new centroid, then iterating again like the above calculation process by calculating the center of cluster 0, cluster 1 and cluster 2 of each centroid using the Euclidean Distance (D) formula in iteration II, the above process is repeated until the

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

iteration does not change the centroid value in Before and after iterations, this study stopped at the 6th iteration after the centroid values did not change. Then the result is that cluster 0 (C0) for not selling there are 13 data, cluster 1 (C1) for best-selling has 25 data and cluster 2 (C2) for very in demand there are 2 data. More details about Mazaya product grouping can be seen in table 5 as follows :

Table 5: Overall Grouping Results

NO	PEMBELIAN	PENJUALAN	SISA STOK	KEPUTUSAN
1	66	21	45	LARIS
2	70	14	56	LARIS
3	55	31	24	LARIS
4	77	64	13	LARIS
5	181	173	8	TIDAK LARIS
6	135	63	72	TIDAK LARIS
7	129	99	30	TIDAK LARIS
8	160	111	49	TIDAK LARIS
9	135	119	16	TIDAK LARIS
10	59	23	36	LARIS
11	44	18	26	LARIS
12	96	68	28	LARIS
13	126	95	31	TIDAK LARIS
14	85	54	31	LARIS
15	149	125	24	TIDAK LARIS
16	126	100	26	TIDAK LARIS
17	72	41	31	LARIS
18	96	81	15	LARIS
19	58	25	33	LARIS
20	34	16	18	LARIS
21	63	25	38	LARIS
22	42	9	33	LARIS
23	63	28	35	LARIS
24	57	27	30	LARIS
25	59	27	32	LARIS
26	48	21	27	LARIS
27	60	33	27	LARIS
28	87	35	52	LARIS
29	121	72	49	TIDAK LARIS
30	36	3	33	LARIS
31	302	257	45	SANGAT LARIS
32	415	357	58	SANGAT LARIS
33	108	64	44	LARIS
34	252	107	145	TIDAK LARIS
35	213	76	137	TIDAK LARIS
36	243	131	112	TIDAK LARIS

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020

37	47	18	29	LARIS
38	79	29	50	LARIS
39	36	27	9	LARIS
40	132	103	29	TIDAK LARIS

5. CONCLUTION

The K-Means Clustering method was successfully applied to classify Mazaya products which consisted of not selling, selling well and being very in demand. The results of this study can facilitate PT. Bougenville Anugrah in optimizing the stock of Mazaya products. Based on the data used to classify Mazaya products from March to July 2019 with a total of 40 pieces of data. Obtaining the results of grouping that cluster 0 (C0) for not selling are 13 data, cluster 1 (C1) is for best-selling as much as 25 data and cluster 2 (C2) for best-selling as much as 2 data. So that for products in cluster 2 the stock can be expanded to reduce stock shortages, and for cluster 0 to reduce the amount of stock in order to minimize costs.

REFERENCE

[1] K. Fatmawati, dan A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi", *CESS (Journal of Computer Engineering System and Science)*, vol. 3 no. 2, hal. 173-178, 2018. doi: 10.24114/cess.v3i2.9661.

[2] I. Parlina, A. P. Windarto, A. Wanto, dan M. R. Lubis, " Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program SDP", *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 1, hal. 87-93, 2018. doi: 10.24114/cess.v3i1.8192.

[3] Gustientiedina, M. H. Adiya, dan Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru", *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, hal. 17-24, 2019. doi: https://doi.org/10.25077/ TEKNOSI.v5i1.2019.17-24.

[4] N. Agustina, dan Prihandoko, "Perbandingan Algoritma K-Means Dengan Algoritma Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan", *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 2, no. 3, hal. 621-626, 2018. doi: https://doi.org/10.29207/resti.v2i3.492.

[5] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)", *SISTEMASI*, vol. 7, no. 3, hal. 238-249, 2018. doi: 10.32520/stmsi.v7i3.388.

[6] D. Triyansyah, dan D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing", *IncomTech, Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 3, hal. 163-182, 2018. doi: 10.22441/incomtech.v8i2.4174.

[7] I. W. A. W. Kusuma, dan R. L. Ellyana, "Penerapan Citra Terkompresi Pada Segmentasi Citra Menggunakan Algoritme K-MEANS", *JUTEI*, vol. 2, no. 1, hal. 65-74, 2018. doi: 10.21460/jutei.2018.21.65.

[8] F. Mahmuda, M. Sitorus, H. Widyastuti, dan D. Kurniawan, "Clustering Profil Pengunjung Perpustakaan Menggunakan Algoritma K-Means", *JAIC*, vol. 1, no. 1, Hal. 14-21, Oct. 2018.

Diterima Redaksi : 18-04-2020 | Selesai Revisi : 18-04-2020 | Diterbitkan Online : 30-04-2020