



OPTIMIZATION OF THE NUMBER OF CLUSTERS ON K-MEDOIDS USING CHEBYCHEV AND MANHATTAN ON GOLD SELLING GROUPING

Dedi Triyanto¹, Deny Kurniawan², Mochamad Wahyudi³

¹²Sistem Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen, Jakarta 10450, Indonesia

³Teknologi Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen, Jakarta 10450, Indonesia

Email: dedi@bsi.ac.id¹, deni@bsi.ac.id², wahyudi@bsi.ac.id³

ARTICLE INFO

ABSTRACT

Article history:

Received: September, 06 2021

Revised: October, 09 2021

Accepted: November, 30 2021

Keywords:

Chebychev Distance;

Clustering;

Data mining;

K-Medoids;

Manhattan Distance.

Gold is a type of precious metal that can maintain value and can be used for exchange. Gold has attractive properties, so many people like to buy gold for jewelry and also for investments that can be resold when they need money quickly. During the COVID-19 pandemic, some sales sectors experienced a decline but gold was still selling well. M. Siregar Gold Shop serves gold jewelry sales. Gold jewelery sales transactions at the M. Siregar gold shop are stored in the database. Every day the transaction data is increasing, so the data is getting more and more. From the mountains of data we can dig up information or generate knowledge. M. Siregar's gold shop has difficulty in knowing the type of gold that is selling well, making it difficult for gold shop owners to determine the right gold supply. This study aims to classify gold sales at the M. Sisregar gold shop so that it is known which types of gold are selling well. This grouping uses the K-Medoids method with the calculation of the distance between the Chebychev distance and the Manhattan distance. The data is taken from the sales of gold at the M. Siregar store from November 2021 to March 2022. To produce an optimal grouping, this grouping is tested with several number of clusters by calculating the distance between Chebychev distance and Manhattan distance by calculating the DBI value of each number of clusters. . The result of the optimal grouping of gold sales is the K-Medoids method with the calculation of the Chebychev distance with the number of clusters = 2 with DBI value = 0.024.ns.

Copyright © 2022 Mantik Journal.
All rights reserved.

1. Introduction

For more than two years, Indonesia has experienced the COVID-19 pandemic which has had a negative impact on the Indonesian economy. Various sales sectors have been greatly impacted by this pandemic. Many shopping centers or malls were closed to comply with health protocols and caused sales to drop sharply.

Many entrepreneurs went bankrupt due to the declining sales trend. In contrast to gold sales, which remained stable, some even experienced an increase.

Clustering is a common unattended machine learning, data sets must be automatically partitioned into clusters, so that objects in the same cluster are more similar, while objects in different clusters are more different[1].

K-Medoids is one method of partitioning, because it uses the most centered object in the cluster to be the center of the cluster of the average value of objects in a cluster[2].

Research [3], the results of forecasting sales of gold bullion at PT Antam (Persero) Tbk with 6 trend methods that have been applied in this study, it is found that the ordinary least square method has a forecast



of gold sales in 2017 of 6721,982 kg with MSE 241671, while the semi-average method obtained 8053,588 kg with an MSE of 214163.11. Of the 6 methods, the ordinary least square method has the smallest MSE.

Research [4], the lowest MAPE results obtained lies in the SES method when the parameter = 0.4 with a value of 56.64. Based on the MAPE value that has been obtained using the three methods of Exponential Smoothing, the results are above 50, so the Exponential Smoothing method is not suitable for calculating gold sales forecasting.

M. Siregar Gold Shop serves gold jewelry sales. Gold jewelery sales transactions at the M. Siregar gold shop are stored in the database. Every day the transaction data is getting bigger and bigger. From the mountains of data we can dig up information or generate knowledge. M. Siregar's gold shop had difficulty knowing the type of gold that was selling well, so it was difficult for the gold shop owner to determine the right gold supply. This study has the aim of classifying gold sales at the M. Sisregar gold shop so that it can be seen which types of gold are selling well.

This study focuses on grouping gold sales using the K-Medoids method with the calculation of the Chebychev distance and Manhatttan Distance. The data is taken from the sales of gold at the M. Siregar store from November 2021 to March 2022.

2. Research Methods

2.1 K-Medoids

K-Medoids Clustering emerged as a solution to the weaknesses of the K-Means Algorithm [5]. K-Medoids is a method used to cluster a group of n objects into a number of k clusters. Further in his research, K-Medoids have reliability in overcoming noise [6].

2.2 Chebycev Distance

Chebychev distance is a distance calculation that is in the K-means or K-Medoids algorithm grouping method. Here is the formula for calculating distance with chebycev distance[7]:

$$d_{che}(x,y) = \max_{1 \leq k \leq d} |x_k - y_k| \quad (1)$$

2.3 Manhattan Distance

The Manhattan distance measurement is based on the difference between the two objects that are added together and the results obtained are in absolute terms. The form of the formula is as follows[8]:

$$d(i,j) = |x_{i,1} - x_{j,1}| + \dots + |x_{i,2} - x_{j,2}| + \dots + |x_{i,n} - x_{j,n}| \quad (2)$$

2.4 Research Framework

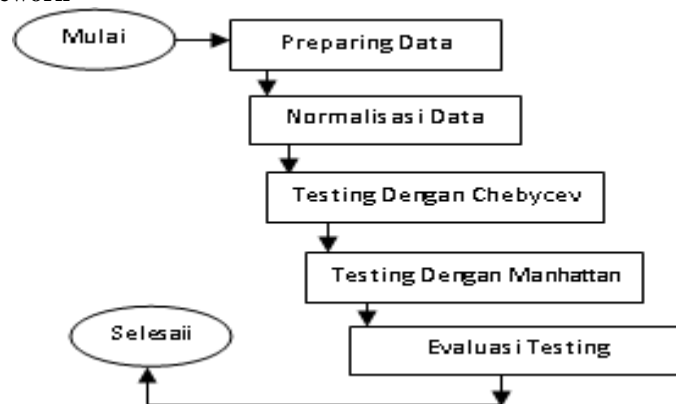


Figure 1. Research Stages

3. Results and Discussion

3.1 Data analysis

The data is taken from the sales of gold at the M. Siregar store from November 2021 to March 2022. This data consists of 51 items of gold type and size of gold as well as sales results from November, December, January, February, and March 2022. Result data sales at the M. Siregar gold shop can be seen in table 1.

TABLE 1
SELLING GOLD AT M. SIREGAR SHOP

No	Gold Type	Size	November	December	January	February	March
1	Flower Wrapping Ring	1 - 2 Mayam	0	0	1	0	0
2	Rail Necklace	1.5 - 5.99 Grams	1	3	0	1	2
3	Rail Necklace	1 - 6 Mayam	10	5	9	0	2
4	Gubsi Earrings	0.5 - 0.72 Gram	5	1	3	0	2
5	Nail Earrings	0.53 Gram	1	0	0	0	0
6	Gold bar	2.04 - 20.75 Gram	7	13	16	2	2
7	Gold bar	21.04 - 28.43 Grams	3	1	2	1	0
8	Gold bar	60.91 Gram	1	0	0	0	0
9	Rante's Head Ring	1 Gram - 2 Mayam	16	11	25	6	16
10	Rante's Head Ring	3 - 5 Mayam	2	1	2	0	0
11	Children's Rings	0.3 - 1 Gram	2	0	1	0	0
12	Flower Bracelet	4 - 8 Mayam	2	0	4	0	1
13	Swivel Ring	1 - 10 Mayam	3	0	3	1	0
14	Flower Ring	1 - 5 Mayam	5	3	6	1	0
15	Wrap Ring	1 - 1.25 Gram	0	1	1	2	1
16	Wrap Ring	- 3 Mayam	1	7	8	8	7
17	Kendari Earrings	2.32 - 4.5 Gram	1	0	4	0	0
18	Centipede Bracelet	2.03 - 29 Grams	0	2	0	0	2
19	Centipede Bracelet	- 10 Mayam	4	4	11	6	2
20	Wrap Ring	5 - 7.84 Grams	1	0	1	0	0
21	Brotan Ring	0.45 - 5 Gram	1	0	4	0	4
22	Brotan Ring	- 2½ Mayam	6	8	10	7	9
23	Bamboo Bracelet	3 - 10 Mayam	10	0	2	2	1
24	Sun Ring	1½ Mayam	1	0	0	0	0
25	Silver Ring	89 Grams	1	0	0	0	0
26	Earrings	0.6 - 2 Gram	1	0	4	0	0
27	Plain Wrap Ring	1 - 3 Mayam	2	1	0	0	0
28	Diamond Ring	2 Mayam	1	0	0	0	0
29	Cartier Bracelet	0.5 Gram	1	0	0	0	0
30	Fashion Ring	1.25 - 3.1 Gram	1	0	1	0	1
31	Channel Ring	0.5 - 0.6 Gram	1	0	1	0	0
32	Diamond Bracelet	7.9 - 10 Grams	1	1	0	0	0
33	Ring Earrings	1 Gram	1	0	0	0	0
34	Bamboo Ring	2 - 5 Mayam	3	0	1	1	1
35	SSS ring	2 Mayam	2	0	0	0	0
36	Channel Bracelet	0.75 Gram	1	0	0	0	0
37	Stamp Ring	5 Mayam	1	0	0	0	0
38	LV Cincin ring	0.6 - 0.64 Gram	1	0	1	1	0
39	Balige ring	5 Mayam	0	5	5	0	0
40	Boxer Ring	0.5 - 1.5 Gram	0	0	6	3	1
41	Children's Bracelet	1 - 19 Grams	0	1	1	0	2
42	Earrings	0.6 - 0.9 Gram	0	0	4	0	0
43	Board Bracelet	5 Mayam	0	0	1	0	0
44	Italian necklace	4.6 Gram	0	0	1	0	0
45	Wire Bracelet	5 mayam	0	0	1	0	0
46	Keroncong Bracelet	3.76 - 6.38 Gram	0	0	1	0	1
47	Ariso Bracelet	2 Grams	0	0	1	0	0
48	Rail Bracelet	2.05 - 3 Grams	0	1	0	1	1
49	Rail Bracelet	2½ - 6 Mayam	0	0	1	0	1
50	Bamboo Necklace	5 - 8 Mayam	0	0	1	0	5
51	Golden Toys	0.6 - 4.2 Gram	0	0	2	0	1

Before the grouping process is carried out using the K-Medoids method with the calculation of the Chebycev distance and Manhattan distance, the data is normalized using the formula:

$$a' = \frac{(a - \min.a)}{(\max.a - \min.a)} \quad (3)$$

Information :

= Normalized data



a = Original data
 min.a = smallest data
 max.a = largest data

The results of normalization can be seen in table 2

TABLE 1
 DATA NORMALIZATION

No	Gold Type	Size	November	December	January	February	March
	Flower Wrapping		0.0000	0.0000	0.0625	0.0000	0.0000
1	Ring	1 - 2 Mayam					
2	Rail Necklace	1.5 - 5.99 Grams	0.0625	0.1875	0.0000	0.0625	0.1250
3	Rail Necklace	1 - 6 Mayam	0.6250	0.3125	0.5625	0.0000	0.1250
4	Gubsi Earrings	0.5 - 0.72 Gram	0.3125	0.0625	0.1875	0.0000	0.1250
5	Nail Earrings	0.53 Gram	0.0625	0.0000	0.0000	0.0000	0.0000
6	Gold bar	2.04 - 20.75 Gram	0.4375	0.8125	1.0000	0.1250	0.1250
		21.04 - 28.43	0.1875	0.0625	0.1250	0.0625	0.0000
7	Gold bar	Grams					
8	Gold bar	60.91 Gram	0.0625	0.0000	0.0000	0.0000	0.0000
		1 Gram - 2	1.0000	0.6875	1.5625	0.3750	1.0000
9	Rante's Head Ring	Mayam					
10	Rante's Head Ring	3 - 5 Mayam	0.1250	0.0625	0.1250	0.0000	0.0000
11	Children's Rings	0.3 - 1 Gram	0.1250	0.0000	0.0625	0.0000	0.0000
12	Flower Bracelet	4 - 8 Mayam	0.1250	0.0000	0.2500	0.0000	0.0625
13	Swivel Ring	1 - 10 Mayam	0.1875	0.0000	0.1875	0.0625	0.0000
14	Flower Ring	1 - 5 Mayam	0.3125	0.1875	0.3750	0.0625	0.0000
15	Wrap Ring	1 - 1.25 Gram	0.0000	0.0625	0.0625	0.1250	0.0625
16	Wrap Ring	- 3 Mayam	0.0625	0.4375	0.5000	0.5000	0.4375
17	Kendari Earrings	2.32 - 4.5 Gram	0.0625	0.0000	0.2500	0.0000	0.0000
	Centipede		0.0000	0.1250	0.0000	0.0000	0.1250
18	Bracelet	2.03 - 29 Grams					
	Centipede		0.2500	0.2500	0.6875	0.3750	0.1250
19	Bracelet	- 10 Mayam					
20	Wrap Ring	5 - 7.84 Grams	0.0625	0.0000	0.0625	0.0000	0.0000
21	Brotan Ring	0.45 - 5 Gram	0.0625	0.0000	0.2500	0.0000	0.2500
22	Brotan Ring	- 2½ Mayam	0.3750	0.5000	0.6250	0.4375	0.5625
23	Bamboo Bracelet	3 - 10 Mayam	0.6250	0.0000	0.1250	0.1250	0.0625
24	Sun Ring	1½ Mayam	0.0625	0.0000	0.0000	0.0000	0.0000
25	Silver Ring	89 Grams	0.0625	0.0000	0.0000	0.0000	0.0000
26	Earrings	0.6 - 2 Gram	0.0625	0.0000	0.2500	0.0000	0.0000
27	Plain Wrap Ring	1 - 3 Mayam	0.1250	0.0625	0.0000	0.0000	0.0000
28	Diamond Ring	2 Mayam	0.0625	0.0000	0.0000	0.0000	0.0000
29	Cartier Bracelet	0.5 Gram	0.0625	0.0000	0.0000	0.0000	0.0000
30	Fashion Ring	1.25 - 3.1 Gram	0.0625	0.0000	0.0625	0.0000	0.0625
31	Channel Ring	0.5 - 0.6 Gram	0.0625	0.0000	0.0625	0.0000	0.0000
32	Diamond Bracelet	7.9 - 10 Grams	0.0625	0.0625	0.0000	0.0000	0.0000
33	Ring Earrings	1 Gram	0.0625	0.0000	0.0000	0.0000	0.0000
34	Bamboo Ring	2 - 5 Mayam	0.1875	0.0000	0.0625	0.0625	0.0625
35	SSS ring	2 Mayam	0.1250	0.0000	0.0000	0.0000	0.0000
36	Channel Bracelet	0.75 Gram	0.0625	0.0000	0.0000	0.0000	0.0000
37	Stamp Ring	5 Mayam	0.0625	0.0000	0.0000	0.0000	0.0000
38	LV Cincin ring	0.6 - 0.64 Gram	0.0625	0.0000	0.0625	0.0625	0.0000
39	Balige ring	5 Mayam	0.0000	0.3125	0.3125	0.0000	0.0000
40	Boxer Ring	0.5 - 1.5 Gram	0.0000	0.0000	0.3750	0.1875	0.0625
	Children's		0.0000	0.0625	0.0625	0.0000	0.1250
41	Bracelet	1 - 19 Grams					
42	Earrings	0.6 - 0.9 Gram	0.0000	0.0000	0.2500	0.0000	0.0000
			0.0000	0.0000	0.0625	0.0000	0.0000
43	Board Bracelet	5 Mayam					
44	Italian necklace	4.6 Gram	0.0000	0.0000	0.0625	0.0000	0.0000
45	Wire Bracelet	5 mayam	0.0000	0.0000	0.0625	0.0000	0.0000
	Keroncong		0.0000	0.0000	0.0625	0.0000	0.0625
46	Bracelet	3.76 - 6.38 Gram					
47	Ariso Bracelet	2 Grams	0.0000	0.0000	0.0625	0.0000	0.0000
48	Rail Bracelet	2.05 - 3 Grams	0.0000	0.0625	0.0000	0.0625	0.0625
49	Rail Bracelet	2½ - 6 Mayam	0.0000	0.0000	0.0625	0.0000	0.0625
50	Bamboo Necklace	5 - 8 Mayam	0.0000	0.0000	0.0625	0.0000	0.3125
51	Golden Toys	0.6 - 4.2 Gram	0.0000	0.0000	0.1250	0.0000	0.0625



3.1.1 Cluster Evaluation Using Chebycev and Manhattan Distance Calculations Based on DBI . Values

The principle of the DBI measurement approach is to maximize the inter-cluster distance and minimize the intra-cluster distance. The smaller the DBI value indicates the most optimal cluster scheme. Here is the formula to find the DBI value [8]:

$$R_{j,k} = \frac{MAE_j + MAE_k}{d(c_j c_k)} \tag{4}$$

$$DBI = \frac{1}{M} \sum_{j=1}^M \max_{j \neq k} R_{j,k} \tag{5}$$

3.2 DBI Value from Test Results

The following is the DBI value of K-Medoids Clustering. To classify gold sales using the calculation of the Chebycev distance with several number of clusters.



Figure 2. Number of Clusters = 2



Figure 3. Number of Clusters = 3

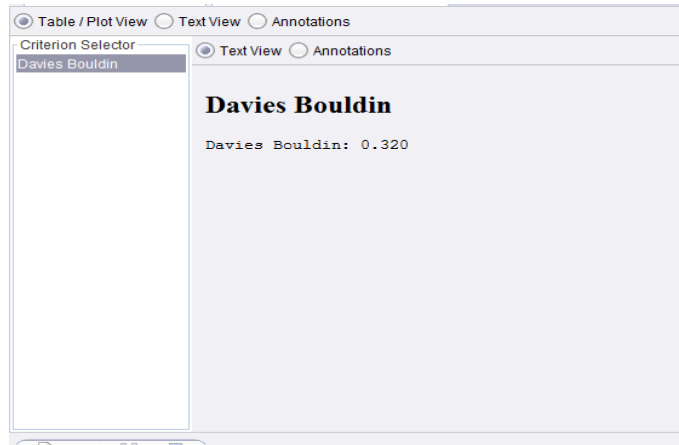


Figure 4. Number of Clusters = 4

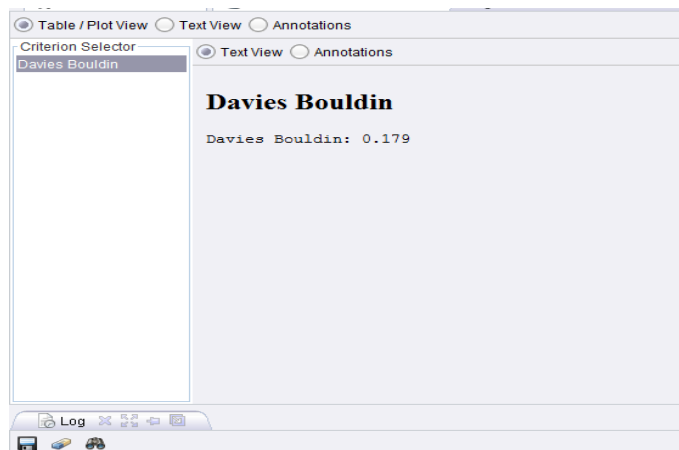


Figure 5. Number of Clusters = 5

The following is the DBI value from K-Medoids Clustering. To classify gold sales using the calculation of the Manhattan distance with several number of clusters.

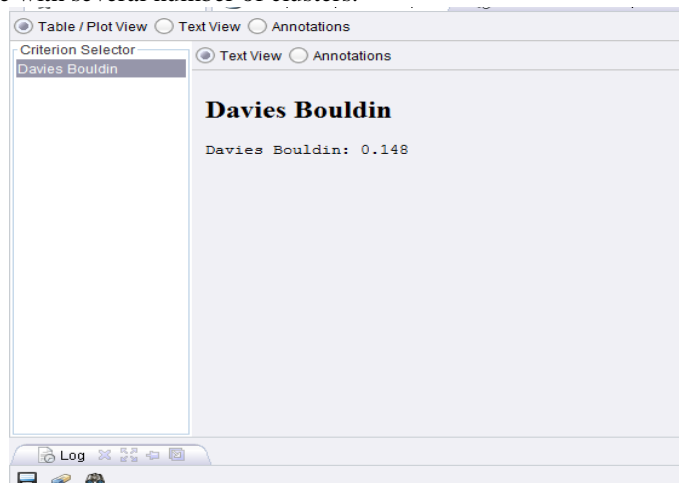


Figure 6. Number of Clusters = 2

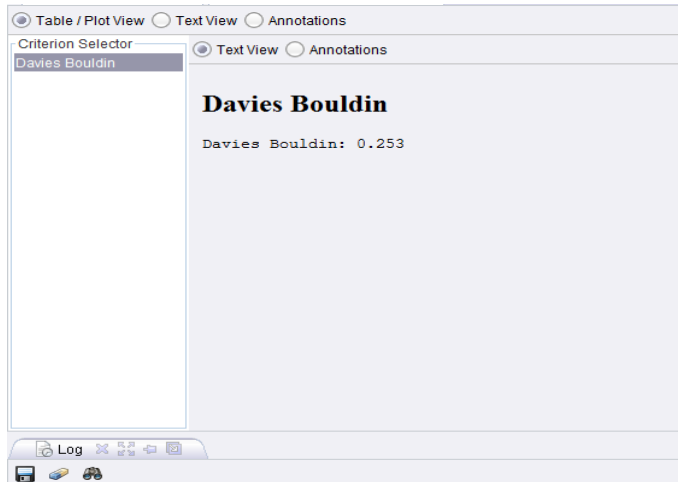


Figure 7. Number of Clusters = 3

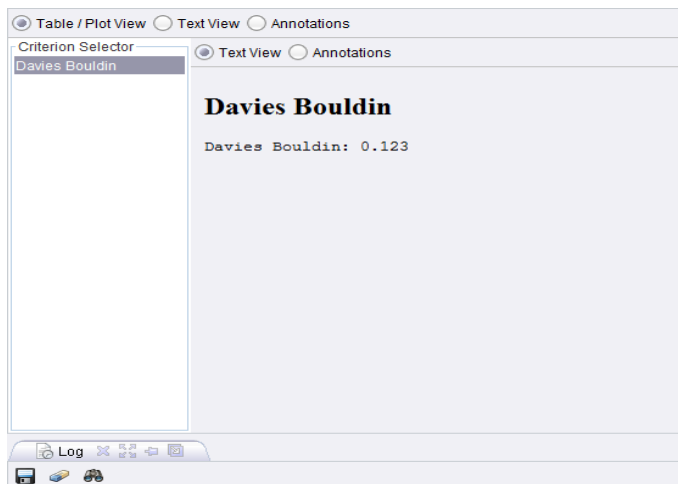


Figure 8. Number of Clusters = 4

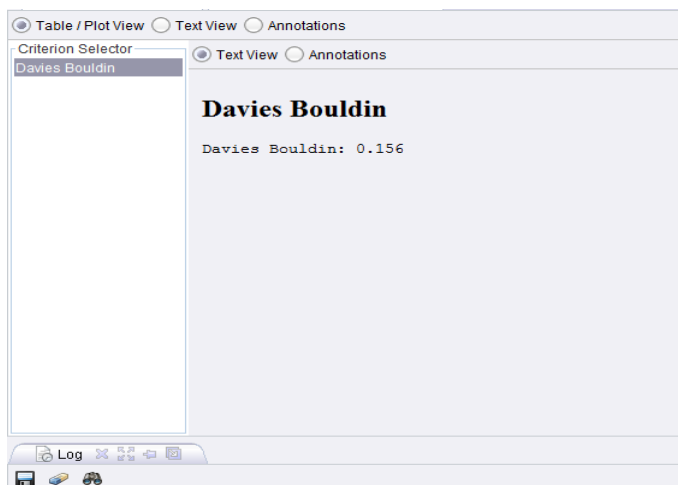


Figure 9. Number of Clusters = 5

From the test results using rapidminer 5.0, it shows that the grouping of gold sales using the K-Medoids method by calculating the Chebycev distance distance with several number of clusters and by calculating the

Manhattan distance with several number of clusters can be seen in Figure 10 and Table 3. The smallest DBI value is 0.024, namely sales grouping. gold method of K-Medoids with the calculation of the distance chebycev distance with the number of clusters = 2.

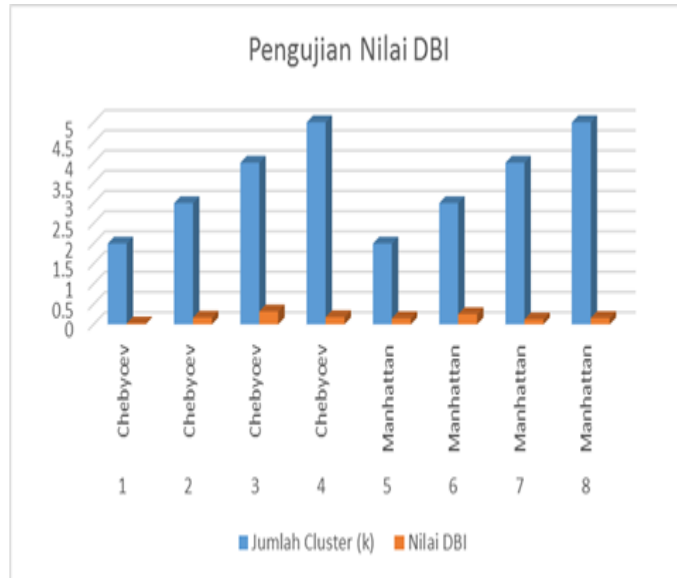


Figure 10. Testing DBI Value

TABLE 1
DBI VALUE TESTING

No	Distance Calculation	Number of Clusters (k)	DBI Value
1	Chebycev	2	0.024
2	Chebycev	3	0.166
3	Chebycev	4	0.32
4	Chebycev	5	0.179
5	Manhattan	2	0.148
6	Manhattan	3	0.253
7	Manhattan	4	0.123
8	Manhattan	5	0.156

After finding the optimal grouping method with the smallest DBI value, it is followed by grouping gold sales. The results of the grouping can be seen in Figure 11.

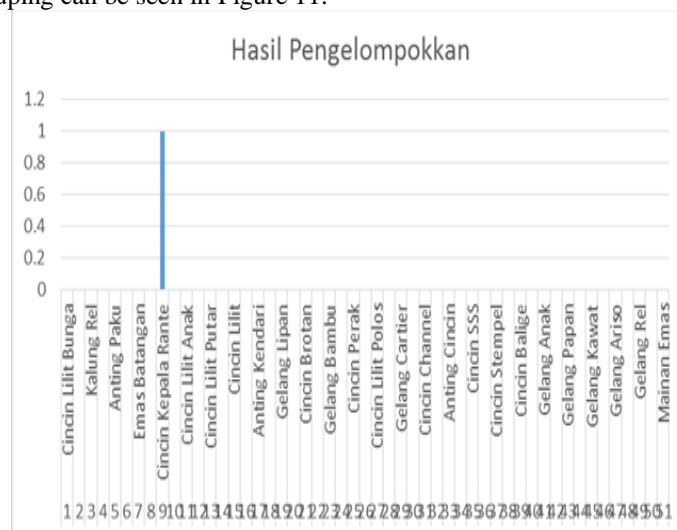


Figure 11. Grouping Results

4. Conclusion

The result of this research is a grouping of gold sales using the K-Medoids method with the calculation of the Chebycev distance and the Manhattan distance. The results of the comparison of DBI values from 2 predetermined distance calculations with several number of clusters found the most optimal grouping, namely the K-Medoids method with the calculation of the Chebycev distance distance with the number of clusters = 2. After finding the most optimal grouping, gold sales are grouped. The results of grouping gold sales obtained two number of clusters from the existing 51 items, namely the high cluster and the low cluster. Custer height is a ring of rante head with a size of 1 gram to 2 mayam. The low cluster is 50 items other than the high cluster. The DBI value of the optimal cluster is 0.024.

References

- [1] E. Schubert and PJ Rousseeuw, Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms, vol. 101, no. 124020371. 2021.
- [2] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementation of K-Medoids and K-Means Algorithms for Grouping Distributed Areas of Disability in Children," J. CoreIT J. Has. researcher. Computer Science. and Technol. Inf., vol. 4, no. 2, p. 64, 2018.
- [3] Y. Effective, P. Pt, and A. Mine, "No Title," vol. 4, no. 3, pp. 236–249, 2017.
- [4] RA Nisyah, N. Hidayat, and AA Supianto, "Forecasting Gold Jewellery Sales Results Using Exponential Smoothing Method (Case Study: Sumenep Gold Shop Rejeki Baru)," vol. 3, no. 7, pp. 7268–7274, 2019.
- [5] A. Lasisi, R. Ghazali, and T. Herawan, "Comparative Performance Analysis of Negative Selection Algorithm with Immune and Classification Algorithms," Adv. Intell. syst. Comput., vol. 287, pp. 441–452, 2014.
- [6] RD Ramadhani and DJ Ak, "K-Means and K-Medoids Evaluation on Small Datasets," Semin. Nas. information. and App., no. September, pp. 20–24, 2017.
- [7] D. Ayu, "Puskesmas Clustering in Kudus District Using the K-Means Method with a Comparison of Euclidean and Chebyshev Distances," vol. 5, pp. 787–798, 2022.
- [8] W. Gie and D. Jollyta, "Comparison of Euclidean and Manhattan for Cluster Optimization Using the Davies Bouldin Index: Covid-19 Status in Riau Region," Pros. Semin. Nas. Ris. Dan Inf. science. 2020, vol. 2, no. April, pp. 187–191, 2020.

