Jurnal Mantik, 5 (4) (2022) 2830-2838

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

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

COMPARISON OF EUCLIDEAN DISTANCE, CAMBERRA DISTANCE, AND CHEBYCHEV DISTANCE IN K-MEANS ALGORITHM BASED ON DBI EVALUATION

Deny Kurniawan[1], Dedi Triyanto[2], Mochamad Wahyudi[3]

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

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

Email: deni@bsi.ac.id[1], dedi@bsi.ac.id[2], wahyudi@bsi.ac.id[3]

ARTICLE INFO

Article history:
Received: December, 05 2022
Revised: January, 09 2022
Accepted: February, 30 2022

Keywords:
Clustering;
Comparison;
Cosmetics Sales;
Data mining;
K-Means.

ABSTRACT

During the COVID-19 pandemic, almost all businesses experienced difficulties. But not all businesses experience difficulties. Cosmetics is a product category that still exists during the pandemic. Many customers buy cosmetics through online sales. Devi Cosmetics is a trading business which is engaged in selling cosmetics. Due to the large number of sales transactions recorded in the neglected database, it is difficult for business managers to find out which cosmetic products are in high demand by customers and make it difficult for business managers to determine the inventory of cosmetic goods correctly. Determination of the incorrect supply of cosmetics resulted in the loss of the store manager, namely many customers who canceled buying cosmetics due to empty supplies. This study uses the K-Means algorithm to classify sales of cosmetic goods. To find out the best grouping results, it is necessary to compare several distance calculation methods. The distance calculation method here uses three methods, namely Euclidean Distance, Camberra Distance, and Chebychev Distance by finding the DBI value of the three methods. The smallest DBI value is the chebychev distance calculation method with a DBI value = 0.254.

Copyright © 2022 Mantik Journal.
All rights reserved.

1. Introduction

During the COVID-19 pandemic, it had a huge impact on cosmetic sales in Indonesia. Since it was mandatory for people to wear masks and stay at home, it has influenced the shift in people's use of cosmetics. People move from decorative make-up to scincare products.

Marketing is important in product sales[1]. The pattern of selling cosmetics has shifted from offline to online sales. The number of transactions that occur makes cosmetic sales data more and more neglected in the database. Store managers have difficulty determining products that are in great demand by customers and difficulty in determining the inventory of cosmetic goods. There are often vacancies in the supply of cosmetic goods caused by the determination of the inventory of goods that are not right.

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[2].

K-means is a non-hierarchical data grouping method that tries to partition the existing data into two or more groups. This method partitions data into groups so that data with the same characteristics are included in the same group and data with different characteristics are grouped into other groups [3]. The purpose of grouping this data is to minimize the objective function set in the grouping process, which generally tries to minimize variation within a group and maximize variation between groups[4][5][6][7].

This study focuses on the grouping of cosmetic sales at Devi Cosmetics stores using the K-Means method with the calculation of the Euclidian distance. The results of the grouping of cosmetic sales are compared

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

with the calculation of the camberra distance and chebychev distance to produce a more optimal grouping based on the DBI value. A good clustering result is the one with the smallest DBI value. The best clustering results are used as recommendations to the Devi Cosmetics store manager to determine which products are in high demand and determine inventory.

2. Research Methods

2.1 Research data

The research data was taken from Devi Cosmetics store in the form of cosmetic sales data for 2022 from January to April 2022. The data consists of 56 cosmetic product items.

TABLE 1
COSMETIC PRODUCT SALES

No	Goods	January	February	March	April
1	Rose Water	4	7	11	1
2	Water Sofles	13	12	15	8
3	Scarlett Body Shower	2	7	10	3
4	Eyelashes	10	14	6	2
5	woe	34	6	6	11
6	Image	26	7	8	2
7	eyeliner	8	3	12	12
8	Henna	2	9	3	10
9	Pause	36	11	9	13
10	JJ Glow	0	4	3	3
11	Larist Cotton	7	4	12	0
12	Fashionable Cotton	6	5	12	1
13	Sariayu Cotton	7	2	0	3
14	Cotton Selection	2	9	8	7
15	Chinese Rubber	7	1	6	6
16	False Nails	0	5	11	2
17	nail polish	4	7	0	1
18	Lameila	1	1	2	1
19	Lavea Creambath	1	5	2	1
20	Eyelash Glue	3	8	4	0
21	Hanasui Lip Cream	18	15	20	7
22	Implora Lip Cream	21	23	23	8
23	Lip cream OMG	0	5	12	21
24	Pink Flash Lipstick	4	0	7	1
25	Maybelline Lipstick	1	9	2	3
26	Scarlett Lotion	23	16	16	10
27	Mabello Scrub	11	4	2	1
28	Marks	4	7	4	2
29	Marina	12	3	6	4
30	Mascara	8	11	16	8
31	Organic Mask	10	13	9	5
32	Saffron Mask	2	8	10	4
33	Veze Mask	5	4	4	2
34	Candlenut Oil	2	1	1	1
35	Ms Glow	20	18	25	9
36	Rexona	9	4	4	3
37	Arabic soap	5	11	8	4
38	Rice Soap	9	4	5	8
39	Hot Soap	4	7	5	4
40	Honey Soap	4	5	10	3
41	Thai soap	5	8	1	3
42	Scarlett	13	15	10	4

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

No	Goods	January	February	March	April
43	Emina Serum	2	0	2	2
44	Hanasui Serum	16	8	15	6
45	Implora Serum	6	3	9	2
46	Serum Red Jelly	15	12	4	12
47	Wardah Serum	3	3	10	2
48	Sheet Mask	11	18	6	12
49	Softlens	12	16	15	6
50	Sponds	10	10	4	5
51	Sunscreen Wardah	6	12	16	6
52	Sunisa	1	6	0	
53	Emina sunscreen	11	11	2	14
54	Tabitha	9	3	5	2
55	Sasha's Hair Vitamins	11	22	2	11
56	Viva Face Tonic	7	6	3	3

2.2 Research Framework

The steps in this research can be seen in Figure 1.

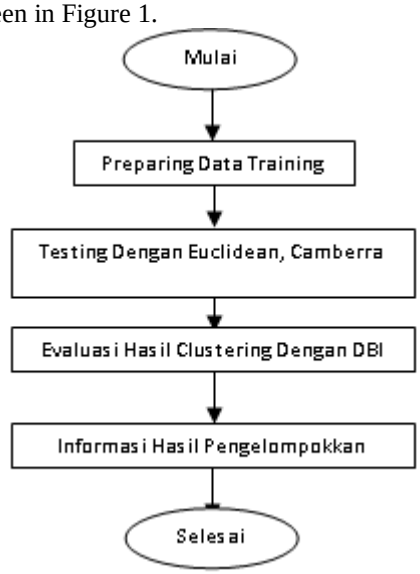

Figure 1. Research Framework

2.3 Euclidean Distance

Euclidean distance is one of the distance calculation methods used to measure the distance from 2 (two) points in Euclidean space (covering two-dimensional, three-dimensional, or even more Euclidean fields) [8][9].

2.4 Camberra Distance

For every 2 vector values to be matched, Canberra Distance divides the absolute difference between 2 values by the sum of the absolute 2 values [3]. The results of the two matched values are then added together to get Canberra Distance[10][9].

2.5 Chebycev Distance

Chebycev distance is a metric in which the distance between two vectors is the maximum difference between their axes[11][12].

3. Results and Discussion

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

3.1 K-Means Algorithm Calculation With Euclidian Distance

Before calculating using the K-Means method with the calculation of the Euclidean distance, the data is normalized first.

TABLE 2
DATA NORMALIZATION

No	Goods	January	February	March	April
1	Rose Water	0.1111	0.1944	0.3056	0.0278
2	Water Sofles	0.3611	0.3333	0.4167	0.2222
3	Scarlett Body Shower	0.0556	0.1944	0.2778	0.0833
4	Eyelashes	0.2778	0.3889	0.1667	0.0556
5	woe	0.9444	0.1667	0.1667	0.3056
6	Image	0.7222	0.1944	0.2222	0.0556
7	eyeliner	0.2222	0.0833	0.3333	0.3333
8	Henna	0.0556	0.2500	0.0833	0.2778
9	Pause	1.0000	0.3056	0.2500	0.3611
10	JJ Glow	0.0000	0.1111	0.0833	0.0833
11	Larist Cotton	0.1944	0.1111	0.3333	0.0000
12	Fashionable Cotton	0.1667	0.1389	0.3333	0.0278
13	Sariayu Cotton	0.1944	0.0556	0.0000	0.0833
14	Cotton Selection	0.0556	0.2500	0.2222	0.1944
15	Chinese Rubber	0.1944	0.0278	0.1667	0.1667
16	False Nails	0.0000	0.1389	0.3056	0.0556
17	nail polish	0.1111	0.1944	0.0000	0.0278
18	Lameila	0.0278	0.0278	0.0556	0.0278
19	Lavea Creambath	0.0278	0.1389	0.0556	0.0278
20	Eyelash Glue	0.0833	0.2222	0.1111	0.0000
21	Hanasui Lip Cream	0.5000	0.4167	0.5556	0.1944
22	Implora Lip Cream	0.5833	0.6389	0.6389	0.2222
23	Lip cream OMG	0.0000	0.1389	0.3333	0.5833
24	Pink Flash Lipstick	0.1111	0.0000	0.1944	0.0278
25	Maybelline Lipstick	0.0278	0.2500	0.0556	0.0833
26	Scarlett Lotion	0.6389	0.4444	0.4444	0.2778
27	Mabello Scrub	0.3056	0.1111	0.0556	0.0278
28	Marks	0.1111	0.1944	0.1111	0.0556
29	Marina	0.3333	0.0833	0.1667	0.1111
30	Mascara	0.2222	0.3056	0.4444	0.2222
31	Organic Mask	0.2778	0.3611	0.2500	0.1389
32	Saffron Mask	0.0556	0.2222	0.2778	0.1111
33	Veze Mask	0.1389	0.1111	0.1111	0.0556
34	Candlenut Oil	0.0556	0.0278	0.0278	0.0278
35	Ms Glow	0.5556	0.5000	0.6944	0.2500
36	Rexona	0.2500	0.1111	0.1111	0.0833
37	Arabic soap	0.1389	0.3056	0.2222	0.1111
38	Rice Soap	0.2500	0.1111	0.1389	0.2222
39	Hot Soap	0.1111	0.1944	0.1389	0.1111
40	Honey Soap	0.1111	0.1389	0.2778	0.0833
41	Thai soap	0.1389	0.2222	0.0278	0.0833
42	Scarlett	0.3611	0.4167	0.2778	0.1111

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

No	Goods	January	February	March	April
43	Emina Serum	0.0556	0.0000	0.0556	0.0556
44	Hanasui Serum	0.4444	0.2222	0.4167	0.1667
45	Implora Serum	0.1667	0.0833	0.2500	0.0556
46	Serum Red Jelly	0.4167	0.3333	0.1111	0.3333
47	Wardah Serum	0.0833	0.0833	0.2778	0.0556
48	Sheet Mask	0.3056	0.5000	0.1667	0.3333
49	Softlens	0.3333	0.4444	0.4167	0.1667
50	Sponds	0.2778	0.2778	0.1111	0.1389
51	Sunscreen Wardah	0.1667	0.3333	0.4444	0.1667
52	Sunisa	0.0278	0.1667	0.0000	0.0000
53	Emina sunscreen	0.3056	0.3056	0.0556	0.3889
54	Tabitha	0.2500	0.0833	0.1389	0.0556
55	Sasha's Hair Vitamins	0.3056	0.6111	0.0556	0.3056
56	Viva Face Tonic	0.1944	0.1667	0.0833	0.0833

After the data is normalized, it is continued with the grouping process using the K-Means algorithm.
1. Determination of the Centroid value:

TABLE 4
DATA CENTROID

Centroid	January	February	March	April
C1	1.0000	0.3056	0.2500	0.3611
C2	0.5000	0.4167	0.5556	0.1944
C3	0.0000	0.1111	0.0833	0.0833

2. Calculating the distance from Centroid
Calculation of distance using Euclidian Distance iteration-1 with the formula:

$$D_{(i,f)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2} \quad (1)$$

The calculation is continued until Dx34,c3.

TABLE 5
CENTROID DISTANCE TO ITERATION 1

Cluster 1	Cluster 2	Cluster 3	Distance shortest
0.9574	0.5393	0.2679	0.2679
0.6753	0.2152	0.5569	0.2152
0.9911	0.5800	0.2187	0.2187
0.7930	0.4698	0.4025	0.4025
0.1800	0.6509	0.9754	0.1800
0.4285	0.4787	0.7407	0.4285
0.8137	0.5069	0.4185	0.4185
0.9643	0.6747	0.2453	0.2453
0.0000	0.6193	1.0690	0.0000
1.0690	0.7607	0.0000	0.0000
0.9078	0.5234	0.3275	0.3275
0.9167	0.5152	0.3068	0.3068
0.9225	0.7381	0.2187	0.2187
0.9610	0.5800	0.2324	0.2324
0.8780	0.6298	0.2422	0.2422
1.0603	0.6395	0.2257	0.2257
0.9880	0.7328	0.1712	0.1712
1.0823	0.8075	0.1076	0.1076
1.0592	0.7602	0.0735	0.0735
0.9985	0.6684	0.1643	0.1643
0.6193	0.0000	0.7607	0.0000
0.6747	0.2531	0.9730	0.2531
1.0412	0.7265	0.5597	0.5597
0.9988	0.6950	0.103	0.103

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Cluster 1	Cluster 2	Cluster 3	Distance shortest
1.0312	0.7163	0.1443	0.1443
0.4410	0.1984	0.8292	0.1984
0.8179	0.6395	0.3118	0.3118
0.9566	0.6461	0.1443	0.1443
0.7505	0.5450	0.3458	0.3458
0.8137	0.3203	0.4867	0.3203
0.7577	0.3859	0.4129	0.3859
0.9809	0.5652	0.2324	0.2324
0.9444	0.6638	0.1443	0.1443
1.0628	0.8094	0.1273	0.1273
0.6672	0.1800	0.9280	0.1800
0.8347	0.6048	0.2515	0.2515
0.8971	0.5107	0.2778	0.2778
0.7949	0.5747	0.2913	0.2913
0.9367	0.6174	0.1521	0.1521
0.9465	0.5638	0.2257	0.2257
0.9354	0.6776	0.1863	0.1863
0.6956	0.3216	0.5122	0.3216
1.0567	0.8003	0.1303	0.1303
0.6174	0.2469	0.5727	0.2469
0.9150	0.5787	0.2390	0.2390
0.6009	0.4803	0.5350	0.4803
0.9919	0.6174	0.2152	0.2152
0.7265	0.4640	0.5604	0.4640
0.7275	0.2205	0.5833	0.2205
0.7688	0.5189	0.3298	0.3298
0.8780	0.3622	0.4631	0.3622
1.0758	0.7949	0.1332	0.1332
0.7217	0.5813	0.4747	0.4747
0.8471	0.6054	0.2591	0.2591
0.7852	0.5813	0.6273	0.5813

3.2 Cluster Evaluation Based on DBI Value

After obtaining the results of the clustering using the K-Means method, then an evaluation of the clustering of the three distance calculations is carried out using the DBI value. The formula for finding the DBI value is as follows:

1. Finding SSW with the formula:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (2)$$

Information :

m_i= The number of data in the i-th cluster

X = Data in cluster

D(x,c) = Distance data to centroid

X_j= Data on the cluster

C_i= Centroid cluster i

2. Find SSB with the formula:

$$SSB_{ij} = d(c_i\ c_j) \quad (3)$$

Information :

c_i= Cluster 1

c_j= Other Clusters

$d(c_i\ c_j)$= Distance from centroid sat to other

3. Find the Ratio with the formula:

$$R_{ij} = \frac{SSw_i + SSW_j}{SSB_{ij}} \quad (4)$$

Information :

R_{ij} = Ratio between clusters

SSW_i= Cluster 1

SSW_j = Cluster 2

SSB_{ij}= Separation of clusters 1 and 2

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

4. Finding DBI with the formula:

DBI = $()\frac{1}{K}\sum_{i=1}^{k} max_{i\neq j}\, R_{ij}$ (5)

Information :

K = existing cluster

$R_{i,j}$= Ratio between clusters i and j

Max = Find the largest inter-cluster ratio

3.3 DBI Value of K-Means Clustering With Euclidean Distance

The following is the DBI value of K-Means Clustering with the calculation of the Euclidian distance.

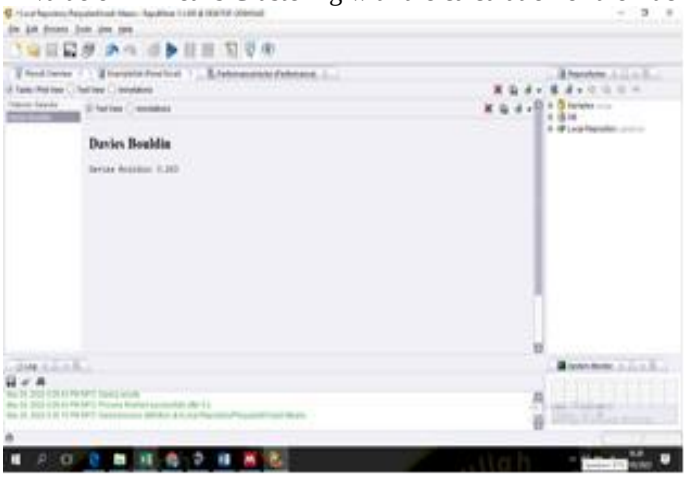

Figure 2. DBI value of Euclidian Distance

3.4 DBI Value of K-Means Clustering With Camberra Distance

The following is the DBI value of K-Means Clustering by calculating the camberra distance.

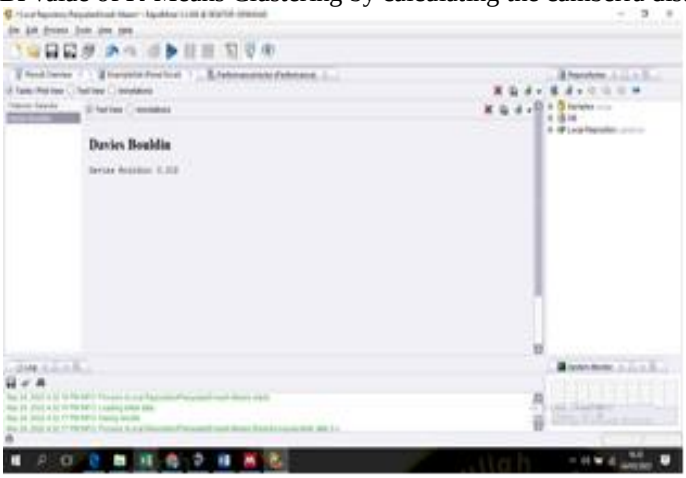

Figure 3. DBI value of Camberra Distance

3.5 DBI Value of K-Means Clustering With Chebycev Distance

The following is the DBI value of K-Means Clustering with the calculation of the Chebychev distance

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

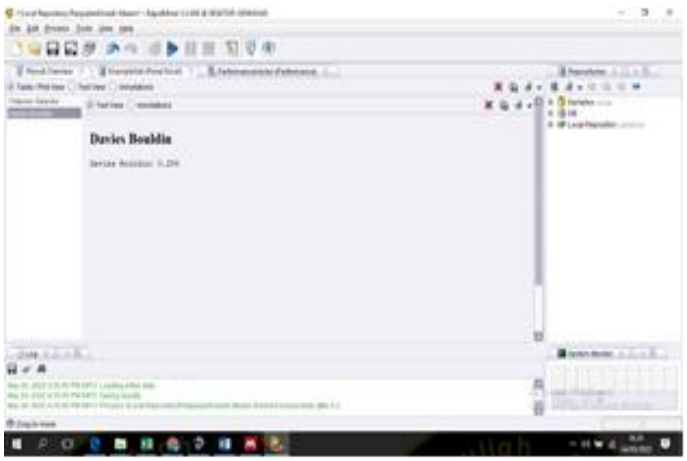

Figure 4. DBI value from Chebycev Distance

3.6 Comparison of DBI Values of 3 Distance Calculations

From the test results using a rapidminer with 3 distance calculations, it can be presented the comparison results in the following comparison table.

TABLE 4
CALCULATION 3 DISTANCE CALCULATION

No	Distance Calculation	Total k	DBI Value
1	Euclidian Distance	2	0.263
2	Euclidian Distance	3	0.304
3	Euclidian Distance	4	0.274
4	Camberra Distance	2	0.312
5	Camberra Distance	3	0.409
6	Camberra Distance	4	0.399
7	Chebydev Distance	2	0.254
8	Chebydev Distance	3	0.301
9	Chebydev Distance	4	0.301

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] K. Di and K. Blitar, “PROMOTION EFFECTIVENESS IN INCREASING COSMETIC PRODUCT SALES IN BLITAR CITY Denok Wahyudi Setyo Rahayu,” vol. 12, 2019.
[2] M. Robani and A. Widodo, "K-Means Clustering Algorithm for Grouping Al-Quran Verses in Indonesian Translation," J. Sist. inf. Business, vol. 6, no. 2, p. 164, 2016.
[3] H. Priyatman, F. Sajid, and D. Haldivany, “Clustering Using K-Means Clustering Algorithm to Predict Graduation Time,” vol. 5, no. 1, pp. 62–66, 2019.
[4] E. Nanda, Solikun, and Irawan, “APPLICATION OF DATA MINING IN GROUPING CORN PRODUCTION BY PROVINCE USING K-MEANS ALGORITHM,” vol. 3, pp. 702–709, 2019.
[5] DPT Hapsari and E. Widodo, "Clustering of Crime Prone Areas in Indonesia Using K-Means Clustering Analysis," Pros. SI MaNIs (Nas. Integr. Mat. and Islamic Values Seminar., vol. 1, no. 1, pp. 147–153, 2017.
[6] NH Kristanto, ACL A, and HB S, “Implementation of K-Means Clustering for Grouping Profitability Ratio Analysis in Working Capital,” Juisi, vol. 02, no. 01, pp. 9–15, 2016.
[7] E. Muningsih, I. Maryani, and VR Handayani, "Application of the K-Means Method and Optimization of the

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Number of Clusters with the Davies Bouldin Index for Clustering Provinces Based on Village Potential," J. Sains and Manaj., vol. 9, no. 1, pp. 95–100, 2021.
[8] AK Clustering, "Comparison of Accuracy of Euclidean Distance, Minkowski Distance, and Manhattan Distance on Chi-Square-based K-Means Clustering Algorithm," no. July, pp. 19–24, 2019.
[9] RI Fajriah, H. Sutisna, BK Simpony, BS Informatics, and U. Bsi, "Comparison of Manhattan and Euclidean Distance Space in K - Means Clustering in Determining Promotion," vol. 4, no. 1, pp. 36–49, 2019.
[10] SR Wurdianarto, S. Novianto, and U. Rosyidah, "COMPARATION OF EUCLIDEAN DISTANCE WITH CANBERRA DISTANCE ON FACE RECOGNITION," vol. 13, no. 1, pp. 31–37, 2014.
[11] WMP Dhuhita, "CLUSTERING USING THE K-MEANS METHOD FOR," vol. 15, no. 2, 2015.
[12] T. Kl, "Lower bounds on the size of spheres of permutations under the Chebychev distance," no. 123, 2010.

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

JournalMantikis licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).