Grouping of Village Status in West Java Province Using the Manhattan, Euclidean and Chebyshev Methods on the K-Mean Algorithm

Gatot Tri Pranoto [1*)], Wahyu Hadikristanto [2], Yoga Religia [3]
[1]Informatics Engineering Study Program, Faculty of Engineering, Pelita Bangsa University
[2]Informatics Engineering Study Program, Faculty of Engineering, Pelita Bangsa University
[3]Informatics Engineering Study Program, Faculty of Engineering, Pelita Bangsa University
email: [1]gatot.pranoto@pelitabangsa.co.id, [3]wahyu.hadikristanto@pelitabangsa.ac.id, [2]yoga.religia@pelitabangsa.ac.id

Abstract – Village Potential Data in 2014 (Podes 2014) in West Java Province is data released by the Central Statistics Agency in collaboration with the Ministry of Villages PDTT in unsupervised form and consists of 5319 village data. The 2014 Podes data in West Java Province is based on the level of village development (village specific) in Indonesia, by using the village as the unit of analysis. However, village funds have not been distributed effectively and accurately in accordance with the conditions and potential of the village due to the lack of clear information about the status of the village. So that there are no priority villages that should receive greater funds and attention from the government. One of the algorithms that can be used for the clustering process is the k-means algorithm. Grouping data using k-means is done by calculating the shortest distance from a data point to a centroid. In this study, a comparison of the distance calculation methods on k-means between Manhattan, Euclidean and Chebyshev will be carried out. Tests will be performed using execution time and davies bouldin index. From this test, the 2014 Village Potential data in West Java Province has been grouped into 5 village statuses by obtaining the number of villages for each cluster , namely cluster as many as 694 villages, cluster as many as 567 villages, cluster as many as 1440 villages. , cluster is 1557 villages and cluster is 1061 villages. For distance calculation, Chebyshev has the most efficient accumulation of time compared to Manhattan and Euclidean. Meanwhile, the Euclidean method has the Davies Index compared to the Manhattan and Chebyshev methods.

Keywords – Village Development, k-means, *Manhattan, Euclidean, Chebyshev, Davies Bouldin index*.

I. INTRODUCTION

The Republic of Indonesia is one of the countries in Southeast Asia which is located between two continents (Asia and Australia) and two oceans (India and the Pacific) which makes Indonesia also known as the Archipelago (Intermediate Archipelago). Indonesia is the largest archipelagic country in the world consisting of 17,504 islands. With a population of 222 million in 2006, Indonesia is the country with the fourth largest population in the world. Indonesia consists of various ethnic groups, languages and different religions. The total area of Indonesia is 1,913,578 km2, and has the second largest biodiversity in the world [10].

The large number of people and the rapid development in cities have an impact on the Indonesian economy. So that the development makes people come to the city to get a job and settle down. This phenomenon is also known as urbanization. Urbanization in Indonesia causes many problems, the impact on the village is the reduction in human resources because the population moves to cities which causes villages in Indonesia to not experience real development. The existence of urbanization is triggered by development facilities between rural and urban areas [14].

The Ministry of Villages, Development of Disadvantaged Regions and Transmigration of the Republic of Indonesia (Kemendes) is a ministry within the Government of Indonesia in charge of developing rural areas and rural areas, empowering rural communities, accelerating development of disadvantaged areas, and transmigration. The Ministry of Villages is under and responsible to the President. Based on Presidential Regulation Number 18 of 2020 concerning the 2020-2024 National Medium-Term Development Plan (RPJMN), the 2020-2024 RPJMN is a strategic document that contains development plans that must be carried out by the government for the next five years. The 2020-2024 RPJMN is used as an official reference for local governments and other stakeholders in implementing development. The RPJMN 2020-2024 was raised with the aim of mapping rural conditions in Indonesia based on their level of development, setting development targets/targets in the next five years that must be achieved by village development actors, and photographing the performance of village development that has been carried out. In the 2020-2024 RPJMN for the area and spatial planning of the sub-sector of village development and rural areas, it contains village development targets that must be achieved in the next five years, namely reducing the number of underdeveloped villages to 5,000 villages and increasing the number of independent villages to at least 2,000 villages in 2024. Realizing this requires mapping the status of the village to be built.

The Ministry of Village based on the Regulation of the Minister of Villages, Development of Disadvantaged Regions, and Transmigration of the Republic of Indonesia number 2 of 2016 concerning the index of developing villages, states that the status of villages is categorized into 5 categories, namely independent villages, developed villages, developing villages, underdeveloped villages and

very underdeveloped villages with the explanation as follows: following:

1) Independent Village or the so-called Sembada Village is an Advanced Village that has the ability to carry out village development to improve the quality of life and life as much as possible for the welfare of the Village community with social resilience, economic resilience, and ecological resilience in a sustainable manner. Independent Village or Madya Village is a Village that has a Developing Village Index greater (>) than 0.8155.

2) Advanced Village or the so-called Pre-Sufficient Village is a Village that has the potential of social, economic and ecological resources, as well as the ability to manage them to improve the welfare of the Village community, quality of human life, and alleviating poverty. Advanced Village or Pre-Madya Village is a Village that has a Developing Village Index less and equal to (≤) 0.8155 and greater (>) than 0.7072.

3) Developing Village or what is called Madya Village is a potential village to become an advanced village, which has the potential of social, economic and ecological resources but has not managed them optimally for improving the welfare of the village community, quality of human life and alleviating poverty. Developing Village or Middle Village is a Village that has a Developing Village Index less and equal to (≤) 0.7072 and greater (>) than 0.5989.

4) Disadvantaged Villages or so-called Pre-Madya Villages are villages that have potential social, economic, and ecological resources but have not, or have not managed them in an effort to improve the welfare of the Village community, the quality of human life and experience poverty in its various forms. Disadvantaged Villages or Pre-Madya Villages are Villages that have a Developing Village Index less and equal to (≤) 0.5989 and greater (>) than 0.4907.

5) Very Disadvantaged Villages or so-called Pratama Villages are villages that experience vulnerability due to natural disasters, economic shocks, and social conflicts so that they are not able to manage the potential of social, economic and ecological resources, and experience poverty in various forms. Very Disadvantaged Villages or Pratama Villages are Villages that have a Developing Village Index less and smaller (≤) than 0.4907.

Village status is in fact inseparable from village development that uses funds from the government. In the Law of the Republic of Indonesia number 6 of 2015 concerning villages, article 86 paragraph 2 states that the Government and Regional Governments are obliged to develop a village information system and development of rural areas. The Ministry of Villages in collaboration with the National Development Planning Agency and the Central Statistics Agency issued data on village potential (Podes) in West Java Province which has 5,319 instances and consists of 42 dependent attributes without labels. Podes data is a measurement method that is compiled based on the level of village development in Indonesia by making the village the unit of analysis. Podes data is compiled with reference to Law Number 6 of 2014 concerning Villages,

which is intended to capture the level of village development in West Java Province and can be used as a reference for policy planning and village development supervision [19].

In information technology, data is an important part that cannot be separated from information retrieval. Data mining is a series of activities used to find new, hidden or unexpected patterns contained in the data. The term data mining is often considered as a synonym for knowledge discovery from data (KDD), namely the discovery of knowledge from data that focuses on the purpose of the mining process [21]. Data mining can be used to perform clustering, classification and association. Clustering is a grouping process that is carried out by finding similar characteristics between data according to certain class groups [8]. In simple terms, clustering can be used to analyze a set of data and generate a set of clustering rules that can be used to group future data.

In the real world sometimes data is not only grouped into status binary (binary class),but needs also to be grouped into the multi-status(multi-class).In the case of multi-class data-sets, grouping will be more difficult than in the case of classes binary. K-mean is an interactive clustering algorithm that partitions the data-set into number of K clusters a predetermined. Ever done a comparative study of clustering, the partition-based clustering hierarchical based and clustering. density-basedIn this study, it was revealed that k-mean which is a partition-based algorithm provides better performance and k-mean is superior for large/lots of data compared to clustering hierarchical and-based density [6]. In addition, several other studies also mention that clustering using the k-means algorithm is faster than algorithms clustering otherand also produces quality clusters when using large datasets [7] [20] [12].

The k-mean algorithm has been used for multi-class grouping which shows effective results and is able to improve performance classification [1]. In another case, two algorithms were compared, namely k-mean and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) which said that the two algorithms are algorithms clustering the most commonly used for grouping data with different criteria, and the test results show that k-mean algorithm is better than DBSCAN in time efficiency analysis [5]. In addition, in research on the comparison of algorithms clustering data mining comparing k-mean, Density based clustering and Hierarchical Clusterings concluded that the k-mean algorithm is superior in time efficiency compared to the other two algorithms and also k-mean is able to distribute cluster instances quite accurately [17]. When the k-mean algorithm performs data grouping, the k-mean algorithm will calculate the closest distance between a data set to appoint centroid. The calculation of the distance on the k-mean algorithm can use Manhattan, Euclidean and Chebyshev. Each method of calculating distance is superior to one another depending on the dataset used [3] [18] [15].

The calculation of the distance in the k-means algorithm can use Manhattan, Euclidean and Chebyshev. Awasthi in his research on the comparison of the Manhattan and

Euclidean distance calculation methods on the k-means algorithm to determine the number of squared errors, the data used in this study was a data-set bank that was tested using the WEKA tool. The test results show that the Manhattan distance calculation method is better than the Euclidean method [3].

In another study, a comparison of 3 methods of calculating distances in the k-means algorithm was carried out, namely Manhattan, Euclidean and Minkowski to find the best distance calculation method, the study was carried out by comparing the results of previous studies which concluded that the Euclidean distance calculation method was better than the Manhattan and Minkowski method [18].

Kouser and Sunita also conducted other research on the comparison of the Manhattan, Euclidean and Chebyshev distance calculation methods on the k-means algorithm to determine the accuracy and mean absolute error. From the tests conducted using the flower data-set, it was found that the Chebyshev distance calculation method is better than the Manhattan and Euclidean methods [15]. From previous studies, it is known that the Manhattan, Euclidean and Chebyshev distance calculation methods are superior to each other depending on the dataset used.

Based on the considerations previously mentioned in this study,will be carried out clustering village status in West Java Provinceusing the k-mean algorithm into 5 village statuses, as well as comparing which distance calculation method is the most effective for grouping village status.

II. RESEARCH METHODOLOGY

A. Research Methods

This research is a quantitative method where there is data on Village Potential in 2014 (Podes 2014) in West Java Province which will be processed. In this study,will be carried out clustering village status in West Java Province, using the k-mean algorithm with the Manhattan, Euclidean and Chebyshev distance calculation methods into 5 village statuses, as well as comparing which distance calculation method has the most efficient accumulation of time and which has the most efficient time accumulation. the value Davies Index is most optimal.

B. Data Collection Methods The data

Used in this study is the potential of villages in West Java Province in 2014 (Podes 2014). Podes 2014 data is secondary data issued by the Central Statistics Agency based on Law Number 6 of 2014 concerning villages. Data Podes 2014 consists of 5,319 instances and has 42 dependent attributes without labels. The description of the 2014 Podes data owned can be seen in Table 1.

Table 1. Image data sets Podes 2014

No	Nama Desa	I1	I2	I3	I4	I5	I6	I7	I8	.	.	I40	I41	I42
1	WANAHERANG	3	3	4	4	3	3	4	4	.	.	5	5	5
2	BOJONG KULUR	3	3	4	4	3	3	4	4	.	.	3	4	2
3	CIANGSANA	3	3	4	4	3	3	4	4	.	.	3	4	5
4	GUNUNG PUTRI	3	3	4	3	3	3	3	4	.	.	3	4	3
5	BOJONG NANGKA	4	3	4	4	3	3	4	4	.	.	3	4	3
6	TLAJUNG UDIK	3	3	4	4	3	3	3	4	.	.	3	4	3
7	CICADAS	3	3	4	4	3	3	3	4	.	.	3	5	3
8	CIKEAS UDIK	3	3	4	4	3	3	3	4	.	.	3	5	2
..	....	.	.	.	.	.	.	.	.	.	.	..	..	..
..	....	.	.	.	.	.	.	.	.	.	.	..	..	..
5318	WARINGINSARI	3	4	4	4	3	3	4	4	.	.	4	4	3
5319	KUJANGSARI	3	4	4	4	3	3	3	3	.	.	4	4	3

In Table 1 there are attributes I1, I2, I3 to I42 where the initial I stands for "Indicator". The value of each attribute is 0 to 5, where the value 0 is the lowest value while the value 5 is the highest value. Information on values 0-5 for each indicator can be seen in Appendix 1.

Data on the potential of villages in West Java Province in 2014 is the result of measurements made based on the level of village development (village specific) in West Java Province by using the village as the unit of analysis. Podes 2014 in West Java Province is used as a reference for the main indicators that make up the index, and data on government administration areas (MDNR Indonesia, 2015) which is used as a reference standard for the number of integrated villages in West Java Province. Podes 2014 in West Java Province is a complex multidimensional concept consisting of dimensions, variables and indicators that are used as measuring tools for village development. In the 2014 Podes data in West Java Province, there are 5 dimensions, 12 variables and 42 indicators. Completely from the dimensions, variables and indicators can be seen in Table 2.

Table 2. Table of Indicators for Compiling Podes 2014

Code	Description Indicator	Variable Indicator	Dimension Indicator
I1	Availability and access to TK/RA/BA	Educational	Services Basic services
I2	Availability and access to SD equivalent		
I3	Availability and access to SMP equal		
I4	Availability and access to SMA equal		
I5	Availability and easy access to hospital	Health Services	
I6	Availability and ease of access to maternity hospitals		
I7	Availability and ease of access to puskesmas		
I8	Availability and ease of access to polyclinics/medical centers		
I9	Availability and ease of access to doctors		

I10	Availability and ease of access to midwives' practices		
I11	Availability and convenience access to poskesdes or polindes		
I12	Availability and ease of access to pharmacies		
I13	Availability of shops, minimarkets, or grocery stores	Economic	Infrastructure Condition of Infrastructure
I14	Availability of markets		
I15	Availability of restaurants, restaurants or food stalls/shops		
I16	Availability of hotel accommodation or lodging		
I17	Availability bank aan		
I18	Electrification	of Energy Infrastructure	
I19	Lighting conditions on main roads		
I20	Fuel for cooking		
I21	Sources of drinking water	Health and Sanitation Infrastructure	
I22	Sources of water for bathing/washing		
I23	Defecation facilities		
I24	Availability and quality of cellular communication facilities	Communication and Information Infrastructure	
I25	Availability of internet facilities and postal or goods delivery		
I26	Traffic and road quality	Means of Transportation	Accessibility / Transportation
I27	Road accessibility		
I28	Availability of public transport		
I29	transport operations		
I30	Travel time per kilometer of transportation to the sub-district office	Transportation accessibility	
I31	Cost per kilometer of transportation to the sub-district office		
I32	Travel time per kilometer of transportation to the regent		
I33	Cost per kilometer of transportation to the regent		
I34	Handling of extraordinary events	Public Health Public	Services
I35	Handling of malnutrition		

I36	Availability of sports facilities	Sports	
I37	Existence of k groups Sports		
I38	Completeness of village government	Independence	of Government Administration
I39	Village autonomy		
I40	assets/wealth		
I41	Quality of human resources of village head	Quality of Human Resources	
I42	Quality of human resources of village secretary		

C. Sample Selection Methods

Data samples random taken from the original data is data 2014 Village Potential in West Java province as many as 15 villages and 42 indicators initialized as I1 to I42 to be grouped using the k-means algorithm, the data obtained are shown in Table 3.

Table 3. Table of Indicators for Compiling Podes 2014

x	Nama Desa	I1	I2	I3	I4	I5	...	...	I41	I42
1	CARIU	3.00	3.00	4.00	4.00	3.00	...	...	4.00	3.00
2	JABON MEKAR	3.00	3.00	4.00	4.00	3.00	...	...	5.00	5.00
3	CIHIDEUNG HILIR	2.00	4.00	4.00	4.00	3.00	...	...	4.00	3.00
4	CISOLOK	2.00	3.00	4.00	4.00	3.00	...	...	4.00	3.00
5	KADUPANDAK	3.00	4.00	5.00	5.00	0.00	...	...	4.00	5.00
6	BANJARAN KULON	4.00	4.00	4.00	4.00	3.00	...	...	5.00	3.00
7	JAGABAYA	4.00	3.00	4.00	3.00	3.00	...	...	4.00	3.00
8	DEPOK	3.00	4.00	4.00	2.00	3.00	...	...	0.00	3.00
9	LINGGAJAYA	2.00	3.00	3.00	4.00	1.00	...	...	4.00	3.00
10	KRAMATMULYA	3.00	3.00	3.00	3.00	3.00	...	...	4.00	3.00
11	PABUARAN WETAN	3.00	3.00	4.00	3.00	3.00	...	...	4.00	5.00
12	KADIPATEN	3.00	4.00	4.00	3.00	3.00	...	...	4.00	0.00
13	JATIBARANG	4.00	4.00	5.00	4.00	3.00	...	...	4.00	5.00
14	KASOMALANG WETAN	4.00	4.00	5.00	5.00	3.00	...	...	4.00	5.00
15	DAWUAN TIMUR	4.00	3.00	2.00	3.00	3.00	...	...	4.00	3.00

From table 3 will be used for grouping into 5 clusters initialized as C1 SD C5. The grouping will be done using the k-means algorithm with three different distance calculation methods, namely Manhattan, Euclidean and Chebyshev.

D. Evaluation

In this study, the evaluation will be carried out by

grouping the village potential data. The grouping will use the k-mean algorithm where the distance calculation will use Manhattan, Euclidean and Chebyshev. The results of the grouping obtained are the grouping of village potential data into 5 clusters, namely cluster 0, cluster 1, cluster 2, cluster 3 and cluster 4. Until this stage it is not a known cluster which can be called a cluster independent village, developed village, developing village, underdeveloped villages and very underdeveloped villages.

In the village potential data, each attribute/indicator has a value of 0 to 5, where a value of 0 is the lowest value while a value of 5 is the highest value. From each cluster obtained has value centroid, where the centroid is the "midpoint" of the cluster. So to determine the status of the village, we can calculate the number of centroids for each cluster, which can be written with the equation:

$$Status\ desa = \sum CI_1, CI_2, ..._{...}, CI_{42}$$
.... 8

From equation 8, CI is the centroid of each indicator and each cluster has 42 indicators. Determination of village status will be sorted based on the sum of the values centroid of each indicator in each cluster, where the lowest sum value will be initialized as very underdeveloped village status and the highest sum value will be initialized as independent village status.

E.   Validation
In this study, validation will be carried out to test the distance calculation method on which k-mean algorithm is most effectively used for grouping village potential data. The test will be carried out using tools RapidMinerto obtain the accumulated time and the value of the Bouldin Index for each distance calculation method used. The best time efficiency is the one that has the minimum accumulated time. Meanwhile, by using the Davies Bouldin Index, a cluster will be considered to have scheme clustering an optimal which has a Davies Bouldin Index minimal.

III.   RESULTS AND DISCUSSION
A.   Testing the proposed
Method Testing the Manhattan, Euclidean and Chebyshev distance calculation methods on the k-means algorithm used to group the 2014 Village Potential data in West Java Province will be carried out using latest model clustering that is validated using execution time and Davies Bouldin Index.

1.   Manhattan Distance
From the use of the k-means algorithm with the Manhattan Calculation method distance to group the 2014 Podes data in West Java Province which amounted to 5319 villages, the number of villages from each obtained cluster was as follows:
·   Cluster 0: 1,005 villages
·   Cluster 1: 1,189 villages
·   Cluster 2: 1,084 villages
·   Cluster 3: 447 villages
·   Cluster 4: 1,594 villages
When viewed from the number of centroids calculated by the equation above and the number of villages in each cluster, the status of the village can be obtained from the k-

means grouping using the Manhattan Calculation method distance as shown in Table 4.

Table 4. Status and Number of Villages Using Manhattan

Cluster	Jumlah Centroid	Status Desa	Jumlah Desa
Cluster 0	3,12	Tertinggal	1.005 desa
Cluster 1	3,35	Maju	1.189 desa
Cluster 2	3,73	Mandiri	1.084 desa
Cluster 3	2,75	Sangat Tertinggal	447 desa
Cluster 4	3,34	Berkembang	1.594 desa

2.   Euclidean Distance
From the use of the k-means algorithm with the Euclidean Calculation method distance to group the 2014 Podes data in West Java Province which amounted to 5,319 villages, the number of villages from each obtained cluster was as follows:
·   Cluster 0: 1,061 villages
·   Cluster 1: 567 villages
·   Cluster 2: 1,557 villages
·   Cluster 3: 1,440 villages
·   Cluster 4: 694 villages
When viewed from the number of centroids calculated by the equation above and the number of villages in each cluster, the status of the village can be obtained from the k-means grouping using the Euclidean Calculation method distance as shown in Table 5.

Table 5. Status and Number of Villages Using Euclidean

Cluster	Jumlah Centroid	Status Desa	Jumlah Desa
Cluster 0	3,71	Mandiri	1.061 desa
Cluster 1	3,19	Tertinggal	567 desa
Cluster 2	3,42	Maju	1.557 desa
Cluster 3	3,24	Berkembang	1.440 desa
Cluster 4	2,87	Sangat Tertinggal	694 desa

3.   Chebyshev Distance
From the use of the k-means algorithm with the Chebyshev Calculation method distance to group the 2014 Podes data in West Java Province, amounting to 5319 villages, the number of villages from each obtained cluster is as follows:
·   Cluster 0: 1,366 villages
·   Cluster 1: 1,004 villages
·   Cluster 2: 306 villages
·   Cluster 3: 1,214 villages
·   Cluster 4: 1,429 villages
When viewed from the number of centroids calculated by the equation above and the number of villages in each cluster, the status of the village can be obtained from the k-means grouping using the Chebyshev Calculation method distance as shown in Table 6.

Table 6. Status and Number of Villages Using Chebyshev

Cluster	Jumlah Centroid	Status Desa	Jumlah Desa
Cluster 0	3,53	Mandiri	1.366 desa
Cluster 1	3,04	Sangat Tertinggal	1.004 desa
Cluster 2	3,42	Berkembang	306 desa
Cluster 3	3,45	Maju	1.214 desa
Cluster 4	3,23	Tertinggal	1.429 desa

B. Testing the proposed

Accumulation of time is carried out by executing 5 times for each distance calculation method used. The 5 executions will then be averaged to obtain the most efficient execution time for each distance calculation method. From the tests that have been carried out, it is obtained that the length of execution time is different, as for the length of execution time from the test of the Manhattan, Euclidean and Chebyshev distance calculation methods that have been carried out, it can be seen in Figure 1.
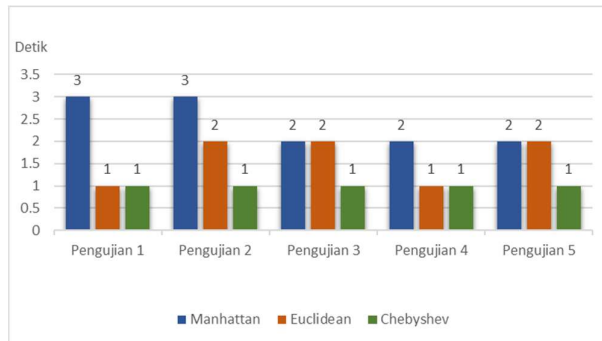

Figure 1. Execution Time

In Figure 1. It can be seen that the execution time of the method Manhattan distance to test 1 to test 5 in a row is 3 seconds, 3 seconds, 2 seconds, 2 seconds and 2 seconds, so when taken on average execution time of a Manhattan distance is 2.4 seconds. Meanwhile, the execution time of the Euclidean Method distance for testing 1 to 5, respectively, is 1 second, 2 seconds, 2 seconds, 1 second and 2 seconds, so that the average execution time of the Euclidean distance is 1.6 seconds. Then the execution time of the Chebyshev Method distance for testing 1 to 5, respectively, namely 1 second, 1 second, 1 second, 1 second and 1 second, so that when taken the average execution time of Chebyshev distance is 1 second. The more easily the execution time required for the Manhattan, Euclidean and Chebyshev methods can be seen in Table 7.

Table 7. Old Time Execution

Pengujian	Waktu Eksekusi		
	Manhattan	Euclidean	Chebyshev
1	3 detik	1 detik	1 detik
2	3 detik	2 detik	1 detik
3	2 detik	2 detik	1 detik
4	2 detik	1 detik	1 detik
5	2 detik	2 detik	1 detik
Rata-rata	**2,4 detik**	**1,6 detik**	**1 detik**

C. Testing the Davies Bouldin Index

In this study, the Davies Bouldin Index (DBI) was used to validate the data in each cluster. Measurement using DBI aims to maximize the distance inter-cluster. By using DBI A cluster will be considered to have scheme clustering an optimal if it has a Davies Index minimal. As for the tests that have been carried out, the values obtained Davies Index from the Manhattan, Euclidean and Chebyshev methods are shown in Figure 2.
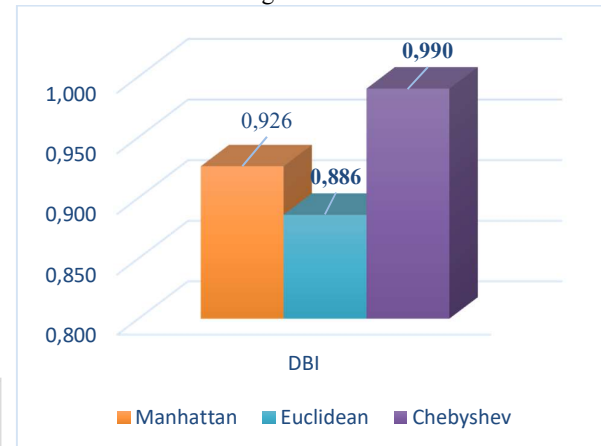

Figure 2. Davies Index of Manhattan, Euclidean and Chebyshev methods

From Figure 2, it can be seen that the value *Davies Index* from the Manhattan method is 0.926, the value *Davies Index* from the Euclidean method is 0.886 and the value *Davies Index* from the Chebyshev method is 0.990. As for the easier value of the *Davies Index* from the Manhattan, Euclidean and Chebyshev methods, it can be seen in Table 8.

Table 8. Status and Number of Villages Using Chebyshev

Davies Bouldin Index		
Manhattan	**Euclidean**	**Chebyshev**
0,926	0,886	0,990

From Table 8 it can be seen that the most optimal value of the Manhattan, Euclidean and Chebyshev methods is the Euclidean Method distance with the value Davies Index of 0.886.

D. Analysis of Test Results

From testing the 2014 Village Potential data grouping method in West Java Province using the k-means algorithm with the distance calculation methods Manhattan, Euclidean and Chebyshev that have been carried out, the results are:

1. The test model used can run well and shows the results in the form of values centroid for each cluster from the methods Manhattan, Euclidean and Chebyshev, so that the status of the village can be determined from the number of centroids in each cluster.
2. The use of the distance calculation method used affects the amount of data in each cluster.
3. The accumulated time obtained from the tests that have been carried out shows that the distance calculation method Chebyshev has the most efficient execution time with an average accumulation time of 1 second.

4. By using the test, the Davies Bouldin Index shows that the distance calculation method Euclidean has the Davies Index most optimal value with a value of 0.886.

From the tests that have been carried out, it can be seen that the 2014 Village Potential data grouping in West Java Province using the k-means algorithm with the distance calculation method Chebyshev has the most efficient accumulation of time compared to Manhattan and Euclidean, while the Euclidean method has the value Davies Index most optimal compared to the method. Manhattan and Chebyshev. So when viewed from the quality of the cluster based on the Davies Index , the obtained cluster status of the village is from the k-means algorithm with the distance calculation method Euclidean as follows:

- Cluster Very Disadvantaged Village As many as 694 villages
- Cluster Disadvantaged Village As many as 567 villages
- Cluster Developing Village Of 1,440 villages
- Cluster Advanced Village as many as 1,557 villages
- Cluster Independent Villageof 1,061 villages

IV. CONCLUSION

From the discussion and evaluation in the previous chapters, the 2014 Village Potential data grouping in West Java Province into 5 groups using the k-means algorithm with the Manhattan, Euclidean and Chebyshev distance calculation methods, the conclusions are:

1. The 2014 Village Potential data in West Java Province has been grouped into 5 village statuses with the obtained number of villages for each cluster , namely cluster as many as 694 villages, cluster as many as 567 villages, cluster as many as 1440 villages, cluster as many as 1557 villages and cluster as many as 1061 villages.

2. The grouping of Village Potential data in 2014 in West Java Province into 5 village statuses using the k-means algorithm with the Chebyshev distance calculation method has the most efficient accumulation of time compared to Manhattan and Euclidean, while the Euclidean method has the Davies Index most optimal

REFERENCES

[1] Al-roby, M. F., & El-halees, A. M. (2013). Classifying Multi-Class Imbalance Data. 37(5), 74–81.

[2] Amandeep Kaur Mann, N. K. (2013). Review Paper on Clustering Techniques. Global Journal of Computer Science and Technology.

[3] Awasthi, R., Tiwari, A. K., & Pathak, S. (2013). Empirical Evaluation On K Means Clustering With Effect Of Distance Functions For Bank Dataset. International Journal of Innovative Technology and Research, 1(3), 233–235.

[4] Mishra, B. K., Rath, A., Nayak, N. R., & Swain, S. (2012). Far efficient K-means clustering algorithm. ACM International Conference Proceeding Series. https://doi.org/10.1145/2345396.2345414

[5] Chakraborty, S., Nagwani, N. K., & Dey, L. (2011). Performance Comparison of Incremental K-means and Incremental DBSCAN Algorithms. International Journal of Computer Applications. https://doi.org/10.5120/3346-4611

[6] Chaudhari, B., & Parikh, M. (2012). A Comparative Study of Clustering Algorithms using Weka Tools. International Journal of Application or Innovation in Engineering and Management (IJAIEM).

[7] Claypo, N., & Jaiyen, S. (2015). Opinion mining for Thai restaurant reviews using K-Means clustering and MRF feature selection. 2015 7th International Conference on Knowledge and Smart Technology (KST), 105–108. https://doi.org/10.1109/KST.2015.7051469

[8] Deepa, V. K., Rexy, J., & Geetha, R. (2013). Rapid development of applications in data mining. 2013 International Conference on Green High Performance Computing, ICGHPC 2013. https://doi.org/10.1109/ICGHPC.2013.6533916

[9] Ding, S., Wu, F., Qian, J., Jia, H., & Jin, F. (2015). Research on data stream clustering algorithms. Artificial Intelligence Review. https://doi.org/10.1007/s10462-013-9398-7

[10] Direktorat Jenderal Pemerintahan Umum, K. D. N. (n.d.). https://www.bps.go.id/statictable/2014/09/05/1366/luas-daerah-dan-jumlah-pulau-menurut-provinsi-2002-2016.html.

[11] Gandhi, G., & Srivastava, R. (2014). Review Paper: A Comparative Study on Partitioning Techniques of Clustering Algorithms. International Journal of Computer Applications, 87(9), 10–13. https://doi.org/10.5120/15235-3770

[12] Ghosh, S., & Kumar, S. (2013). Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. International Journal of Advanced Computer Science and Applications. https://doi.org/10.14569/ijacsa.2013.040406

[13] Grabusts, P. (2011). The choice of metrics for clustering algorithms. Vide. Tehnologija. Resursi - Environment, Technology, Resources. https://doi.org/10.17770/etr2011vol2.973

[14] Harahap, F. R. (2013). Dampak Urbanisasi Bagi Perkembangan Kota Di Indonesia. Society, 1(1), 35–45. https://doi.org/10.33019/society.v1i1.40

[15] Kouser, K., & Sunita, S. (2013). A comparative study of K Means Algorithm by Different Distance Measures. International Journal of Innovative Research in Computer and Communication Engineering.

[16] KumarSagar, H., & Sharma, V. (2014). Error Evaluation on K-Means and Hierarchical Clustering with Effect of Distance Functions for Iris Dataset. International Journal of Computer Applications. https://doi.org/10.5120/15066-3429

[17] Pratap, S., Kushwah, S., Rawat, K., & Gupta, P. (2012). Analysis and Comparison of Efficient Techniques of Clustering Algorithms in Data Mining. 3, 109–113.

[18] Singh, A., Yadav, A., & Rana, A. (2013). K-means with Three different Distance Metrics. International Journal of Computer Applications. https://doi.org/10.5120/11430-6785

[19] Soleh, A. (2017). Strategi Pengembangan Potensi Desa. Jurnal Sungkai, 5(1), 35–52.

[20] Verma, M., Srivastava, M., Chack, N., Diswar, A. K., & Gupta, N. (2012). A Comparative Study of Various Clustering Algorithms in Data Mining. International Journal of Engineering Research and Applications Www.Ijera.Com.

[21] Xu, L., Jiang, C., Wang, J., Yuan, J., & Ren, Y. (2014). Information security in big data: Privacy and data mining. IEEE Access. https://doi.org/10.1109/ACCESS.2014.2362522