Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13148
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application of Data Mining for Clustering Human Development Index Based on West Java Province 2017-2022

Widiarina[1)*],Kartika Mariskhana [2)], Ita Dewi Sintawati [3)]
[1,2,3)] Universitas Bina Sarana Informatika,Indonesia, Indonesia
[1)], widiarina.wda@bsi.ac.id [2)] kartika.kma@bsi.ac.id, [3)] ita.ids@bsi.ac.id

Submitted : Nov 3, 2023 | **Accepted** : Nov 20, 2023 | **Published** : Jan 1, 2024

Abstract: Human development is used as a parameter to see development from the human side. The Human Development Index (HDI) explains how people get sufficient income, adequate health and education. Geographically, Indonesia is an archipelagic country where each province is spread across various islands separated by sea. Making the disparity in human development between provinces relatively high. The gap that occurs is still a problem that must be resolved immediately, because the gap in the human development index can hamper the government's goal of equalizing human welfare in Indonesia. −One of the problems related to population that West Java Province still has to face is the problem of imbalance in population distribution. Incomplete population distribution causes problems with population density and population pressure in an area. This research uses data sources from the West Java Province Central Statistics Agency (BPS). The data used in this research is data from 2017-2022 which consists of 27 regencies and cities of West Java Province. Therefore, researchers utilized the K-Means algorithm in clustering 27 Regencies and Cities of West Java Province. The data will be processed by clustering into 3 clusters, namely the high population area level cluster, the medium population area level cluster and the low population area level cluster. This research classifies population density using Ms. software. Excel and RapidMiner. The iteration process took place 3 times so that the results obtained were 8 regencies and cities with high population area clusters (C0), 0 regencies and cities with medium population area clusters (C1) and 19 regencies and cities with low population area clusters (C1). C2).
Keywords: Clustering, Human Development Index, Data Mining, K-Means

INTRODUCTION

The advancement of information and technology has impacted people's lives on a worldwide scale in the current globalization age. Therefore, it should come as no surprise that modern individuals use technology and communication tools on a regular basis(Anshori, 2018). People rely on information technology because it makes it simpler for them to get the information they need. Information is essential for conducting research, decision-making, business planning, and other activities. This has led to a lot of data as well as the public's thirst for knowledge because data is a collection of facts derived from occurrences that occur occasionally in people's daily lives(Daniati & Prasetyo, 2022). After that, the data is transformed into information with a pattern or shape and processed through a system that can generate reliable information that people can use to support tasks successfully and efficiently. This has led to a lot of data as well as the public's thirst for knowledge because data is a collection of facts derived from occurrences that occur occasionally in people's daily lives(Zamroni, 2020). After that, the data is transformed into information with a pattern or shape and processed

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13148 p-ISSN : 2541-044X

through a system that can generate reliable information that people can use to support tasks successfully and efficiently.

The Human Development Index (HDI) is a metric used to assess the performance of regional development; it considers factors such as adjusted income, knowledge and education, and life expectancy and health(Pamungkas Wijayanto, 2022). Geographically speaking, Indonesia is an archipelagic nation, with each province being dispersed over numerous islands that are divided by water. increasing the relatively large difference in human development between provinces. The government's objective of achieving equal human welfare throughout Indonesia may be hampered by the disparity in the human development index, so this work still needs to be done right away(Hamid, 2018). To take group-based solutions, it is necessary to group each province based on its human development index value. The Central Statistics Agency (BPS), a non-ministerial government agency directly answerable to the president for gathering data on various fields in Indonesia for government and public needs, is in charge of compiling and calculating data for the country's human development index (HDI). Both the Indonesian public data portal, http://data.go.id, and the official BPS website, http://bps.go.id, offer access to the data gathered from BPS. The high accuracy of the human development index (HDI) data gathered by BPS can be attributed to the fact that each province's data is gathered through direct field observations and expert statistical computations(Permadi, 2022). A comparative assessment of life expectancy, literacy, education, and living standards for every nation on the planet is provided by the Human Development Index (HDI). In addition to determining a nation's status as developed, developing, or underdeveloped, the HDI is used to assess how economic policies affect people's quality of life. Thus, the author's research focuses on Indonesia's Human Development Index (HDI) broken down by district. Naturally, there are a number of challenges involved in measuring and calculating Human Development Index data, including reading the data and determining which districts have the highest or lowest HDI values due to the data's impracticality. In addition, tabular data presents less informative information, which makes it challenging to read the data in particular due to the large volume of data. The absence of a clustering method from BPS for displaying HDI data by Indonesian district as graphs or diagrams is another barrier(Bariyah, 2022). Thus, in order to group HDI levels into image data in the form of graphs or diagrams, the author used the K-Means algorithm to segment or cluster the Human Development Index (HDI) per district in Indonesia for this final assignment research. and offer insightful justifications, as well as quickly determine the HDI level(Suhanda, 2020).

LITERATURE REVIEW

The following are some earlier research that the writers cite:

1. A study conducted by Hermawan and colleagues. It is possible to apply data mining with clustering techniques and the k-means algorithm to group Indonesian provinces according to the Human Development Index value. by designating as an attribute the mean HDI score for the years 2016–2021. The same research results were obtained whether the calculations were done manually or using RapidMiner. There were six clusters total, with two provinces in the cluster with a very good HDI value, three provinces in the cluster with a good HDI value, seventeen provinces in the cluster with a more than good HDI value, eight provinces with good HDI scores, three provinces with poor HDI scores, and one province with very poor HDI scores. Using the Davies-Bouldin Index (DBI) to test the outcomes of the number of clusters performance evaluation, the best value of -0.283 was found for the number of clusters 6(Hermawan & Hasugian, 2022).
2. The Algomerative method employing distance similarity with the Ward Method is the Hierarchical method utilized in the research of Emir Luthfi, et al. The validity of the Dunn Index (DN), Davies Bouldin Index (DB), and Calinski-Harabasz Index (CH) will be examined in order to compare the outcomes of the three methods and determine which is the most important and optimal number of clusters/groups. The standard deviation ratio value, which seeks to obtain the minimum standard deviation value within groups (SW) and the maximum standard deviation value between groups (SB), is the best algorithm method to compare and determine. Using K-Medoids yielded the best model, which is more clearly seen when standard deviation ratios are compared. This is then used in sentiment analysis for Indonesian district and city regions based on

*name of corresponding author

each region's HDI figures to determine which regions have the highest and lowest HDI figures in 2019(Luthfi & Wijayanto, 2021).

3. Study conducted by Nabiilah Khoirunnisaa and colleagues. This study focuses on using the k-means algorithm to find trends and classify 34 Indonesian provinces according to the indicators that make up the 2022 HDI. Life expectancy (UHH), expected length of schooling (HLS), average length of schooling (RLS), and expenses are some of these indicators. The goal of this HDI grouping is to determine which HDI variables should be given top priority during development. Based on K-Means Cluster Analysis, two groups (clusters) were formed, according to the analysis results. Provinces in Cluster 1 have high to very high scores for adjusted expenditure, UHH, HLS, and RLS. Provinces with medium to high scores on UHH, HLS, RLS, and adjusted expenditure make up Cluster 2(Kaylista et al., 2023).

The K-means algorithm is used by the author to cluster data in this study. James B. MacQueen popularized the K-Means method. The goal of the k-means method is to cluster objects into k clusters (k < n), where k has a predetermined value Data(Ahsan, 2023). The first step in the clustering process is to identify the data to be clustered. Xij (i = 1,.., n), j = 1,.., m), where m is the attribute (variable) and n is the quantity of data to be clustered. Each cluster's center is chosen at random at the start of the iteration. ckj(k=1,...,k) ; j=1,...,m. Next, the centroid—the distance—between each cluster and each set of data is computed. Euclidian distance is used to calculate distances, specifically(Azmi, 2020):

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2}$$

Information :
d = distance of object to centroid di-k
i = number of attributes (data dimensions)
y = centroid
x = data

K-means clustering can be effective when applied to small data sets. Big data sets need to be clustered so that each entity or other data point is identical to all other entities in the cluster. The k-means method has the following benefits(Djamro, 2018):
1) It is simple to use.
2) The algorithm is quick to execute.
3) Capable of classifying a set of patterns into related groupings.

METHOD

The stages in this research include research steps. The framework in this research is described as follows:

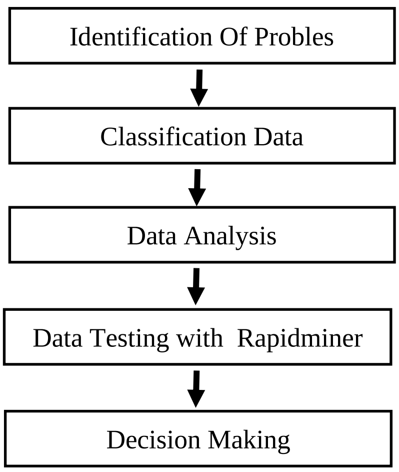

Figure 1: Research Framework

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13148
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

Secondary data obtained from the official publication website of government institutions, namely the Central Statistics Agency (BPS), was used in this study which was obtained via www.bps.go.id. The data used is Human Development Index (HDI) data based on provinces in Indonesia in 2022. Next, the data will be processed based on what was previously obtained using the k-means algorithm through the use of data processing tools, namely Rapidminer Data Mining, to determine the number of HDI clusters. appropriate. The data used in this research is statistical data for 2017 – 2022 for the Regency and City of West Java Province.

Table 1. Human Development Index Value Data 2017 - 2022

Number	West Java Province	Human Development Index						Average
		2022	2021	2020	2019	2018	2017	
1	Bogor	71,20	70,60	70,40	70,65	69,69	69,13	70,28
2	Sukabumi	67,64	67,07	66,88	66,87	66,05	65,49	66,67
3	Cianjur	65,94	65,56	65,36	65,38	64,62	63,7	65,09
4	Bandung	73,16	72,73	72,39	72,41	71,75	71,02	72,24
5	Garut	67,41	66,45	66,12	66,22	65,42	64,52	66,02
6	Tasikmalaya	66,84	65,9	65,67	65,64	65	64,14	65,53
7	Ciamis	71,45	70,93	70,49	70,39	69,63	68,87	70,29
8	Kuningan	70,16	69,71	69,38	69,12	68,55	67,78	69,12
9	Cirebon	70,06	69,12	68,75	68,69	68,05	67,39	68,68
10	Majalengka	68,56	67,81	67,59	67,52	66,72	65,92	67,35
11	Sumedang	72,69	71,8	71,64	71,46	70,99	70,07	71,44
12	Indramayu	68,55	67,64	67,29	66,97	66,36	65,58	67,07
13	Subang	69,87	69,13	68,95	68,69	68,31	67,73	68,78
14	Purwakarta	71,56	70,98	70,82	70,67	69,98	69,28	70,55
15	Karawang	71,74	70,94	70,66	70,86	69,89	69,17	70,54
16	Bekasi	75,22	74,45	74,07	73,99	73,49	72,63	73,98
17	Bandung Barat	69,04	68,29	68,08	68,27	67,46	66,63	67,96
18	Pangandaran	69,03	68,28	68,06	68,21	67,44	66,6	67,94
19	Kota Bogor	77,17	76,59	76,11	76,23	75,66	75,16	76,15
20	Kota Sukabumi	75,4	74,6	74,21	74,31	73,55	73,03	74,18
21	Kota Bandung	82,5	81,96	81,51	81,62	81,06	80,31	81,49
22	Kota Cirebon	75,89	75,25	74,89	74,92	74,35	74	74,88
23	Kota Bekasi	82,46	81,95	81,5	81,59	81,04	80,3	81,47
24	Kota Depok	81,86	81,37	80,97	80,82	80,29	79,83	80,86
25	Kota Cimahi	78,77	78,06	77,83	78,11	77,56	76,95	77,88
26	Kota Tasikmalaya	73,83	73,31	73,04	72,84	72,03	71,51	72,76
27	Kota Banjar	72,55	71,92	71,7	71,75	71,25	70,79	71,66

In Table 1. There is data on the Human Development Index for 27 Regencies and Cities of West Java Province for 2017-2022. Pre-processing of data is carried out to facilitate the process of searching for information from the results of data collection. Pre-processing of data carried out in research includes: Data Cleaning At this stage, data that is not needed is deleted, namely by deleting

*name of corresponding author

Indonesian data in the dataset, because Indonesian data is data on the average value of the Human Development Index in West Java Province. Data Transformation At this stage to make it easier at the calculation stage with the k-means dataset algorithm which has as many as 2 attributes (Regency/City, Province and HDI). The average HDI value was calculated from 2017-2022. So that the attributes become one with the average value. The results of data pre-processing can be seen in table 2.

Table 2. Dataset After Data Preprocessing

Number	West Java	Average
1	Bogor	70,28
2	Sukabumi	66,67
3	Cianjur	65,09
4	Bandung	72,24
5	Garut	66,02
6	Tasikmalaya	65,53
7	Ciamis	70,29
8	Kuningan	69,12
9	Cirebon	68,68
10	Majalengka	67,35
11	Sumedang	71,44
12	Indramayu	67,07
13	Subang	68,78
14	Purwakarta	70,55
15	Karawang	70,54
16	Bekasi	73,98
17	Bandung Barat	67,96
18	Pangandaran	67,94
19	Kota Bogor	76,15
20	Kota Sukabumi	74,18
21	Kota Bandung	81,49
22	Kota Cirebon	74,88
23	Kota Bekasi	81,47
24	Kota Depok	80,86
25	Kota Cimahi	77,88
26	Kota Tasikmalaya	72,76
27	Kota Banjar	71,66

After the data set in table 2 is ready, clustering modeling is carried out using the k-means algorithm. Based
on the data mining processing steps with the k-means algorithm, the following stages will be carried out:

a. The first step that needs to be done is determining the number of clusters. Where 3 clusters are determined, consisting of the first cluster (C0) with the highest value, the second cluster (C1) with the middle value, the third cluster (C2) with the lower value.
b. Selecting the initial centroid, at this stage it will be chosen randomly from the dataset in table 2. The initial cluster center can be seen in table 3.

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13148
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 3 Initial Center Cluster - Iteration 1

Number	Cluster	Regency	IPM
1	C0	Kota Bandung	81,49
2	C1	Karawang	70,54
3	C2	Cianjur	65,09

Table 3, Initial central cluster HDI C0 Bandung City 81.49, HDI C1 Karawang 70.54, HDI C2 Cianjur 65.09.

The results of the entire calculation can be observed as grouping results according to iteration 1.

Table 4. Cluster Center Distance Calculation Results - Iteration 1

Number	West Java	IPM	C0	C1	C2	Shortest Distance	C0	C1	**C2**
1	Bogor	70,28	11,21	0,26	5,19	0,26		1	
2	Sukabumi	66,67	14,82	3,87	1,58	1,58			**1**
3	Cianjur	65,09	16,4	5,45	0	0			**1**
4	Bandung	72,24	9,25	1,7	7,15	1,7		1	
5	Garut	66,02	15,47	4,52	0,93	0,93			1
6	Tasikmalaya	65,53	15,96	5,01	0,44	0,44			1
7	Ciamis	70,29	11,2	0,25	5,2	0,25		1	
8	Kuningan	69,12	12,37	1,42	4,03	1,42		1	
9	Cirebon	68,68	12,81	1,86	3,59	1,86		1	
10	Majalengka	67,35	14,14	3,19	2,26	2,26			1
11	Sumedang	71,44	10,05	0,9	6,35	0,9		1	
12	Indramayu	67,07	14,42	3,47	1,98	1,98			1
13	Subang	68,78	12,71	1,76	3,69	1,76		1	
14	Purwakarta	70,55	10,94	0,01	5,46	0,01		1	
15	Karawang	70,54	10,95	0	5,45	0		1	
16	Bekasi	73,98	7,51	3,44	8,89	3,44		1	
17	Bandung Barat	67,96	13,53	2,58	2,87	2,58		1	
18	Pangandaran	67,94	13,55	2,6	2,85	2,6		1	
19	Kota Bogor	76,15	5,34	5,61	11,06	5,34	1		
20	Kota Sukabumi	74,18	7,31	3,64	9,09	3,64		1	
21	Kota Bandung	81,49	0	10,95	16,4	0	1		
22	Kota Cirebon	74,88	6,61	4,34	9,79	4,34		1	
23	Kota Bekasi	81,47	0,02	10,93	16,38	0,02	1		
24	Kota Depok	80,86	0,63	10,32	15,77	0,63	1		
25	Kota Cimahi	77,88	3,61	7,34	12,79	3,61	1		
26	Kota Tasikmalaya	72,76	8,73	2,22	7,67	2,22		1	
27	Kota Banjar	71,66	9,83	1,12	6,57	1,12		1	

Based on Table 4. Determine the new centroid by counting the number of selected clusters (C0 has 5 regencies/cities, C1 has 15 regencies/cities, C2 has 6 regencies/cities) then distributes them by the

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13148
e-ISSN : 2541-2019
p-ISSN : 2541-044X

number of selected clusters. Thus, the results of the distance from each object in the 1st iteration are obtained, then proceed to the 2nd iteration according to the calculations below.

Table 5. New Cluster – Iteration 2

Determination of the initial center of the cluster	
The 1st new cluster	79,57
2nd new cluster	52,61
The 3rd new cluster	67,15

In Table 5. Formation of the initial cluster center for the 2nd interaction, consisting of the 1st new cluster is 79.57, the 2nd new cluster is 52.61 and the 3rd new cluster is 67.15.

Table 6. Cluster Center Distance Calculation Results - Iteration 2

Number	West Java	IPM	C0	C1	C2	Shortest Distance	C0	C1	**C2**
1	Bogor	70,28	9,29	17,668	3,1267	3,126666667			**1**
2	Sukabumi	66,67	12,9	14,058	0,4833	0,483333333			**1**
3	Cianjur	65,09	14,48	12,478	2,0633	2,063333333			**1**
4	Bandung	72,24	7,33	19,628	5,0867	5,086666667			**1**
5	Garut	66,02	13,55	13,408	1,1333	1,133333333			1
6	Tasikmalaya	65,53	14,04	12,918	1,6233	1,623333333			1
7	Ciamis	70,29	9,28	17,678	3,1367	3,136666667			1
8	Kuningan	69,12	10,45	16,508	1,9667	1,966666667			1
9	Cirebon	68,68	10,89	16,068	1,5267	1,526666667			1
10	Majalengka	67,35	12,22	14,738	0,1967	0,196666667			1
11	Sumedang	71,44	8,13	18,828	4,2867	4,286666667			1
12	Indramayu	67,07	12,5	14,458	0,0833	0,083333333			1
13	Subang	68,78	10,79	16,168	1,6267	1,626666667			1
14	Purwakarta	70,55	9,02	17,938	3,3967	3,396666667			1
15	Karawang	70,54	9,03	17,928	3,3867	3,386666667			1
16	Bekasi	73,98	5,59	21,368	6,8267	5,59	1		
17	Bandung Barat	67,96	11,61	15,348	0,8067	0,806666667			1
18	Pangandaran	67,94	11,63	15,328	0,7867	0,786666667			1
19	Kota Bogor	76,15	3,42	23,538	8,9967	3,42	1		
20	Kota Sukabumi	74,18	5,39	21,568	7,0267	5,39	1		
21	Kota Bandung	81,49	1,92	28,878	14,337	1,92	1		
22	Kota Cirebon	74,88	4,69	22,268	7,7267	4,69	1		
23	Kota Bekasi	81,47	1,9	28,858	14,317	1,9	1		
24	Kota Depok	80,86	1,29	28,248	13,707	1,29	1		
25	Kota Cimahi	77,88	1,69	25,268	10,727	1,69	1		
26	Kota Tasikmalaya	72,76	6,81	20,148	5,6067	5,606666667			1
27	Kota Banjar	71,66	7,91	19,048	4,5067	4,506666667			1

*name of corresponding author

 e-ISSN : 2541-2019
p-ISSN : 2541-044X

Based on Table 6. Determine the new centroid by counting the number of selected clusters (C0 has 8 districts and cities, C1 has 0 districts and cities, C2 has 19 districts and cities) then divides it by the number of selected clusters. Thus, the results of the distance from each object in the 1st iteration are obtained, then proceed to the 2nd iteration according to the calculations below.

Table 7. New Cluster - Iteration 3

Determination of initial cluster center Iteration 3	
The 1st new cluster	79,57
2nd new cluster	52,61
The 3rd new cluster	79,57

In Table 7. The formation of the initial cluster center for the 3rd interaction, consisting of the 1st new cluster is 79.57, the 2nd new cluster is 52.61 and the 3rd new cluster is 79.57.

Table 8. Cluster Center Distance Calculation Results - Iteration 3

Number	West Java	IPM	C0	C1	C2	Shortest Distance	C0	C1	C2
1	Bogor	70,28	9,29	17,7	9,29	9,29			1
2	Sukabumi	66,67	12,9	14,1	12,9	12,9			1
3	Cianjur	65,09	14,5	12,5	14,5	12,4775			1
4	Bandung	72,24	7,33	19,6	7,33	7,33			1
5	Garut	66,02	13,6	13,4	13,6	13,4075			1
6	Tasikmalaya	65,53	14	12,9	14	12,9175			1
7	Ciamis	70,29	9,28	17,7	9,28	9,28			1
8	Kuningan	69,12	10,5	16,5	10,5	10,45			1
9	Cirebon	68,68	10,9	16,1	10,9	10,89			1
10	Majalengka	67,35	12,2	14,7	12,2	12,22			1
11	Sumedang	71,44	8,13	18,8	8,13	8,13			1
12	Indramayu	67,07	12,5	14,5	12,5	12,5			1
13	Subang	68,78	10,8	16,2	10,8	10,79			1
14	Purwakarta	70,55	9,02	17,9	9,02	9,02			1
15	Karawang	70,54	9,03	17,9	9,03	9,03			1
16	Bekasi	73,98	5,59	21,4	5,59	5,59	1		
17	Bandung Barat	67,96	11,6	15,3	11,6	11,61			1
18	Pangandaran	67,94	11,6	15,3	11,6	11,63			1
19	Kota Bogor	76,15	3,42	23,5	3,42	3,42	1		
20	Kota Sukabumi	74,18	5,39	21,6	5,39	5,39	1		
21	Kota Bandung	81,49	1,92	28,9	1,92	1,92	1		
22	Kota Cirebon	74,88	4,69	22,3	4,69	4,69	1		
23	Kota Bekasi	81,47	1,9	28,9	1,9	1,9	1		
24	Kota Depok	80,86	1,29	28,2	1,29	1,29	1		
25	Kota Cimahi	77,88	1,69	25,3	1,69	1,69	1		
26	Kota Tasikmalaya	72,76	6,81	20,1	6,81	6,81			1
27	Kota Banjar	71,66	7,91	19	7,91	7,91			1

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13148
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Based on Table 8. There are similarities in the results of the calculation of the distance to the center of the cluster - Iteration 3 with the results of the calculation of the distance to the center of the cluster - Iteration 2, so the calculation results were stopped because the positions of C0, C1 and C2 had the same value as the Regency and City cluster positions in the cluster (C0 has 8 districts and Cities, C1 has 0 Districts and Cities, C2 has 19 Regencies and Cities).

At this stage, testing will be carried out using the RapidMiner tool. RapidMiner is software created using Java language and is open which is used as a solution for analyzing data mining, text mining and predictive analysis. for testing the Human Development Index dataset. With the number of clusters that have been determined, there are a total of 3 clusters. Where the dataset processed is 27 data. The plot of data processing results can be seen in Figure 2.

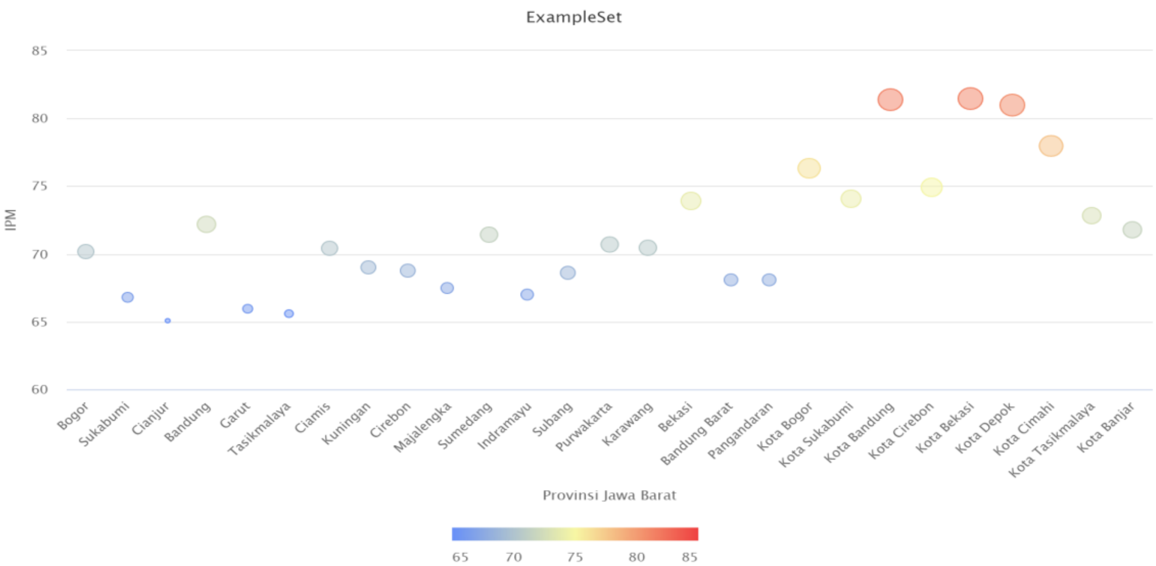

Figure 2. Plot of implementation results on rapidminer

In Figure 2, the results of the implementation of clustering pots are displayed, including:

a. Cluster 0 is in the category with very good HDI scores for Bekasi Regency, Bogor City, Sukabumi City, Bandung City, Cirebon City, Bekasi City, Depok City and Cimahi City.
b. Cluster 1 is in the category with no good HDI scores.
c. Cluster 2 is in the category with low HDI values, for Bogor Regency, Sukabumi Regency, Cianjur Regency, Bandung Regency, Garut Regency, Tasik Regency, Ciamis Regency, Kuningan Regency, Cirebon Regency, Majalengka Regency, Sumedang Regency, Indramayu Regency, Subang Regency , Purwakarta Regency, Karawang Regency, West Bandung Regency, Pangandaran Regency, Tasikmalaya City and Bandar City.

DISCUSSIONS

There are similarities in the Plot graphic model based on the research findings of Hermawan et al. Application of Data Mining for Clustering the Human Development Index Based on Provinces in Indonesia, The Plot graphic model has certain similarities. Points where both the x and y values are positive and trend upward share similarities with the highest HDI. in light of Table 8. There are parallels between the outcomes of the distance to the cluster center - Iteration 3 and the outcomes of the distance to the cluster center - Iteration 2, Because C0, C1, and C2's positions matched the cluster's regency and city cluster positions (C0 has eight districts and cities, C1 has zero districts and cities, and C2 has nineteen regencies and cities), the computation results were halted.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13148
e-ISSN : 2541-2019
p-ISSN : 2541-044X

CONCLUSION

the results of the implementation of clustering pots are displayed, including: Cluster 0 is in the category with very good HDI scores for Bekasi Regency, Bogor City, Sukabumi City, Bandung City, Cirebon City, Bekasi City, Depok City and Cimahi City.Cluster 1 is in the category with no good HDI scores.Cluster 2 is in the category with low HDI values, for Bogor Regency, Sukabumi Regency, Cianjur Regency, Bandung Regency, Garut Regency, Tasik Regency, Ciamis Regency, Kuningan Regency, Cirebon Regency, Majalengka Regency, Sumedang Regency, Indramayu Regency, Subang Regency , Purwakarta Regency, Karawang Regency, West Bandung Regency, Pangandaran Regency, Tasikmalaya City and Bandar City.

REFERENCES

Ahsan. (2023). Peringkasan teks multi dokumen berita berbahasa Indonesia menggunakan Fasttext Dan K-Means Clustering. *Universitas Islam Negeri Maulana Malik Ibrahim*. http://etheses.uin-malang.ac.id/52460/

Anshori, S. (2018). Pemanfaatan Teknologi Informasi Dan Komunikasi Sebagai Media Pembelajaran. *Civic-Culture: Jurnal Ilmu Pendidikan PKn Dan Sosial Budaya*, *9924*, 88–100. file:///C:/Users/HP/Downloads/70-Article Text-536-1-10-20191223.pdf

Azmi, F. (2020). Initial Centroid Optimization of K-Means Algorithm Using Cosine Similarity. *Journal of Informatics and Telecommunication Engineering*, *3*(2), 224–231. https://doi.org/10.31289/jite.v3i2.3211

Bariyah, N. (2022). Pendidikan, Kesehatan dan Penanggulangan Kemiskinan di Kalimantan Barat: Menuju Sustainable Development Goals. *Jurnal Ilmu Sosial Dan Humaniora*, *11*(1), 93–110. https://doi.org/10.23887/jish.v11i1.39343

Daniati, R. R., & Prasetyo, H. D. (2022). Pengaruh Herding Dan Overconfidence Terhadap Keputusan Investasi. *Jurnal Revenue : Jurnal Ilmiah Akuntansi*, *3*(1), 10–14. https://doi.org/10.46306/rev.v3i1.92

Djamro, R. A. (2018). Penerapan Algoritma K-Means Dalam Memilih Tanah Yang Tepat Untuk Tanaman Padi. *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, *VII*(1), 12–20.

Hamid, H. (2018). *Manajemen Pemberdayaan Masyarakat*. De La Macca.

Hermawan, & Hasugian, H. (2022). Penerapan Data Mining Untuk Clustering Indeks Pembangunan Manusia Berdasarkan Provinsi Di Indonesia. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, *1*(1), 525–532.

Kaylista, N. N. ., Khoirunnisaa, N., Viewianti E.N.F, G., & Yusuf, A. Y. P. (2023). Implementasi Algoritma K-Means Clustering Menggunakan Aplikasi Orange Untuk Mengetahui Pola Indeks Pembangunan Manusia Tahun 2022. *Journal of Information and Information Security (JIFORTY)*, *4*(1), 65–76. http://ejurnal.ubharajaya.ac.id/index.php/jiforty

Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hirearchical , k-means , dan k-medoids clustering dalam pengelompokkan indeks pembangunan manusia Indonesia Comparative analysis of hirearchical , k-means , and k-medoids clustering and methods in grouping Indonesia ' s human. *Inovasi*, *17*(4), 770–782.

Pamungkas Wijayanto. (2022). Pengaruh Upah Minimum, Indeks Pembangunan Manusia (Ipm), Jumlah Penduduk Dan Pertumbuhan Ekonomi Terhadap Tingkat Pengangguran Terbuka Di Provinsi D.I Yogyakarta Pada Tahun 2016 - 2021. *Universitas Islam Indonesia*.

Permadi, H. (2022). Solusi untuk Meningkatkan Knowledge Management Readness di Badan Pusat Statistik Kabupaten/Kota. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *9*(1), 97. https://doi.org/10.25126/jtiik.2022914918

Suhanda, Y. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, *6*(2), 12–20. https://doi.org/10.37012/jtik.v6i2.299

Zamroni, A. (2020). Penerapan Sistem Informasi Manajemen Pendidikan dalam Proses Pembelajaran di Sekolah Menengah Pertama. *Jurnal Manajemen Pendidikan Islam E-ISSN: On Process*, *1*, 11–21.

*name of corresponding author