JEECS (Journal of Electrical Engineering and Computer Sciences)
Vol. 10, No. 1, June 2025, pp. 49-58
e-ISSN: 2579-5392 p-ISSN: 2528-0260

Application of K-means Clustering Data Mining in Grouping Data of People with Disabilities

Moh. Baha'uddin, Zaehol Fatah

Department of Information Technology, Faculty of Science and Technology, Universitas Ibrahimy, Jl. KHR. Syamsul Arifin No. 1-2, Situbondo, 68374, East Java, Indonesia

Article Info

Article history:
Received: 17 November 2024
Revised: 31 December 2024
Accepted: 3 January 2025

Keyword:
Data mining
Clustering
K-Means
Social welfare
People with disabilities

Abstract

Data mining is critical in enabling organizations to derive reliable insights from data. Social welfare remains a significant challenge in Indonesia, particularly for people with disabilities, emphasizing the need for targeted strategies. However, developing research has not used natural characteristics according to disability problems. This study utilizes the K-Means Clustering algorithm to analyze and categorize the population of people with disabilities in East Java. The attributes include the type of disability, population size, and regional distribution. We employs a dataset from the East Java Central Bureau of Statistics, comprising 342 data points across eight attributes, including region, disability type, and year. The analysis involves data preprocessing, transformation, clustering, and evaluation using the Davies-Bouldin Index (DBI). The results identify two optimal clusters, achieving the lowest DBI score of 0.097, indicating high cluster quality. Cluster 0 represents regions with fewer people with disabilities, while Cluster 1 highlights areas with higher populations. These findings provide a foundation for developing more focused and inclusive welfare programs tailored to regional needs, enhancing the quality of life for people with disabilities.

Corresponding author:
Moh. Baha'uddin, gussaintek21@gmail.com

DOI: https://doi.org/10.54732/jeecs.v10i1.6

1. Introduction

Advances in information and communication technology (ICT) have brought significant changes in various aspects of life, including the management of social data for the welfare of the community [1]. Social welfare is one of the main goals of the Indonesian nation, as stated in the 1945 Constitution of the Republic of Indonesia [2]. However, until now, this goal has not been fully achieved. One of the main challenges is the existence of People with Social Welfare Problems (MSMEs), especially in urban areas. MSMEs can interfere with community comfort, social stability, and urban development [3].

One of the significant groups in MSMEs is people with disabilities, namely individuals with special needs who have different characteristics compared to people in general, both in terms of physical, mental, and [4]. People with disabilities are a heterogeneous population, including individuals with physical, mental, or a combination of both [5]. In recent years, the increase in the number of people with disabilities in Indonesia, especially in East Java, requires data-based strategic handling.

Previous research has shown the importance of a technology-based approach to social data management. The research of [6] stated that grouping data on persons with disabilities based on disability categories can help increase the effectiveness of service programs. The researcher of [7] noted that a technology-based approach can reduce inequality in the accessibility of social services for people with disabilities. Also, the research of [8] show that the K-Means algorithm is effective in grouping heterogeneous data in the social service sector. This is reinforced by [9] who note that grouping based on the K-Means algorithm can help governments design more targeted inclusive policies. Research by [10] adds that the visualization of clustering data can support a deeper understanding of the distribution and needs of people with disabilities in various regions.

This study aims to cluster the population of people with disabilities in East Java based on disability categories using the K-Means algorithm. This approach is expected to provide a more specific understanding of the needs of each group, thus supporting efficient and inclusive service program planning. Thus, the results of this grouping can be the basis for governments and related institutions to design data-driven policies that support equality, accessibility, and social inclusion.

2. Research Methodology

2.1. Data Collection

Data collection in this research was carried out to obtain and analyze information related to the population with disabilities in East Java. The data used was obtained from the East Java Central Statistics Agency (BPS) from the link jatim.bps.go.id which includes records of the number of disabled people based on disability categories per district/city in 2019. The documents used include annual reports detailing disability categories, such as blind, deaf, physically disabled, mentally retarded , disabled, and disabled by former leprosy. This data is then processed using the RapidMiner application for further analysis.

2.2. Data Processing

The data processing in this research aims to group the population of people with disabilities in East Java based on disability categories using the K-Means Clustering method. The steps in the data processing include:

a. Data Preparation
   Data collection was conducted using official records provided by the East Java Central Bureau of Statistics (BPS). The dataset contains information on the number of people with disabilities categorized by type, such as visually impaired, hearing impaired, physically disabled, and others, for each city/regency in 2019. The data was formatted in an Excel file (.xlsx) and imported into RapidMiner for further analysis.
b. Clustering Process with K-Means
   In this phase, the number of clusters is determined (for example, k = 3) to represent specific categories of disabilities. The data is grouped based on the distance to centroids using the K-Means algorithm, continuing the iteration process until the results converge.
c. Cluster Evaluation
   Once the clusters are formed, the quality of the clustering is assessed using the Davies-Bouldin Index (DBI). The best clustering model is selected by choosing the one with the smallest DBI value after conducting experiments with various numbers of clusters.
d. Regional Analysis
   Following the formation of the clusters, analysis is performed to identify the characteristics of each region in East Java, focusing on the distribution of different disability categories across various areas.
e. Clustering Results
   The results of the data processing categorize the population of people with disabilities into clusters. For instance, Cluster 1 represents regions with a high number of visually impaired individuals, Cluster 2 reflects areas with a combination of physical and intellectual disabilities, and Cluster 3 highlights regions with other disability categories.

These steps, from data collection to the evaluation of clusters, provide a comprehensive overview of the distribution of disabilities in East Java. This analysis plays a crucial role in helping local governments design more inclusive policies and programs based on this data-driven approach

2.3. Clustering

Clustering is a key method in data analysis used to group objects into specific clusters based on similar characteristics. In clustering, objects within the same cluster exhibit higher similarity compared to objects in other clusters [11]. These clusters are sequentially combined based on predefined criteria until all data points are grouped into one cluster or until a specific termination criterion is met [12].

Clustering is the process of grouping data points where the class attribute has not been previously described. Conceptually, clustering aims to maximize intra-cluster similarity while minimizing inter-cluster similarity. For example, a set of objects can be initially clustered into several groups, which are then refined into a more structured set, based on certain classification groups. Essentially, clustering results in a number of clusters (groups). Clustering refers to grouping records, observations, or entities that share similarities.

A cluster is a collection of records that are similar to one another, while being distinct from records in other clusters. Clustering attempts to partition the entire dataset into groups that have relatively high similarity, where the similarity within a cluster is maximized, and the similarity between records in different clusters is minimized [13].

2.4. Disability

People with disabilities are citizens entitled to respect, protection, and fulfillment of their basic rights by the state. The human rights of people with disabilities are guaranteed by the Indonesian Constitution of 1945 (UUD NRI 1945) and various regulations concerning disabilities. One of these rights is participation in efforts to prevent and protect against violence, particularly for women with disabilities [14].

2.5. K-Means

The K-Means method involves determining and configuring the number of clusters to serve as centers or references in the data grouping process using the K-Means library [15]. Before applying the K-Means algorithm, the data must undergo preprocessing. K-Means is categorized as a partitioning clustering method that divides data into distinct segments. This algorithm is highly popular due to its simplicity and effectiveness in grouping large datasets while handling outliers [16].

From Figure 1, you can see the flow diagram of the K-Means method, which begins by determining the desired number of clusters. Once the number of clusters is decided, the next step is to determine the initial positions of the cluster centers (centroids). Then, the distance between each data object and each cluster center is calculated using the Euclidean distance formula to identify the shortest distance between each data point and the existing centroids. Based on this minimum distance, objects are grouped into the appropriate clusters.

The cluster centers are then updated by calculating the average position of the objects within each cluster (the new centroid). If any objects need to be reassigned to a different cluster, the process is repeated. This iterative process continues until convergence is achieved. The following are the steps of the K-Means algorithm used:

- Determine k as the number of clusters to be formed.
- Randomly determine the initial k centroids.
- Calculate the distance of each object to each centroid of the clusters using the Euclidean Distance method, as shown in equation (1):

$$d_{Euclidean}(y,x) = \sqrt{\sum_{i=1}^{n}(x_1 - y_1)^2} \qquad (1)$$

- Assign each object to the closest centroid.
- Perform an iteration, then determine the new centroid position using equation (2):

$$C = \frac{\sum m}{n} \qquad (2)$$

  where :
  C : centroid of the data
  m: members of the data that belong to the closest cluster
  n : the number of data points that are members of a specific cluster.
- Repeat Step 3 if the new centroid positions are not consistent.

2.6. Davies-Bouldin Index (DBI)

The Davies-Bouldin Index is a method used to measure the validity of clusters in the clustering process. Cohesion is defined as the total closeness between data points and the cluster center within the respective cluster.

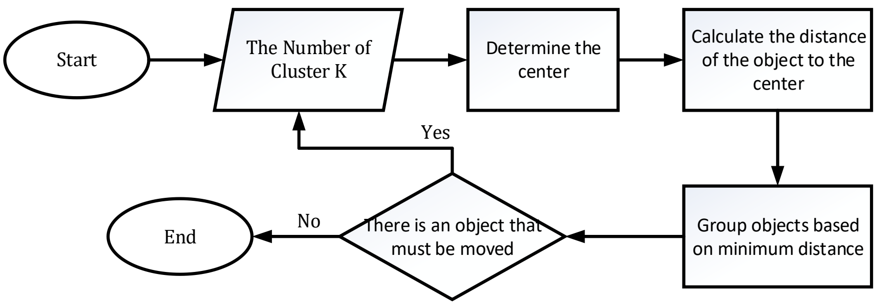


Figure 1. K-Means Flowchart

The Davies-Bouldin Index (DBI) was introduced by David L. Davies and Donald W. Bouldin in 1979. It is another metric for evaluating clustering algorithms. This index is defined as the average similarity measure between each cluster and the most similar cluster. Similarity is the ratio of within-cluster distance to between-cluster distance. The minimum value of DBI is 0, and a smaller value (closer to 0) represents a better model that produces better clusters [17].

The Davies Bouldin index (DBI) formula is used to assess the quality of clustering results. This index measures the extent to which clusters are separated from each other and the extent to which they are dense. The lower the DBI value, the better the clustering results (clusters are denser and more clearly separated). The following is the formula for calculating the Davies-Bouldin Index in a clustering [18].

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} \left(\frac{d(c_i, c_j)}{a_i + a_j}\right) \tag{3}$$

where :

- K is the number of clusters.
- a_i is the size (density) of cluster i, usually calculated as the average distance between each point in the cluster and the cluster center (centroid).
- d (c_i,c_j) is the distance between cluster centers c_i and c_j.
- Max i=j that we look for pairs of clusters that have the maximum ratio between density and distance between centers.

Explanation:

- This formula calculates the DBI value by adding up the maximum value of the ratio of the distance between cluster centers to the number of standard deviations for each different pair of clusters.
- Lower DBI values indicate better clustering.

2.7. Research Flow Diagram

In this research, the methodology flowchart is used to clearly and systematically depict the steps involved, from data collection to result visualization. This flowchart provides an overview of the research process, which involves various stages in data processing and the application of the K-Means Clustering method.

The following are the stages involved in the research methodology:

1. Data Collection
   The first stage involves collecting data, which is done by accessing information available from the Indonesian Central Statistics Agency (BPS). This dataset includes the number of people with disabilities based on disability categories in East Java in 2019. It serves as the foundation for further analysis to identify patterns among the disabled population.
2. Data Processing
   Once the data is collected, it undergoes processing in the next stage. This step involves validation and pre-processing to ensure the dataset is free from missing values and inconsistencies. By ensuring data quality, this stage lays the groundwork for accurate analysis.
3. Data Transformation
   In this stage, the disability categories are converted into numerical form so they can be analyzed using the K-Means Clustering method. This step is critical because K-Means requires numerical data to compute distances and create clusters.
4. K-Means Application
   This stage involves applying the K-Means Clustering algorithm to group the data based on disability categories. The number of clusters (K) is determined beforehand, representing different disability groups. The algorithm iteratively adjusts the cluster centroids until they stabilize.
5. DBI Evaluation
   After forming the clusters, their quality is evaluated using the Davies-Bouldin Index (DBI). A lower DBI value indicates better cluster compactness and separation, which helps determine the optimal clusters.
6. Result Visualization
   The final stage is result visualization, where the clustering outcomes are displayed in graphical forms, such as scatter plots. This provides a clear representation of the distribution of disability clusters across different categories.

The Figure 2 is the flowchart that depicts these stages.

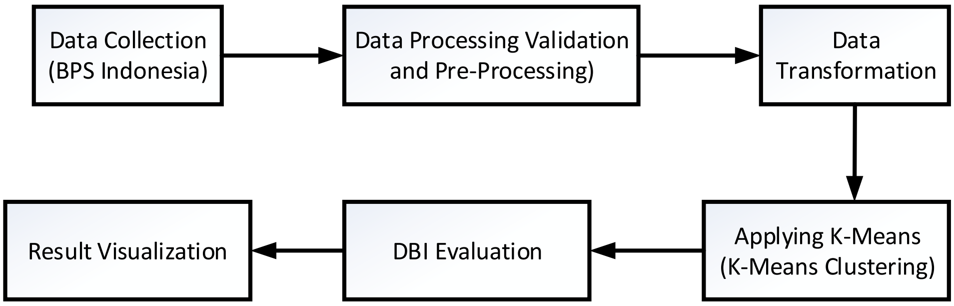


Figure 2. Research Methodology Flowchart

3. Results and Discussions

3.1. Data Selection

The dataset used in this study consists of 342 entries, with attributes including No, Province Name, City/District Name, Disability Category, Population Count, Unit, and Year. These attributes form the basis for the clustering analysis. The data was sourced from the official website of Statistics Indonesia (Badan Pusat Statistik Indonesia), which provides comprehensive information on people with disabilities in East Java, ensuring its relevance and accuracy. The dataset is presented in Table 1.

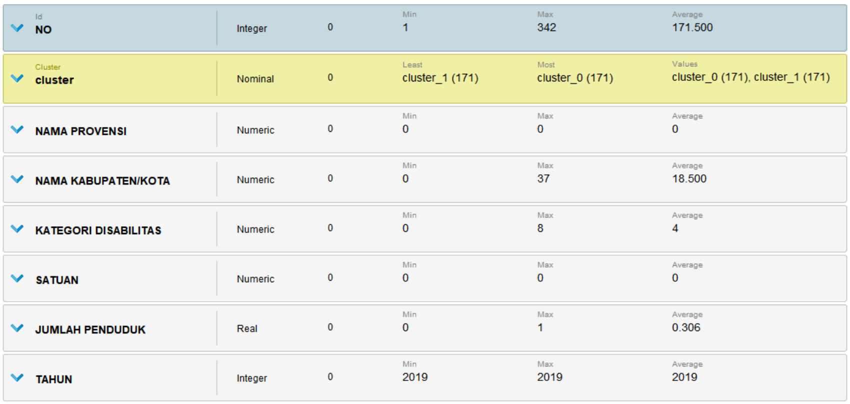
This dataset includes key metrics such as disability categories, population per category, and the year of data collection. All entries have been confirmed to contain no missing values for crucial columns such as "Number," "Province Name," "City/District Name," "Disability Category," "Population," "Unit," and "Year." The absence of missing data indicates that the dataset is clean and complete, ready for further analysis without additional data processing. The result is presented in Table 2.

Table 1. Data Import in the Read Excel Operator

No	Province *(Polynominal)*	City *(Polynominal)*	Disability Category *(Polynominal)*	Population *(Integer)*	Unit *Polynominal)*	Year *Integer)*
1	Jawa Timur	Pacitan	Penyandang Cacat	171	Jiwa	2019
2	Jawa Timur	Pacitan	Tunanetra	132	Jiwa	2019
3	Jawa Timur	Pacitan	Tunarungu	114	Jiwa	2019
4	Jawa Timur	Pacitan	Tunawicara	117	Jiwa	2019
5	Jawa Timur	Pacitan	Tunarungu-Wicara	114	Jiwa	2019
6	Jawa Timur	Pacitan	Tunadaksa	153	Jiwa	2019
7	Jawa Timur	Pacitan	Tunagrahita	142	Jiwa	2019
8	Jawa Timur	Pacitan	Tunalaras	91	Jiwa	2019
9	Jawa Timur	Pacitan	Cacat Eks Sakit Kusta	14	Jiwa	2019
10	Jawa Timur	Ponorogo	Penyandang Cacat	306	Jiwa	2019
...	...	...	...	...	...	...
...	...	...	...	...	...	...
342	Jawa Timur	Kota Batu	Cacat Eks Sakit Kusta	0	Jiwa	2019

Table 2. Data in the Set Role Operator

Row No.	Province	City	Disability Category	Population	Unit	Year
1	Jawa Timur	Pacitan	Penyandang Cacat	171	Jiwa	2019
2	Jawa Timur	Pacitan	Tunanetra	132	Jiwa	2019
3	Jawa Timur	Pacitan	Tunarungu	114	Jiwa	2019
4	Jawa Timur	Pacitan	Tunawicara	117	Jiwa	2019
5	Jawa Timur	Pacitan	Tunarungu-Wicara	114	Jiwa	2019
6	Jawa Timur	Pacitan	Tunadaksa	153	Jiwa	2019
7	Jawa Timur	Pacitan	Tunagrahita	142	Jiwa	2019
8	Jawa Timur	Pacitan	Tunalaras	91	Jiwa	2019
9	Jawa Timur	Pacitan	Cacat Eks Sakit Kusta	14	Jiwa	2019
10	Jawa Timur	Ponorogo	Penyandang Cacat	306	Jiwa	2019
...	...	...	...	...	...	...
...	...	...	...	...	...	...
342	Jawa Timur	Kota Batu	Cacat Eks Sakit Kusta	0	Jiwa	2019



Attribute	Type	Missing	Min / Least	Max / Most	Average / Values
Id NO	Integer	0	Min 1	Max 342	Average 171.500
Cluster cluster	Nominal	0	Least cluster_1 (171)	Most cluster_0 (171)	Values cluster_0 (171), cluster_1 (171)
NAMA PROVENSI	Numeric	0	Min 0	Max 0	Average 0
NAMA KABUPATEN/KOTA	Numeric	0	Min 0	Max 37	Average 18.500
KATEGORI DISABILITAS	Numeric	0	Min 0	Max 8	Average 4
SATUAN	Numeric	0	Min 0	Max 0	Average 0
JUMLAH PENDUDUK	Real	0	Min 0	Max 1	Average 0.306
TAHUN	Integer	0	Min 2019	Max 2019	Average 2019

Figure 3. Data Preprocessing Results

This table shows the dataset after role assignment, ensuring that the attributes have been correctly grouped for the clustering process. Variables such as "Disability Category" and "Population" were specifically adjusted for analysis, providing the necessary data for clustering.

3.2. Data Preprocessing

In the data processing step, the dataset was evaluated to ensure no missing values. All critical columns were confirmed to be complete, eliminating the need for further data cleaning. Attributes like "Disability Category" and "Population" are crucial for the analysis, with the former used to group individuals with disabilities by type, and the latter providing the population distribution across regions. These attributes were processed to prepare the dataset for clustering analysis.

The dataset is already complete, which eliminates the need for imputation or deletion of missing data, making it highly suitable for clustering analysis. The next step in data processing involves converting categorical variables into numeric forms, as clustering algorithms require numeric data. The result is presented in Figure 3.

3.2. Data Transformation

The Data Transformation process involved using the Normalize and Nominal to Numerical operators to prepare the data for clustering analysis. The "Population Count" attribute was normalized to ensure consistency across the dataset, making it suitable for K-Means clustering. The Nominal to Numerical operator was used to convert categorical variables, such as "Disability Category," into numerical values, allowing the clustering algorithm to process them effectively. Table 3 below shows the transformed dataset, where categorical data has been converted into numerical values, making it compatible with the clustering process.

Table 3. Results of the Nominal to Numerical Operator

Row No.	Province	City	Disability Category	Population	Unit	Year
1	0	0	0	0	171	2019
2	0	0	1	0	132	2019
3	0	0	2	0	114	2019
4	0	0	3	0	117	2019
5	0	0	4	0	114	2019
6	0	0	5	0	153	2019
7	0	0	6	0	142	2019
8	0	0	7	0	91	2019
9	0	0	8	0	14	2019
10	0	1	0	0	306	2019
...	...	...	...	...	...	...
...	...	...	...	...	...	...
342	0	37	8	0	0	2019

3.3. Data Mining Model

The clustering process utilized the K-Means algorithm, where the number of clusters (K) was determined through evaluation using the Davies-Bouldin Index (DBI). After running tests with K ranging from 2 to 10, the optimal value of K was determined based on the lowest DBI value, which indicated the best clustering performance. The clustering test scenario is presented in Figure 4.

3.4. Evaluation

Evaluation was conducted using the Davies-Bouldin Index (DBI), which measures the compactness and separation between clusters. A lower DBI value indicates better clustering results. The analysis revealed that K=2 provided the lowest DBI, indicating that two clusters represented the dataset's structure more effectively. In this study, evaluation was performed using the Davies-Bouldin Index (DBI) as a measure of clustering quality. DBI assesses the quality of clusters formed by considering the distance between clusters and intra-cluster distance. Table 4 presents the DBI values for each tested cluster count.

The DBI value serves as a useful metric for assessing clustering quality. A lower DBI value indicates better clustering results. This study found that K=2 yielded the lowest DBI, indicating that this clustering model was the most optimal.

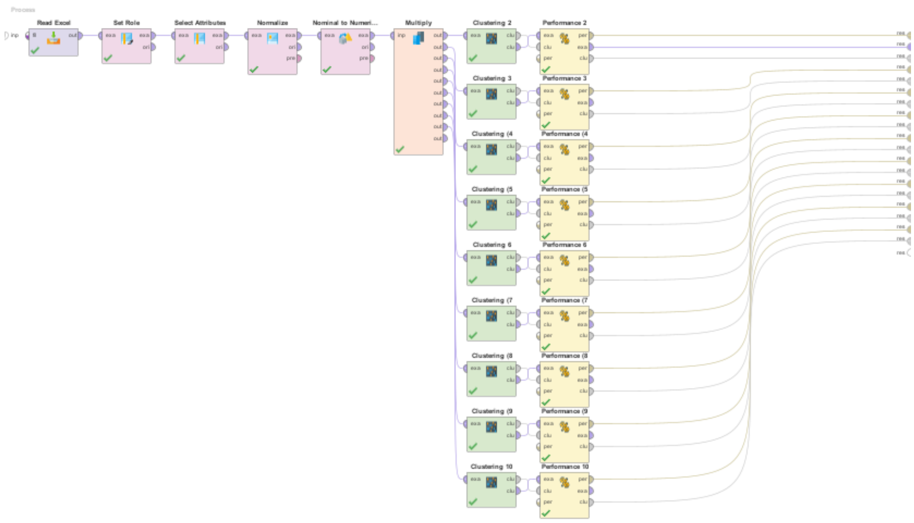


Figure 4. Data Mining Model

Table 4. DBI Values for Each Number of Clusters

Cluster	DBI Value
K2	0.097
K3	0.110
K4	0.125
K5	0.140
K6	0.157
K7	0.156
K8	0.153
K9	0.149
K10	0.160

Table 5. Clustering results

Cluster	Data Grouping
2	Group 0 : 171 data
	Group 1 : 171 data
3	Group 0 : 117 data
	Group 1 : 108 data
	Group 2 : 117 data
4	Group 0 : 90 data
	Group 1 : 81 data
	Group 2 : 91 data
	Group 3 : 90 data

5	Group 0 : 72 data
	Group 1 : 72 data
	Group 2 : 63 data
	Group 3 : 72 data
	Group 4 : 63 data
6	Group 0 : 54 data
	Group 1 : 54 data
	Group 2 : 63 data
	Group 3 : 54 data
	Group 4 : 63 data
	Group 5 : 54 data
7	Group 0 : 54 data
	Group 1 : 54 data
	Group 2 : 34 data
	Group 3 : 54 data
	Group 4 : 58 data
	Group 5 : 34 data
	Group 6 : 54 data
8	Group 0 : 33 data
	Group 1 : 54 data
	Group 2 : 31 data
	Group 3 : 37 data
	Group 4 : 54 data
	Group 5 : 48 data
	Group 6 : 53 data
	Group 7 : 32 data
9	Group 0 : 54 data
	Group 1 : 40 data
	Group 2 : 37 data
	Group 3 : 36 data
	Group 4 : 34 data
	Group 5 : 37 data
	Group 6 : 35 data
	Group 7 : 34 data
	Group 8 : 35 data
10	Group 0 : 33 data
	Group 1 : 36 data
	Group 2 : 42 data
	Group 3 : 36 data
	Group 4 : 28 data
	Group 5 : 33 data
	Group 6 : 33 data
	Group 7 : 38 data
	Group 8 : 27 data
	Group 9 : 36 data

The testing was performed using DBI on two clusters with a total of 342 data points. The results showed an average centroid distance of 6.117, with an average intra-cluster distance of 6.116 for cluster_0 and 6.119 for cluster_1. The DBI value obtained from measuring the distance between points within clusters was 0.97. The clustering result is presented in Table 5.

From the evaluation results, the lowest DBI value was achieved at K = 2, with a score of 0.097. This value is close to zero, indicating that two clusters provided the most optimal and distinct grouping of the data. Therefore, the clustering model used in this study was conducted with two clusters.

3.5. Results Visualization

The clustering results were visualized using a scatter plot, shown in Figure 5, where each point represents a district or city. The vertical axis indicates the population of individuals with disabilities, and the clusters are distinguished by color. Cluster 0 (blue) represents districts with a lower population of individuals with disabilities, while Cluster 1 (green) shows areas with higher concentrations of individuals with disabilities.

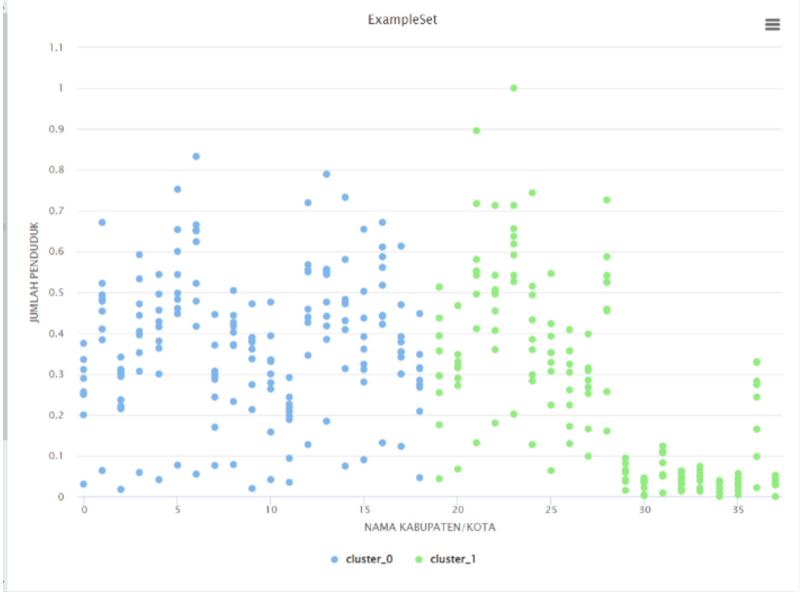


Figure 5. Scatter Bubble Visualization of Clustering Results

Figure 5 illustrates the visualization of clustering results in a scatter bubble plot. Each point represents a district or city, with the vertical axis showing the population of individuals with disabilities. In this graph, the clusters are divided into two groups: Cluster 0 and Cluster 1. Cluster 0 (depicted in blue) represents areas with lower populations of individuals with disabilities. This cluster includes 19 districts/cities, such as Pacitan, Ponorogo, Trenggalek, Tulungagung, and others. Cluster 1 (depicted in green) represents areas with higher populations of individuals with disabilities. This cluster also includes 19 districts/cities, such as Magetan, Ngawi, Bojonegoro, Tuban, and others. This visualization simplifies the observation of patterns in the distribution of individuals with disabilities in each region, allowing for the identification of districts/cities with higher and lower numbers of individuals with disabilities.

4. Conclusion

The application of the K-Means algorithm in grouping data on people with disabilities in East Java has shown its effectiveness in identifying areas based on the number of people with disabilities, resulting in two main groups, namely areas with a low number of people with disabilities and areas with a higher number, with an optimal configuration at K = 2 which was achieved through the lowest Davies-Bouldin Index (DBI) value with a value of 0.097, indicating good cluster quality; These findings provide a valuable basis for the government and related agencies to design inclusive programs that are more targeted according to the needs of people with disabilities, especially in areas with high concentrations, so as to improve their welfare, inclusion and social participation, while providing deeper insight into population distribution. people with disabilities to support the development of effective and adaptive policy strategies in the future.

While key suggestions include a focus on interventions in high concentration areas, further research to include demographic factors such as age and type of disability, close collaboration with local organizations to create more relevant programs, and regular monitoring and evaluation to ensure the sustainability and effectiveness of such programs.

References

[1] L. N. Indahsari, I. Yunita, and Z. Fatah, “Information System Design Using Cutomer Relationship Management (CRM) Method at Paglak Petung Cafe and Art in Banyuwangi District,” *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 617–622, 2024, doi: https://doi.org/10.52436/1.jutif.2024.5.4.2263.

[2] A. Rahman, “Persepsi Masyarakat Terhadap Penyandang Disabilitas Di Kelurahan Bongki Kecamatan Sinjai Utara Kabupaten Sinjai,” vol. 13, pp. 93–99, 2021.

[3] I. A. Darmawan, M. F. Randy, I. Yunianto, M. M. Mutoffar, and M. T. P. Salis, “Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial,” *Sebatik*, vol. 26, no. 1, pp. 223–230, 2022, doi: 10.46984/sebatik.v26i1.1622.

[4] A. I. Sari, H. S. Tambunan, W. Saputra, I. S. Damanik, and I. S. Saragih, “Implementasi Algoritma K-Means Dalam Pengelompokan Penyandang Disabilitas Menurut Kecamatan Kabupaten

Simalungun," in *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2020, vol. 2, pp. 54–61.

[5] H. Hidayaturrahman, "Pelaksanaan Hak Penyandang Disabilitas Di Kota Pekanbaru Berdasarkan Undang–Undang Nomor 8 Tahun 2016," *Jurnal Gagasan Hukum*, vol. 4, no. 01, pp. 14–28, 2022, doi: 10.31849/jgh.v4i01.10439.

[6] P. Lestari, N. Suarna, and W. Prihartono, "Implementasi Data Mining Clustering Dalam Mengelompokkan Penduduk Penyandang Disabilitas Menggunakan Algoritma K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 958–966, 2024, doi: 10.36040/jati.v8i1.8445.

[7] S. Hanafi, Y. Djabbar, M. Fahri, S. P. Jasmin, and M. Zulhidayat, "Tantangan dalam Implementasi Kebijakan Perlindungan Hak Asasi Manusia bagi Penyandang Disabilitas di Provinsi DKI Jakarta," *Jurnal Hukum dan HAM Wara Sains*, vol. 2, no. 6, pp. 509–516, 2023, doi: 10.58812/jhhws.v2i6.446.

[8] A. Rahmawati, R. Roekhudin, and A. Prastiwi, "Effect of good corporate governance and corporate social responsibility on firm value moderate by profitability," *International Journal of Research in Business and Social Science*, vol. 10, no. 4, pp. 59–66, 2021, doi: 10.20525/ijrbs.v10i4.1194.

[9] S. Arrohman and Z. Fatah, "Prediksi Diabetes Menggunakan Algoritma Klasifikasi K-Nearest Neighbors (K-NN) pada Perempuan Indian Pima," *Gudang Jurnal Multidisiplin Ilmu*, vol. 2, no. 10, pp. 220–226, 2024, doi: https://doi.org/10.59435/gjmi.v2i10.986.

[10] A. Sujjada, G. P. Insany, and S. Noer, "Analisis Clustering Data Penyandang Disabilitas Menggunakan Metode Agglomerative Hierarchical Clustering dan K-means," *Jurnal Teknologi dan Manajemen Informatika*, vol. 10, no. 1, pp. 1–12, 2024, doi: 10.26905/jtmi.v10i1.10654.

[11] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024.

[12] S. Trihandaru, *Clustering Untuk Data Suara Studi Kasus Mel-Frequency Cepstral Coefficients (MFCC) dan Long Short Term Memory (LSTM) dengan Internet of Things (IoT) untuk Klasifikasi Suara*. Uwais Inspirasi Indonesia, 2024.

[13] R. M. Sari, A. Rizka, N. A. Putri, and A. Efriana, *Perhitungan Metode Clustering*. Payakumbuh: Serasi Media Teknologi, 2024.

[14] I. A. Firdaus *et al.*, *Strategi Pengembangan Kota Ramah Disabilitas*. Surabaya: Cipta Media Nusantara, 2022.

[15] P. A. Gatto and R. M. Awangga, *Pengelompokan Kedisiplinan Pegawai Berdasarkan Absensi Menggunakan Algoritma K-Means*. Penerbit Buku Pedia, 2023.

[16] Nurhayati Nurhayati, *Pemodelan K- Means Algoritma Dan Big Data Analysis (Pemetaan Data Mustahiq)*, Cetakan I. Pascal Books, 2022.

[17] Kuntoro, *Data Science dalam Perspektif Statistika*. Airlangga University Press, 2024, 2024.

[18] I. F. Fauzi, M. G. Resmi, and T. I. Hermanto, "Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit," *Jurnal Masyarakat Informatika Unjani*, vol. 7, no. 2, pp. 1–15, 2023.