

## Analisis *Clustering* Data Penyandang Disabilitas Menggunakan Metode *Agglomerative Hierarchical Clustering* dan *K-means*

Alun Sujjada<sup>1</sup>, Gina Purnama Insany<sup>2</sup>, Silvia Noer<sup>3</sup>

University Nusa Putra, Indonesia

### Article Info

#### Article History

Received: 17-07-2023

Revises : 12-10-2023

Accepted: 05-12-2023

#### Keywords

Penyandang disabilitas;  
*Agglomerative Hierarchical Clustering*;  
*K-means*;  
EDA (*Exploratory Data Analysis*);  
DBI *Davies Bouldin-Index*

#### ✉ Corresponding Author

**Alun Sujjada**

Universitas Nusa Putra

Tel. +62 818-0504-0468

[alun.sujjada@nusaputra.ac.id](mailto:alun.sujjada@nusaputra.ac.id)

### ABSTRACT

*Disability issues are still a major concern in society due to the discrimination often faced by people with disabilities. Many of them have abilities that are equal to individuals without physical limitations. Through this case study, this research aims to Cluster disability data by considering three types of disabilities: physical, visual, and hearing, and hearing and speech using agglomerative hierarchical Clustering and k-means methods. This research was conducted by analyzing data from people with disabilities in 7 provinces in Indonesia. K-means to group data and agglomerative hierarchical Clustering as a centroid determinant in k-means. to enrich the results of data analysis, the EDA (*Exploratory Data Analysis*) process is used to identify outliers and anomalies. The results of the data analysis show that there are three main Clusters. The first Cluster has a high level of disability and includes 62 cities and districts, the second Cluster has a medium level of disability with 37 cities and districts, and the third Cluster has a low level of disability with 27 cities and districts. The best evaluation using the Davies Bouldin Index method resulted in two Clusters, indicating a better quality of Cluster division. The results of this study provide a better understanding of the distribution of disability in Indonesia, which can be used as a foundation to improve inclusion and accessibility for people with disabilities. Further recommendations can be made based on these findings to improve their situation in terms of employment and education.*

### PENDAHULUAN

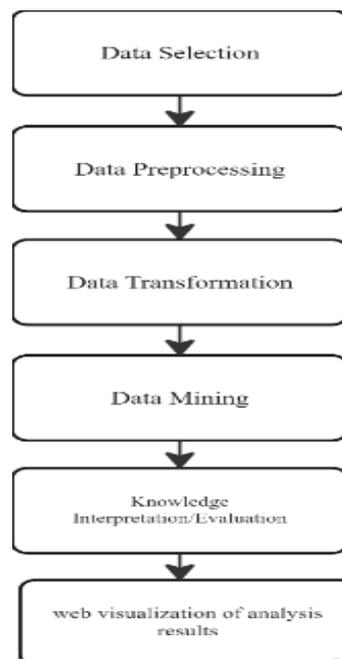
Masalah disabilitas masih menjadi perhatian utama dalam masyarakat, terutama terkait dengan diskriminasi yang dialami oleh penyandang disabilitas sehari-hari. Meskipun hukum di Indonesia (Undang-Undang No. 8 Tahun 2016) telah memberikan perlindungan, masih ada hambatan dalam mewujudkan hak-hak mereka. Penelitian ini bertujuan untuk menganalisis tingkat disabilitas dengan mengambil *sample* data di 7 provinsi Indonesia, termasuk Jawa Barat, Jawa Tengah, Jawa Timur, Bali, Jambi, Bengkulu, dan D.I. Yogyakarta. Penelitian ini berfokus pada tiga variabel utama: jumlah disabilitas dengan cacat fisik, cacat netra atau buta, cacat rungu, dan wicara [1]. Terdapat pola *Clustering* yang dapat diidentifikasi dalam tingkat disabilitas di provinsi-provinsi untuk dijadikan sebagai sampel. Peneliti mengantisipasi bahwa provinsi-provinsi tertentu mungkin memiliki tingkat disabilitas yang lebih tinggi atau lebih rendah dalam kategori-kategori yang disebutkan. Selain itu, terdapat kemungkinan bahwa pola *Clustering* dapat memberikan wawasan yang berguna dalam merancang kebijakan yang lebih baik untuk mendukung hak-hak penyandang disabilitas, terutama dalam aspek akses pendidikan dan pekerjaan.

Melalui analisis *Clustering* menggunakan algoritma *K-means* dan *agglomerative hierarchical Clustering*, serta evaluasi dengan *Davies–Bouldin Index*[2], penelitian ini berusaha memberikan wawasan yang berguna untuk merancang kebijakan yang lebih baik dalam memenuhi hak-hak penyandang disabilitas. Masih ada upaya yang harus dilakukan Indonesia untuk memenuhi hak-hak penyandang disabilitas, terutama dalam dimensi eksistensi, politik, ekonomi, masyarakat, budaya, dan layanan publik dengan meningkatkan pemahaman tentang masalah ini. Diharapkan dapat tercipta akses yang lebih baik untuk pendidikan dan pekerjaan yang layak bagi penyandang disabilitas.

Selain itu, perlu ditekankan bahwa jumlah penyandang disabilitas di Indonesia masih terus meningkat, dan hambatan terhadap hak-hak mereka perlu ditangani dengan serius[3]. Dengan bantuan dari analisis data, diharapkan penelitian ini dapat memberikan kontribusi yang lebih besar untuk kesadaran dan tindakan yang lebih baik dalam mendukung penyandang disabilitas di negara ini.

## METODE

### Tahapan Penelitian



**Gambar 1.** Tahapan Penelitian

#### a. *Data Selection*

Pada tahap ini pengumpulan data dilakukan menggunakan Data sekunder dari beberapa alamat *website* pemerintahan.

#### b. *Preprocessing Data*

Pada tahap ini yang dilakukan adalah proses pembersihan data, termasuk penghapusan data duplikat, *missing value*, serta memperbaiki kesalahan data. Imputasi data akan digunakan untuk menangani *missing value*.

#### c. *Data Transformation*

Pada tahap ini untuk data *transformation* dilakukan untuk menghilangkan *outlier*. Peneliti meningkatkan kualitas dataset dan membantu pelatihan model dengan metode pengembangan sistem *EDA (Exploratory Data Analysis)* agar pemrosesan data yang memadai menghasilkan prediksi yang lebih akurat.

#### d. *Data Mining*

Tahapan ini merupakan tahapan untuk mengaplikasikan metode *agglomerative hierarchical Clustering* untuk memilih *centroid* pada *Cluster* metode *K-means*.

e. *Knowledge Interpretation/Evaluation*

Pada tahap ini, hasil data mining diinterpretasikan ke dalam gambar atau tabel dan dievaluasi untuk mengetahui kualitas dan kesesuaian kelompok dengan data *Cluster* dengan menggunakan metode perbandingan *elbow* dan *Davies-Bouldin* indeks.

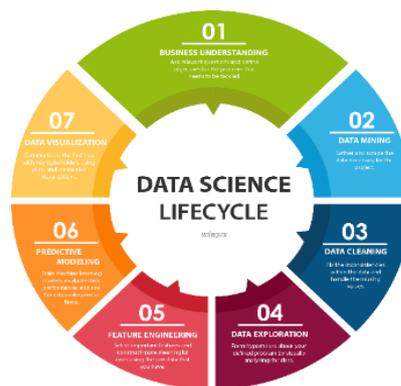
f. *Web Visualization of analys result*

Memvisualisasikan hasil klasifikasi data penyandang disabilitas menggunakan sistem berbasis *web* yang dibangun dengan *framework streamlit*.

**Metode Pengumpulan Data**

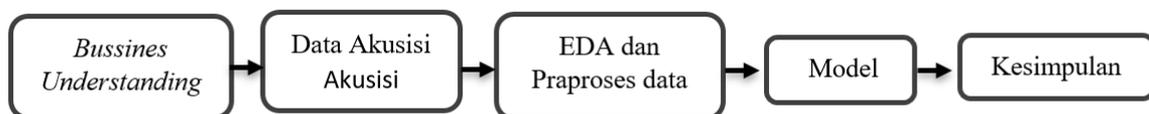
Peneliti mengumpulkan dan menganalisis data dari berbagai sumber tertulis, seperti artikel, jurnal, dan laporan penelitian. Selain itu peneliti juga mengumpulkan data tentang disabilitas dari tujuh provinsi di Indonesia tanpa menggunakan observasi atau wawancara. Berbagai sumber tertulis lainnya termasuk *website* statistik, *website* open data, dan artikel. Hal ini dilakukan untuk menyampaikan perkembangan terkini dalam pemahaman dan perawatan penyandang disabilitas.

**Metode Pengembangan Sistem**



Gambar 2. EDA (*Exploratory Data Analysis*)

Pada penelitian ini peneliti menerapkan pendekatan dalam penelitian dengan pengembangan sistem EDA (*Exploratory Data Analysis*). *Exploratory Data Analysis* merupakan eksplorasi data analisis yang digunakan untuk menganalisis data menggunakan berbagai teknik khususnya secara grafis. Tujuannya adalah untuk mendeteksi *outlier* dan *anomaly* data, mengenali struktur data dasar, menguji asumsi data. Pada Gambar 3 menunjukkan keseluruhan tahap akan diimplementasikan dengan mempergunakan bahasa *Python*. [4]



Gambar 3. Metode Pengembangan Sistem

***Business Understanding***

Tahapan ini berfokus pada pemahaman mendalam tentang data yang akan dianalisis melalui tahapan *Understanding Data* dalam EDA [5]. Tujuan utama dari tahapan ini adalah untuk mendapatkan pemahaman yang lebih baik tentang data yang akan dianalisis, serta mempersiapkan dasar untuk langkah-langkah analisis selanjutnya. *Business Understanding* pada tahapan penelitian ini, mengarah pada pemahaman yang lebih baik tentang data penyandang disabilitas serta untuk memperoleh informasi yang berguna dalam mengidentifikasi pengelompokan jumlah penyandang disabilitas di tujuh provinsi di Indonesia. Identifikasi jumlah penyandang disabilitas dikelompokkan berdasarkan 137 kota dan kabupaten dengan 3 jenis variabel disabilitas, yaitu cacat fisik, cacat netra atau buta, cacat

rungu dan wicara. Pengelompokan data penyandang disabilitas menggunakan analisis *Clustering* algoritma *K-means* dan *agglomerative hierarchical Clustering*.

### Data Akuisisi

Data Akuisisi [6] merupakan alat atau prosedur yang digunakan untuk mengumpulkan informasi yang digunakan untuk memahami dan mencatat karakteristik, kebutuhan, dan pengalaman penyandang disabilitas secara akurat dan menyeluruh.

### EDA dan Pra Proses Data

Pada tahap ini dilakukan eksplorasi data dengan menggunakan fungsi statistik, matematik, dan divisualisasikan dalam bentuk grafik. Hal ini bertujuan untuk mempermudah pemahaman terhadap data dan pola dasar data [7].

### Model

Untuk membangun model dalam studi kasus pengelompokan penyandang disabilitas menggunakan metode EDA, berikut merupakan tahapan yang harus dilakukan:

- Pemilihan Dataset.
- *Transformasi* Dataset.
- Verifikasi Dataset.
- Pemilihan Alat Visualisasi atau Pemodelan.
- Analisis Visualisasi atau Pemodelan.

### Hasil Analisis Visualisasi

Pada tahapan ini yang dilakukan adalah mempresentasikan hasil temuan yang relevan untuk mendapatkan pemahaman yang lebih baik tentang pengelompokan data penyandang disabilitas dan informasi yang dihasilkan.

## HASIL DAN PEMBAHASAN

Penelitian ini menyimpulkan hasil terhadap penelitian sebelumnya yang telah mengidentifikasi pola-pola dalam studi kasus dan metode *clustering* terkait penyandang disabilitas. Salah satu teknik yang efektif dalam proses ini terletak pada penggunaan berbagai teknik pengolahan data salah satunya dengan menggunakan pendekatan EDA (*Explory Data Analys*). Pendekatan tersebut mampu menghilangkan *outlier* dan anomali, untuk menentukan jumlah *Cluster* yang optimal. Melalui perbandingan dengan penelitian sebelumnya, temuan penelitian ini menekankan pentingnya mengoptimalkan jumlah Cluster. Penemuan ini juga membantu memahami pola data penyandang disabilitas, yang berdampak pada pembuatan kebijakan yang lebih baik yang mendukung hak-hak mereka, terutama yang berkaitan dengan akses ke pendidikan dan kesempatan kerja.

### Analisis Masalah

Big Data muncul karena menggunakan data yang sangat besar sebagai informasi semakin sulit. Akibatnya, menjadi lebih mudah untuk menghasilkan manfaat bagi manusia.[8]

### Data Akuisisi

Sampel data yang digunakan adalah data set penyandang disabilitas di 7 Provinsi di Indonesia. Dalam data set tersebut terdiri dari 54 baris dan 6 kolom yang terdiri dari variable Cacat fisik, Cacat Netra atau Buta dan Cacat Rungu atau Wicara.

### EDA & Praproses

#### a. *Import Library*

Dalam mengidentifikasi *outlier* dan *anomali* dengan metode EDA [9] dibutuhkan *Library* yang telah diinstal ke dalam program. *Library* ini berisikan kumpulan fungsi-fungsi yang digunakan untuk menjalankan perintah yang lebih detail. *Source code* yang digunakan untuk memanggil beberapa package.

#### b. *Import Data*

Proses selanjutnya adalah *input* data dari file excel ke program *Google Collab*. Tipe data yang digunakan adalah .xlsx.

**c. Missing value imputation**

*Missing value Imputation* adalah proses menambahkan nilai yang hilang atau tidak cukup dalam dataset kemudian digantikan dengan nilai yang memadai atau tervalidasi.

```
#Missing value imputation
data.isna().sum()

NAMA_KOTA_DAN_KABUPATEN    0
CACAT_FISIK                 0
CACAT_NETRA_ATAU_BUTA      0
CACAT_RUNGU_ATAU_WICARA    0
dtype: int64
```

**Gambar 4.** *Missing Value Imputation*

Pada gambar diatas Jumlah *Missing value* pada tiap variabel bernilai 0 atau tidak ada nilai yang hilang di setiap kolom, yang perlu diperhatikan adalah `isna()` digunakan untuk mendeteksi *missing value* pada data.

**d. Descriptive Statistics**

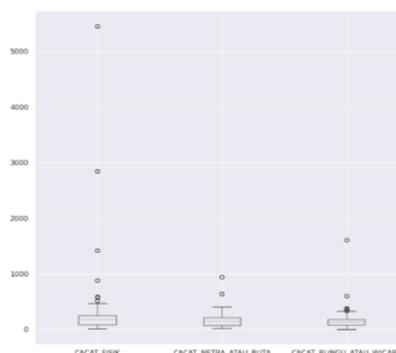
*Descriptive Statistics* adalah prosedur untuk mendeskripsikan, mengilustrasikan, dan meringkas data dalam bentuk numerik atau grafik sehingga pengguna dapat memahami karakteristik mendasar dari data yang diamati terutama dalam mendeteksi *outlier* dan anomali pada data (Gambar 5).

Seperti yang ditunjukkan pada Gambar 6 hasil mendeteksi *outlier*. Jumlah maksimum cacat Fisik adalah 5455, sedangkan jumlah rata-ratanya adalah 254.029. Kemudian dari nilai min, 25%, 50%, 75% terdapat beberapa kolom yang terindikasi mengandung *outlier*, begitu halnya dengan variabel yang lain. Untuk lebih memudahkan dalam mendeteksi *outlier* adalah mempergunakan visualisasi *box-plot* seperti yang ditunjukkan pada Gambar 6.

```
#Descriptive Statistics
data.describe(include='all')
```

	NAMA_KOTA_DAN_KABUPATEN	CACAT_FISIK	CACAT_NETRA_ATAU_BUTA	CACAT_RUNGU_ATAU_WICARA
count	135	135.000000	135.000000	135.000000
unique	135	NaN	NaN	NaN
top	KABUPATEN BOGOR	NaN	NaN	NaN
freq	1	NaN	NaN	NaN
mean	NaN	254.029630	151.740741	148.059259
std	NaN	533.188812	116.831257	154.958338
min	NaN	10.000000	12.000000	3.000000
25%	NaN	85.500000	68.500000	76.000000
50%	NaN	181.000000	144.000000	132.000000
75%	NaN	253.000000	210.000000	177.500000
max	NaN	5455.000000	943.000000	1604.000000

**Gambar 5.** *Descriptive Statistic*



**Gambar 6.** Visualisasi *Box-Plot*

**e. Menghilangkan outlier menggunakan metode statistic**

Data pada penelitian ini tidak mengikuti distribusi normal, sehingga penulis akan menghitung titik data outlier menggunakan metode statistik yang disebut rentang interkuartil (IQR). Dengan menggunakan IQR, poin data outlier adalah yang berada di bawah  $Q1 - 1.5 IQR$  atau di atas  $Q3 + 1.5 IQR$ . Q1 merupakan persentil ke-25 dan Q3 merupakan persentil ke-75 dari kumpulan data, dan IQR mewakili rentang interkuartil yang dihitung dengan Q3 dikurangi Q1 ( $Q3 - Q1$ ).

Langkah pertama yaitu menghitung IQR untuk masing-masing kolom seperti pada gambar 7 di bawah ini. Peneliti dapat menilai penyebaran data dalam kolom dan melihat kemungkinan outlier dengan menerapkan aturan dalam IQR. Outlier biasanya disebut sebagai data yang berada di luar rentang  $Q1 - 1.5 * IQR$  hingga  $Q3 + 1.5 * IQR$ .

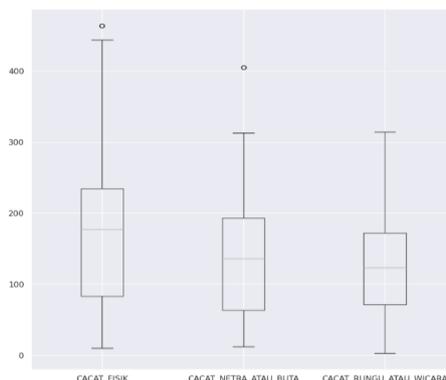
Dengan menghitung batas atas dan bawah untuk outlier, peneliti mendeteksi nilai yang berada di luar atas dan bawah ini sebagai kemungkinan outlier di setiap kolom. Sehingga dalam konteks ini, outlier adalah nilai yang berada di luar rentang yang ditetapkan oleh batas bawah dan atas yang dihitung menggunakan rumus.

Hasil outlier peneliti visualisasikan menggunakan format excel pada Gambar 7.

No	NAMA_KOTA_DAN_KABUPATEN	CACAT_FISIK	CACAT_NETRA_ATAU_BUTA	CACAT_RUNGU_ATAU_WICARA
3	Kab. Sukabumi	201	111	125
6	Kab. Garut	358	155	224
17	Kab. Bekasi	214	155	166
123	Kota Denpasar	383	278	225
124	Kab. Badung	268	310	255
127	Kab. Bangli	260	218	114
128	Kab. Klungkung	392	139	176
131	Kab. Jemberana	464	405	178
133	BANTUL	371	194	355

Gambar 7. Data Outlier

Hasil akhir dalam menggunakan metode EDA, jumlah data yang terhapus adalah 9 dan jumlah data tanpa outlier 126 data dari keseluruhan data 135, peneliti menyajikan data tanpa outlier sekaligus juga menghapus data outlier menggunakan Source code yang ditunjukkan pada Gambar 7. hasil visual menggunakan boxplot kolom tertentu dari Data Frame data no outliers.



Gambar 8. Visual Menggunakan Box Plot tanpa Outlier

**Clustering**

Tahap Clustering pada penelitian ini menggunakan algoritma Agglomerative Hierarchical Clustering dan K-means.

**a. Agglomerative Hierarchical Clustering**

Metode *Agglomerative Hierarchical Clustering* [10] dari prosesnya membuat *Cluster* dengan berbagai tingkat pemisahan antara jarak yang dikelompokkan menggunakan metode “*complete linkage*”. Sementara metode pendekatan menggunakan “*average*” untuk *Average Linkage*, dan “*Single*” untuk *single linkage*.

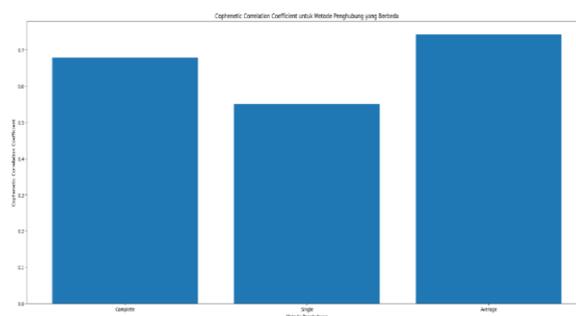
Untuk mengevaluasi seberapa baik *dendrogram* pengelompokan hierarki, sesuai dengan matriks jarak asli antara titik data. *Cophenetic correlation coefficient* (CCC) dapat ditentukan dengan berbagai metode penghubungan dalam *hierarchical Clustering*. Berikut hasil evaluasi ditunjukkan pada gambar Tabel 1.

**Tabel 1.** Nilai Evaluasi *Cophenetic Correlation Coefficient*

Metode	Nilai CCC
<i>Complete Linkage</i>	0.6793315507486207
<i>Single Linkage</i>	0.5509368333077486
<i>Average Linkage</i>	0.7424539173686395

Pendekatan *average linkage* yang digunakan untuk membandingkan ketiga cara tersebut menghasilkan nilai korelasi *cophenetic* paling besar (mendekati angka satu), dengan nilai korelasi sebesar 0,07424539173686395. Hal ini menunjukkan bahwa metode optimum untuk penelitian ini adalah metode *average linkage*. Hasil visual evaluasi *Cophenetic correlation coefficient* dengan menggunakan *Plot Bar* ditunjukkan pada Gambar 9.

Menghitung nilai rata-rata anggota *Cluster* dengan hasil metode evaluasi dari *Cophenetic correlation coefficient* yaitu dengan metode *Average linkage*. Jumlah *cluster* yang ditentukan adalah 3 *cluster*. Nilai *centroid* dari *clustering* dengan hasil *Average linkage* adalah sebagai berikut:



**Gambar 9.** *Cophenetic correlation coefficient*

**Tabel 2.** Nilai *centroid* dari *Clustering* dengan hasil *Average linkage*

<i>Cluster</i>	Cacat fisik	Cacat Netra atau buta	Cacat rungu atau wicara
<i>Cluster 1</i>	376	94	112
<i>Cluster 2</i>	161	102	159
<i>Cluster 3</i>	201	111	125

### b. *K-means Clustering*

*Cluster* dibentuk menggunakan algoritma *K-means* [11] dengan mengidentifikasi titik pusat. Pendekatan pengelompokan *agglomerative hierarchical Clustering* yang telah diselesaikan sebelumnya adalah untuk menentukan *centroid* awal. Setelah data *Cluster* menjadi statis, iterasi dihentikan. Oleh karena itu, tidak diperlukan lagi iterasi dari algoritma pengelompokan *K-means* (Tabel 3).

**Tabel 3.** Hasil *Cluster K-Means* dan *Agglomerative* dengan Metode *Average Linkage*

No	Nama_Kota_Kabupaten	Cacat_Fisik	Cacat_Netra_Atau_Buta	Cacat_Rungu_Atau_Wicara	Jarak Terdekat	Cluster
1	Kabupaten Bogor	444	189	223	11948	2
2	Kabupaten Sukabumi	201	111	125	0	1
3	Kabupaten Cianjur	376	94	112	0	1
4	Kabupaten Garut	358	155	224	7231	2
5	Kabupaten Tasikmalaya	266	117	152	379	2
6	Kabupaten Kuningan	161	102	159	0	2
7	Kabupaten Cirebon	89	32	67	6156	1
8	Kabupaten Majalengka	266	79	158	635	2
9	Kabupaten Sumedang	185	99	173	229	2
10	Kabupaten Indramayu	311	147	265	13411	2
11	Kabupaten Subang	83	55	60	4518	1
12	Kabupaten Purwakarta	84	50	67	4253	1
13	Kabupaten Karawang	220	105	122	64	1
14	Kabupaten Bekasi	214	155	166	2911	2

15	Kabupaten Bandung Barat	91	39	53	6791	1
16	Kota Bogor	10	76	68	2626	1
17	Kota Sukabumi	52	22	36	11284	1
18	Kota Bandung	420	237	314	42509	2
19	Kota Cirebon	88	45	82	3589	1
20	Kota Bekasi	177	121	206	2586	2
21	Kota Depok	239	78	201	2418	2
22	Kota Cimahi	284	103	141	403	3
23	Kota Tasikmalaya	62	91	105	372	1
24	Kota Banjar	53	64	139	1952	2

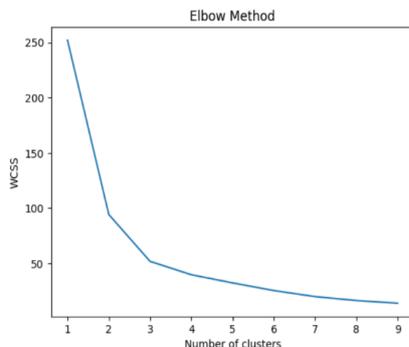
Jumlah data pada masing masing *Cluster* ditunjukkan pada tabel 4 di bawah ini. Langkah selanjutnya melakukan keakuratan hasil perhitungan *K-means* dan *agglomerative hierarchical Clustering*.

**Tabel 4.** Jumlah *Cluster*

<i>Cluster</i>	Jumlah
<i>Cluster 1</i>	62
<i>Cluster 2</i>	37
<i>Cluster 3</i>	27

### c. Evaluasi dengan Metode *Elbow*

Metode *Elbow* [12] merupakan cara untuk menentukan jumlah *cluster* yang tepat untuk algoritma pengelompokan *K-means*. Caranya adalah dengan menemukan titik "sudut" pada grafik sumbu x (jumlah *Cluster*) vs sumbu y (inersia atau SSE - *Sum of Squared Errors*). Gambar 10 menunjukkan hasil dari analisis jumlah *Cluster* menggunakan metode *elbow* [13] yaitu (jumlah *Cluster*) vs sumbu y (inersia atau SSE - *Sum of Squared Errors*). Gambar 10 menunjukkan hasil dari analisis jumlah *Cluster* menggunakan metode *elbow*. Nilai inersia menurun dengan bertambahnya jumlah *Cluster*. Pada jumlah *Cluster* tertentu, penurunan nilai inersia mulai melambat. Gambar 10 menunjukkan bahwa penurunan nilai inersia berhenti setelah enam cluster.



Gambar 10. Evaluasi hasil *Elbow*

Berdasarkan analisis hasil *elbow* tersebut, tiga atau empat *Cluster* tampaknya memiliki jumlah yang ideal dengan nilai inersia yang masih sangat rendah. Nilai inersia untuk setiap *Cluster* ditunjukkan dalam Tabel 5:

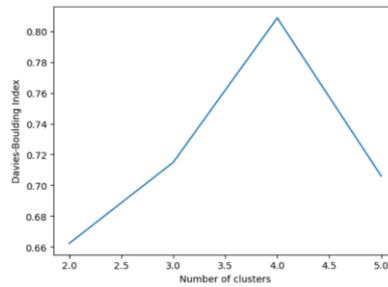
Tabel 5. Nilai Inertia setiap *Cluster*

Jumlah <i>Cluster</i>	Nilai Inertia
1	252.00000000000006
2	94.18280274837247
3	51.76056497679471
4	39.64427779721061
5	32.2315534631135
6	25.30593893010806
7	19.81613617543217
8	16.25059225886231
9	13.827321749229423

Hasil analisis evaluasi dengan menggunakan grafik *elbow* tidak menunjukkan adanya titik siku yang jelas. Ketika penurunan inersia tidak bervariasi secara signifikan saat jumlah *Cluster* bertambah maka hasil evaluasi menggunakan metode *elbow* tidak menghasilkan solusi yang jelas. Untuk menentukan jumlah *Cluster*, peneliti menggunakan metode evaluasi *Davies Bouldin Index* (DBI), untuk memastikan jumlah *Cluster*.

**D. Evaluasi *Cluster* dengan *Davies Bouldin Index* (DBI)**

Dalam penentuan jumlah *Cluster* dengan menggunakan nilai *Davies-Bouldin Index*[14] hasil yang diberikan berupa nilai untuk setiap jumlah *Cluster* yang diuji. Komponen dari setiap hasil terdiri dari jumlah *Cluster* yang diuji dan nilai indeks *Davies-Bouldin*. Hasil evaluasi dengan menggunakan indeks *Davies-Bouldin* ditunjukkan pada Gambar 11 berikut:



**Gambar 11.** Hasil *cluster* dengan metode DBI

Jumlah nilai *Davies Bouldin Index* untuk setiap *Cluster* adalah sebagai berikut:

**Tabel 6.** Jumlah *Cluster* dan nilai DBI

Jumlah <i>Cluster</i>	Jumlah Nilai DBI
2	0.6622695068270446
3	0.7149135734754731
4	0.8087492384948665
5	0.7060136965689485

Dalam temuan ini, hasil evaluasi menunjukkan bahwa penggunaan dua *Cluster* merupakan solusi yang optimal. Hal ini didasarkan pada fakta bahwa jumlah *Cluster* dua memiliki nilai DBI [15] terendah, yang mengindikasikan pembagian *Cluster* yang paling baik dalam analisis data yang dilakukan. Dalam situasi ini, penggunaan dua *Cluster* dianggap lebih baik daripada pendekatan lain yang memerlukan jumlah *Cluster* yang lebih besar. Penggunaan dua *Cluster* ini dapat memiliki konsekuensi praktis yang signifikan. Pemangku kepentingan dan pengambil keputusan dapat mengambil tindakan dengan mengetahui bahwa *cluster* yang dihasilkan dari analisis data memiliki kualitas terbaik.

#### DAFTAR PUSTAKA

- [1] Basysyar Fadhil, M. (2021). *Clustering Data Disabilitas Menggunakan Algoritma K-Means Di Kabupaten Cirebon*. STMIK GICI. vol. 9.
- [2] Muningsih, E. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains dan Manajemen*. vol. 9, no. 1.
- [3] Akbar, M. P. (2023). Kesetaraan Akses Bagi Penyandang Disabilitas. Accessed: Jun. 05.
- [4] Enterprise, J. (2019). *Phyton untuk Programmer Pemula*.
- [5] Setiawan Suparno, D. (2021). Target Pasar Menggunakan Metode EDA, K-Means, Hierarchial Clustering, Confusion Matrix..
- [6] Najwa, N. F. (2021). Akuisisi Data Media Sosial Pemerintah Untuk Menganalisis Keterbukaan Informasi Penyebaran Covid-19. *Jurnal Sioteknologi*. vol. 20, no. 1. 46–55, Apr.
- [7] Wahyuni, E. D. (2019). Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining. pp. 263–269.
- [8] Radhi, M. (2021). Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook. *Jurnal Sistem Informasi dan Ilmu Komputer Prima*. vol. 4, no. 2.

- [9] Ilmu, J., Sosial, K. (2019). Penyandang Disabilitas Di Indonesia: Perkembangan Istilah Dan Definisi.
- [10] Wahyudi, N. (2021). Komparasi Algoritma K-Means, K-Medoid, Agglomerative Clustering Terhadap Genre Spotify. vol. 7, no. 1.
- [11] Wahyuni D. (2019). Analisa Clustering Pada Data Pelanggaran Lalulintas Di Pengadilan Negeri Dumai Dengan Menggunakan Metode K-Means.
- [12] Arientawati. (2023). Analisis Pengelompokan Gangguan TIK Pada Sistem Pencatatan Layanan Menggunakan Algoritma K-Means dan Metode Elbow. Techno.COM. vol. 22, May.
- [13] Dewi Dewa, A. I. C. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. JURNAL MATRIX. vol. 9, Nov.
- [14] Mulyono, U. W. (2020). Klasterisasi Perkara Pelanggaran Lalu Lintas Menggunakan Algoritma K-Means Dan Davies-Bouldin Index. Vol. 5.
- [15] Septiani, I. W. (2022). Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. Jurnal Sistem Komputer dan Informatika (JSON), vol. 3, no. 4, p. 556, Jul.