

## Pengelompokan Wilayah Indonesia Berdasarkan Komponen Indeks Pembangunan Manusia dengan Pendekatan Algoritma K-Means Clustering

Nur Rukhan Saputra<sup>1\*</sup>, Ghuftron Zaida Muflih<sup>2</sup>

<sup>1,2</sup>Teknik, Teknik Informatika, Universitas Ma'arif Nahdlatul Ulama, Kebumen, Indonesia

E-mail: <sup>1\*</sup>nurrukhang@gmail.com, <sup>2</sup>ghuftron.zaida@umnu.ac.id

(\* : corresponding author)

### Abstrak

Indeks Pembangunan Manusia (IPM) adalah alat penting untuk mengukur kesejahteraan masyarakat secara holistik. Penelitian ini bertujuan mengelompokkan kabupaten/kota di Indonesia ke dalam tiga kluster pembangunan manusia (rendah, sedang, tinggi) menggunakan algoritma K-Means Clustering berdasarkan indikator pendidikan, kesehatan, dan standar hidup, meliputi rata-rata lama sekolah, angka harapan hidup, pendapatan bulanan rata-rata, dan pengeluaran per kapita, dengan data dari Badan Pusat Statistik (BPS) tahun 2023. Pendekatan klusterisasi mempermudah analisis data IPM yang kompleks, memberikan gambaran pola disparitas antar wilayah yang lebih jelas. Untuk mengevaluasi kualitas kluster, digunakan Davies-Bouldin Index (DBI), yang mengukur sejauh mana kluster saling terpisah dan memiliki keterhubungan dalam kluster, dengan nilai DBI yang lebih kecil menunjukkan pemisahan kluster yang lebih baik. Penelitian ini menghasilkan nilai DBI sebesar 1,17, yang menunjukkan pemisahan kluster yang baik. Hasil pengelompokan ini menyediakan wawasan awal yang bermanfaat bagi pembuat kebijakan dalam merancang langkah pembangunan yang lebih terarah dan sesuai dengan kebutuhan masing-masing kluster, sehingga dapat mendukung langkah strategis untuk mengatasi disparitas pembangunan di Indonesia.

**Kata kunci:** clustering, indeks pembangunan manusia, IPM, k-means

### Abstract

*The Human Development Index (HDI) is an essential tool for measuring the welfare of a society holistically. This study aims to group districts/cities in Indonesia into three human development clusters (low, medium, high) using the K-Means Clustering algorithm based on indicators of education, health, and living standards, including average years of schooling, life expectancy, average monthly income, and per capita expenditure, with data sourced from the Central Statistics Agency (BPS) in 2023. The clustering approach simplifies the analysis of complex HDI data, providing a clearer overview of disparity patterns across regions. To evaluate cluster quality, the Davies-Bouldin Index (DBI) is used, measuring the extent of separation between clusters and cohesion within clusters, with smaller DBI values indicating better cluster separation. This study produced a DBI value of 1.17, indicating good cluster separation. The clustering results offer valuable insights for policymakers to design more targeted and focused development strategies based on the needs of each cluster, thereby supporting strategic steps to address development disparities in Indonesia.*

**Keywords:** clustering, indeks pembangunan manusia, IPM, k-means

## 1. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) adalah instrumen penting untuk mengevaluasi kesejahteraan individu di suatu wilayah. Indeks Pembangunan Manusia (IPM) adalah metrik komposit yang menilai dan membandingkan tingkat pembangunan manusia di berbagai negara. Program Pembangunan Perserikatan Bangsa-Bangsa (UNDP) memperkenalkan Indeks Pembangunan Manusia (IPM) pada tahun 1990 untuk memberikan pandangan yang lebih menyeluruh tentang kesejahteraan manusia di luar indikator ekonomi seperti PDB. Elemen-elemen utama IPM dinilai berdasarkan tiga kriteria. Secara khusus, kesehatan, pendidikan, dan kondisi kehidupan saling mempengaruhi satu sama lain.

Komponen kesehatan sangat penting dalam Indeks Pembangunan Manusia (IPM) [4]. Faktor penentu kesehatan mempengaruhi angka harapan hidup di suatu daerah. Pendidikan merupakan komponen penting dalam Indeks Pembangunan Manusia (IPM). Indikator ini menunjukkan pencapaian pendidikan suatu wilayah. Dimensi pendidikan tercermin dari angka

harapan lama sekolah dan rata-rata lama sekolah. Aspek terakhir adalah tingkat kehidupan, yang secara signifikan memengaruhi IPM. Standar hidup dinilai dari pengeluaran per kapita dan upah minimum di tingkat kabupaten atau kota, dengan pengeluaran per kapita menunjukkan daya beli individu di suatu wilayah [5].

Indeks Pembangunan Manusia (IPM) digunakan sebagai alat untuk menilai capaian pembangunan yang fokus pada kesejahteraan masyarakat. IPM mengukur kinerja pemerintah serta kemampuan masyarakat dalam mengakses pendapatan, pendidikan, dan kesehatan. IPM dipilih karena mencakup tiga dimensi utama yang saling berkaitan, yaitu kesehatan (diukur melalui angka harapan hidup laki-laki dan perempuan), pendidikan (diukur melalui rata-rata lama sekolah dan harapan lama sekolah), serta daya beli (diukur melalui pengeluaran per kapita). Ketiga dimensi ini memberikan gambaran komprehensif tentang pembangunan masyarakat di suatu wilayah. Namun, IPM tidak mencakup secara spesifik aspek lain seperti kepadatan penduduk, tingkat kemiskinan, atau tingkat pengangguran. Meskipun demikian, IPM tetap menjadi indikator yang efektif untuk menilai kemajuan pembangunan masyarakat karena fokusnya pada tiga dimensi dasar yang relevan [6].

*Clustering* adalah cara *machine learning* yang dipakai guna pengelompokan data ke dalam beberapa kelompok berdasarkan kesamaannya [7]. Memungkinkan data yang terdapat pada kelompok pertama memiliki karakter yang memiliki perbedaan dibandingkan dengan data di kelompok lain. Terdapat banyak jenis dari metode *clustering* seperti *partitional clustering* dan *hierarchical clustering*. Pada jenis *partitional clustering* terdapat teknik *k-means clustering*, *k-medoids* dan *fuzzy-c-means*. Sedangkan pada jenis *hierarchical clustering* terdapat teknik *divisif* dan *agglomerative* [8].

*K-Means* adalah suatu pendekatan klusterisasi data yang bersifat *partitional cluster* dan berusaha untuk membagi data yang tersedia menjadi satu atau lebih [9]. Beberapa karakteristik dari *K-Means* seperti sangat cepat dalam pemrosesan kluster, memungkinkan suatu kluster tidak memiliki anggota, hasil bersifat unik dan sangat sensitif dalam pemilihan centroid awal secara acak [10]. Algoritma *k-means* beroperasi dengan menentukan jumlah kluster terlebih dahulu, lalu menyebarkan data secara acak, menetapkan centroid atau rata-rata data dalam setiap kluster, dan mengelompokkan data berdasarkan centroid terdekat. Proses ini diulang dengan menghitung ulang centroid hingga data dikelompokkan pada centroid terdekat. Langkah-langkah tersebut dilakukan terus menerus hingga tidak ada data yang berpindah kluster atau perubahan pada centroid tidak lagi signifikan [11].

Penelitian terkait penggunaan *K-Means Clustering* diantaranya penelitian yang dilakukan oleh [12], menemukan tiga kluster motif pengunjung Festival JHF#2 menggunakan *K-Means Clustering*. Penelitian oleh [13] mengelompokkan item berdasarkan karakteristik yang serupa antar data. Penelitian oleh [14] menerapkan metode *K-Means Clustering* untuk melakukan analisa perilaku masyarakat dalam membuang sampah. Penelitian oleh [15] menunjukkan bahwa metode *K-Means Clustering* efektif untuk melakukan analisa data cuaca di Kabupaten Cilacap. Penelitian oleh [16] didapatkan bahwa *K-Means Clustering* dapat mengelompokkan UMKM di kota Palembang menjadi dua kluster yakni kluster dengan pertumbuhan tinggi dan kluster dengan pertumbuhan stabil. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada penggunaan Indeks Pembangunan Manusia dari seluruh kabupaten/kota di Indonesia yang dijadikan sebagai dataset penelitian.

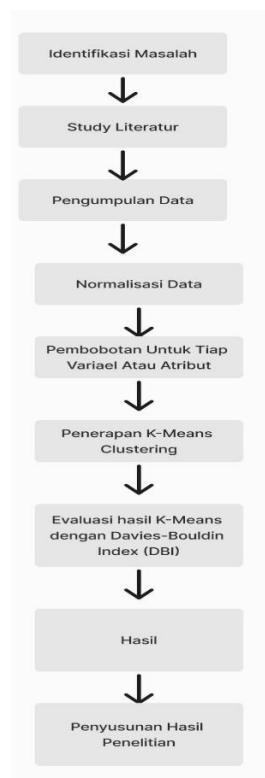
Penelitian ini bertujuan untuk mengelompokkan kabupaten dan kota di Indonesia ke dalam tiga kluster pembangunan manusia (rendah, sedang, tinggi) berdasarkan indikator utama Indeks Pembangunan Manusia (IPM), yaitu pendidikan, kesehatan, dan kondisi hidup layak. Pengelompokan wilayah dilakukan dengan metodologi pengelompokan *K-Means*. Klasifikasi ini sangat penting untuk menjelaskan tren pembangunan di Indonesia, karena data Indeks Pembangunan Manusia (IPM) dari Badan Pusat Statistik (BPS) sering kali hanya berupa angka-angka yang terisolasi, tanpa adanya agregasi yang menyoroti ketidaksetaraan antar wilayah. Analisis ini menggunakan data yang bersumber dari BPS Indonesia dan BPS provinsi pada tahun 2023, yang mencakup faktor-faktor seperti Harapan Lama Sekolah, Rata-rata Lama Sekolah, Angka Harapan Hidup Laki-laki, Angka Harapan Hidup Perempuan, Pengeluaran Per Kapita

Sebulan, dan Pendapatan Rata-Rata Sebulan. Penelitian ini bertujuan untuk menyaring data yang rumit menjadi informasi yang mudah dipahami, memungkinkan deteksi tren ketidaksetaraan, dan memberikan wawasan strategis untuk membantu para pembuat kebijakan dalam menetapkan tujuan pembangunan yang lebih terfokus. Selain itu, temuan-temuan penelitian ini diharapkan dapat membantu para pembuat kebijakan dalam menyusun strategi peningkatan kesejahteraan yang disesuaikan dengan klaster-klaster regional yang telah ditetapkan; namun, keberhasilan peningkatan kesejahteraan tersebut bergantung pada integrasi hasil penelitian dengan kebijakan yang akan diberlakukan.

## 2. METODE PENELITIAN

### 2.1. Tahapan Penelitian

Penelitian ini mencakup beberapa langkah. Mulai dengan identifikasi masalah, studi literatur, pengumpulan data, normalisasi data, pemberian bobot untuk tiap variabel atau atribut, penerapan *k-means clustering*, evaluasi hasil *k-means* hingga tahap penyusunan hasil penelitian dalam bentuk tabel dan diagram skater-plot. Setiap tahap ini dapat dilihat pada gambar 1 berikut.



Gambar 1. Langkah penelitian

Pada tahapan identifikasi masalah, penulis menemukan bahwa data Indeks Pembangunan Manusia yang terdapat BPS masih berupa data tabel individual. Maka penelitian ini mencoba untuk menyederhanakan data tersebut kedalam bentuk kelompok kabupaten/kota dalam tiga buah klaster (rendah, sedang dan tinggi). Kemudian, pada tahapan studi literatur penulis mencari dasar pengetahuan terkait data, metode yang akan di gunakan dalam penelitian. Langkah selanjutnya, pada tahap pengumpulan data, penulis melakukan pengumpulan data dari BPS berdasarkan variabel atau atribut yang akan di jadikan sebagai indikator pengelompokan. Setelah data terkumpul, penulis selanjutnya melakukan normalisasi data untuk memastikan tidak ada duplikasi, *missing value* data serta tidak ada variabel atau atribut yang terlalu mendominasi. Kemudian, setelah data di rasa sudah sesuai, kemudian setiap variabel atau atribut di berikan bobot guna memberikan nilai seberapa penting variabel atau atribut tersebut. Setelah di rasa

sesuai, maka tahapan k-means dapat di lakukan. Tahapan k-means di lakukan sebagai metode untuk mengelompokan kabupaten/kota kedalam tiga buah klaster (rendah, sedang tinggi). Setelah klaster berhasil di bentuk, maka tahap evaluasi hasil *clustering* di lakukan. Langkah ini di lakukan untuk melihat seberapa optimal klaster yang terbentuk berdasarkan metode k-means. Setelah itu, hasil penelitian pun di dapatkan. Langkah terakhir adalah menyusun hasil penelitian kedalam bentuk tabel klaster dan kabupaten/kota yang menjadi anggotanya.

## 2.2. Pengumpulan Data

Pada penelitian ini data yang diteliti didapat melalui laman resmi Badan Pusat Statistik (BPS) Indonesia serta BPS di setiap provinsi di Indonesia. Data yang dikumpulkan mencakup seluruh kabupaten/kota di Indonesia pada tahun 2023 sejumlah 514 kabupaten/kota dengan jumlah variabel yang akan di teliti sejumlah 6 variabel atau 6 indikator . Indikator yang digunakan meliputi Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Angka Harapan Hidup Laki-Laki, Angka Harapan Hidup Perempuan, Rata-Rata Pendapatan per Bulan dan Pengeluaran Per Kapita per Bulan.

## 2.3. Algoritma K-Means

Dalam *unsupervised learning*, salah satu algoritma yang dapat digunakan adalah *K-Means Clustering*. algoritma tersebut kerap dipakai dalam pengelompokan sekumpulan data berdasarkan kemiripan karakteristiknya [17]. *K-Means* ini menyebarkan data ke dalam beberapa klaster atau kelompok. Setiap langkah pada *K-Means Clustering* dapat dilakukan sebagai berikut:

### 1. Menetapkan jumlah *clustering*

Dalam *K-Means* Langkah pertama adalah menetapkan jumlah klaster (*K*). Selanjutnya, titik pusat pada kumpulan data dipilih secara acak sebanyak jumlah klaster dan digunakan sebagai *centroid*.

### 2. Menghitung jarak tiap titik dengan centroid

Setelah centroid ditentukan, langkah selanjutnya adalah menghitung jarak *Euclidean*. Jarak *Euclidean* merupakan jarak antara tiap data dengan masing-masing centroid yang telah di tentukan sebelumnya [18]. Jarak *Euclidean* dinyatakan dengan persamaan (1):

$$d(x_i, c_j) = \sqrt{\sum_{m=1}^n (x_{im} - c_{jm})^2} \quad (1)$$

Keterangan:

Persamaan (1) menghitung jarak antara titik set data dengan centroid dengan:

$d(x_i, c_j)$  adalah jarak dari titik data  $x_i$ , dengan centroid  $c_j$ .

$x_{im}$  adalah nilai fitur  $m$  dari titik data  $x_i$ .

$c_{jm}$  adalah nilai fitur  $m$  dari centroid  $c_j$ .

$n$  adalah jumlah variabel.

Jarak tersebut digunakan untuk menentukan seberapa dekat titik data dengan centroid.

Hal tersebut membantu dalam menentukan *cluster*.

### 3. Menentukan klaster menggunakan persamaan (2)

$$Cluster(x_i) = \underset{j}{\operatorname{argmin}} d(x_i, c_j) \quad (2)$$

Keterangan

Setiap titik data  $x_i$  akan dikelompokan dengan centroid terdekatnya. Fungsi *argmin* akan mencari indeks  $j$  dari centroid yang meminimalkan jarak  $d(x_i, c_j)$ .

### 4. Memperbarui centroid menggunakan persamaan (3)

$$c_j = \left( \frac{1}{n_j} \right) \sum_{x \in Cluster_j} x \quad (3)$$

Keterangan:

Setelah seluruh data dikelompokkan ke dalam klaster, perhitungan centroid baru dilakukan berdasarkan rerata dari semua titik data pada setiap klaster.

$c_i$  adalah centroid baru untuk klaster  $j$ .

$n_j$  adalah jumlah titik data pada klaster  $j$ .

$x \in \text{Cluster}_j$  menunjukkan bahwa hanya mengambil titik data yang termasuk dalam klaster  $j$ .

5. Mengulangi proses

Langkah 2 hingga 4 diulang hingga centroid tidak mengalami perubahan yang signifikan.

6. Fungsi objek menggunakan persamaan (4)

$$J = \sum_{j=1}^K \sum_{x_i \in \text{Cluster}_j} d(x_i, c_j)^2 \quad (4)$$

Keterangan

Fungsi bertujuan untuk meminimalkan jumlah kuadrat jarak antara titik data dengan centroid terdekatnya.

$J$  adalah total jarak kuadrat dalam klaster. Fungsi tersebut membantu dalam mengevaluasi

#### 2.4. Davies-Bouldin Index (DBI)

Pada proses DBI, tingkat optimalisasi klaster yang telah ditentukan sebelumnya akan dianalisis menggunakan metode *Davies-Bouldin Index (DBI)* [19]. Langkah ini dipakai guna menilai mutu suatu klaster. Semakin mendekati 0 nilai yang diperoleh dari *Davies-Bouldin Index (DBI)*, semakin baik pula mutu dari klaster yang dibentuk melalui *K-Means Clustering* [20]. DBI dinyatakan dengan persamaan (5)

$$DB = \left(\frac{1}{N}\right) \sum_{i=1}^N \max_{j \neq i} (R_{ij}) \quad (5)$$

Dengan:

DB merupakan *Daevies-Bouldin Index*.

$N$  merupakan total jumlah klaster.

$\sum_{i=1}^N$  menyatakan perhitungan dilakukan dengan menjumlahkan semua klaster dimulai dari klaster pertama ( $i=1$ ) hingga klaster terakhir ( $i=N$ ).

$\max_{j \neq i}$  menghitung nilai maksimal dari rasio  $R_{ij}$  antar klaster  $i$  dan  $j$ .

$R_{ij}$  rasio yang dipakai guna mengukur rasio kemiripan antara klaster  $i$  dan  $j$ , dinyatakan dengan persamaan (6)

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (6)$$

Dengan:

$S_i$  dan  $S_j$  jarak antara titik data pada klaster  $i$  atau  $j$  dengan centroid.

$R_{ij}$  mengukur jarak antara centroid  $i$  dengan centroid  $j$ .

*K-Means Clustering* menjadi cara analisis data yang dipakai untuk penelitian ini. Algoritma *K-Means* membagi tiap kabupaten/kota menjadi tiga klaster berdasarkan kedekatan jarak antar data dalam ruang multi dimensi yang diwakili oleh tiap indikator.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Data

Statistik berikut ini mewakili variabel-variabel IPM yang digunakan dalam investigasi ini. Analisis ini menggunakan enam variabel data IPM: Angka Harapan Hidup Perempuan, Angka Harapan Hidup Laki-laki, Rata-rata Lama Sekolah, Harapan Lama Sekolah, Rata-rata Pendapatan Bulanan, dan Pengeluaran Per Kapita Bulanan, seperti yang diilustrasikan pada Tabel 1.

Tabel 1. Data IPM

| Provinsi | Kabupaten/<br>Kota | Angka<br>Harapan<br>Hidup Laki-<br>Laki | Angka<br>Harapan<br>Hidup<br>Perempuan | Harapan<br>Lama<br>Sekolah | Rata-Rata<br>Lama<br>Sekolah | Pengeluaran<br>Per Kapita<br>Tiap Bulan | Rata-Rata<br>Pendapatan Per<br>Bulan |
|----------|--------------------|-----------------------------------------|----------------------------------------|----------------------------|------------------------------|-----------------------------------------|--------------------------------------|
| ACEH     | Simeulue           | 63,72                                   | 67,59                                  | 14,28                      | 9,81                         | 640.500                                 | 3.413.666                            |
| ACEH     | Aceh<br>Singkil    | 65,84                                   | 69,82                                  | 14,35                      | 8,7                          | 781.167                                 | 3.413.666                            |
| .....    | .....              | .....                                   | .....                                  | .....                      | .....                        | .....                                   | .....                                |
| PAPUA    | Intan Jaya         | 64,22                                   | 68,11                                  | 7,69                       | 3,27                         | 511.667                                 | 3.470.000,00                         |
| PAPUA    | Deiyai             | 64,04                                   | 67,92                                  | 9,85                       | 3,27                         | 425.333                                 | 3.470.000,00                         |

Tiap variabel tersebut menjelaskan hal-hal seperti: Angka Harapan Hidup Laki-Laki mencerminkan rata-rata usia laki-laki yang dihitung sejak lahir dalam tahun, menunjukkan tingkat kesehatan laki-laki di suatu wilayah. Angka Harapan Hidup Perempuan merupakan variabel yang mencerminkan tingkat kesehatan wanita di suatu wilayah serta mencerminkan harapan usia wanita yang dihitung sejak kelahirannya dalam tahun. Harapan Lama Sekolah mencerminkan harapan potensi lama seseorang dapat mengenyam pendidikan formal dimulai dari usia tertentu dinyatakan dalam tahun serta memberikan gambaran tingkat akses terhadap pendidikan di masa depan. Rata-Rata Lama Sekolah mencerminkan rata-rata lamanya seseorang mengenyam pendidikan formal saat ini yang dinyatakan dalam tahun. Pengeluaran Per Kapita Tiap Bulan menyatakan pengeluaran tiap individu dalam satu bulan yang dinyatakan dalam ratus ribu. Rata-Rata Pendapatan Tiap Bulan menyatakan pendapatan rata-rata tiap individu dalam satu bulan yang dinyatakan dalam juta rupiah.

### 3.2. Data Preprocessing

Berdasarkan data yang telah diperoleh, terdapat beberapa variabel yang memerlukan penyesuaian, yaitu Pengeluaran Per Kapita Tiap Bulan dan Rata-Rata Pendapatan Per Bulan. Kedua variabel tersebut diubah ke dalam skala 1 hingga 100, di mana setiap satuan pada skala tersebut merepresentasikan satu juta rupiah. Kemudian, pada tahap ini, setiap variabel yang akan digunakan diberi bobot. Bobot yang diberikan adalah sebagai berikut: Harapan Lama Sekolah (0.20), Rata-Rata Lama Sekolah (0.15), Angka Harapan Hidup Perempuan (0.125), Angka Harapan Hidup Laki-Laki (0.125), Rata-Rata Pendapatan per Bulan (0.15) dan Pengeluaran Per Kapita per Bulan (0.25). Setelah pembobotan dilakukan pada setiap variabel, variabel Angka Harapan Hidup Perempuan dijumlahkan dengan Angka Harapan Hidup Laki-Laki, kemudian dibagi dua untuk menghasilkan variabel baru yang disebut Angka Harapan Hidup. Kemudian dilakukan normalisasi pada variabel Pendapatan Per Kapita Tiap Bulan dan Rata-Rata Pendapatan Per Bulan agar kedua variabel tersebut tidak terlalu mendominasi hasil dari clustering yang akan dilakukan. Kedua variabel tersebut di berikan skala antara 1 hingga 100 dengan setiap seratus ribu dinyatakan dengan angka satu. Berikut adalah hasil tabel data yang telah diberikan pembobotan dan telah dilakukan normalisasi data dapat di liat pada tabel 2.

Tabel 2. hasil normalisasi

| Provinsi | Kabupaten/Kota | Harapan<br>Lama<br>Sekolah | Rata-Rata<br>Lama Sekolah | Pengeluaran Per<br>Kapita Tiap<br>Bulan | Rata-Rata<br>Pendapatan<br>Per Bulan | Angka Harapan<br>Hidup |
|----------|----------------|----------------------------|---------------------------|-----------------------------------------|--------------------------------------|------------------------|
| ACEH     | Simeulue       | 2,856                      | 1,4715                    | 1,60125                                 | 5,120499                             | 8,206875               |
| ACEH     | Aceh Singkil   | 2,87                       | 1,305                     | 1,9529175                               | 5,120499                             | 8,47875                |
| .....    | .....          | .....                      | .....                     | .....                                   | .....                                | .....                  |
| PAPUA    | Intan Jaya     | 1,538                      | 0,4905                    | 1,2791675                               | 5,205                                | 8,270625               |
| PAPUA    | Deiyai         | 1,97                       | 0,4905                    | 1,0633325                               | 5,205                                | 8,2475                 |
| PAPUA    | Kota Jayapura  | 3,052                      | 1,776                     | 3,1816675                               | 5,205                                | 8,89                   |

### 3.3. Penerapan K-Means Clustering

#### a. Menentukan jumlah kluster

Dalam penelitian ini jumlah kluster yang akan di bentuk adalah sejumlah tiga kluster yakni

klaster rendah, sedang dan tinggi.

b. Menghitung jarak *euclidean*

Jumlah klaster yang akan dibentuk adalah sejumlah tiga klaster, maka ditentukan sejumlah tiga centroid sebagai titik pusat klaster. Adapun kolom yang akan dijadikan sebagai centroid adalah kolom pada kabupaten Simeulue, Lebak dan Keerom. Berikut merupakan tabel 3 yang menyatakan centroid yang digunakan:

Tabel 3. *Centroid* Yang Digunakan

| Provinsi         | Kabupaten/Kota | Harapan Lama Sekolah | Rata-Rata Lama Sekolah | Pengeluaran Per Kapita Tiap Bulan | Rata-Rata Pendapatan Per Bulan | Angka Harapan Hidup |
|------------------|----------------|----------------------|------------------------|-----------------------------------|--------------------------------|---------------------|
| ACEH             | Simeulue       | 2,856                | 1,4715                 | 1,60125                           | 5,120499                       | 8,206875            |
| SUMATERA SELATAN | Lahat          | 2,488                | 1,284                  | 2,2522925                         | 4,377                          | 8,365625            |
| PAPUA            | Keerom         | 2,492                | 1,2855                 | 1,9885425                         | 5,205                          | 8,410625            |

Perhitungan menghitung jarak *euclidean* antara centroid dengan salah satu titik data. Jarak yang akan dihitung adalah jarak antara centroid Simeulue dengan titik data kabupaten Aceh Sinkil. Langkah tersebut dilakukan untuk menghitung jarak data ke tiap tiap centroid. Kemudian data di kelompokkan berdasarkan centroid terdekatnya. Maka didapatkan jarak untuk tiap data terhadap centroid, Centroid Simeulue dinyatakan tabel 4.

Tabel 4. Jarak Terhadap Centroid Simeulue:

| Kabupaten/Kota | Harapan Lama Sekolah (Simeulue) | Rata-Rata Lama Sekolah (Simeulue) | Pengeluaran Per Kapita Tiap Bulan (Simeulue) | Rata-Rata Pendapatan Per Bulan (Simeulue) | Angka Harapan Hidup (Simeulue) |
|----------------|---------------------------------|-----------------------------------|----------------------------------------------|-------------------------------------------|--------------------------------|
| Simeulue       | 0                               | 0                                 | 0                                            | 0                                         | 0                              |
| Aceh Singkil   | 0,014                           | 0,1665                            | 0,3516675                                    | 0                                         | 0,271875                       |
| .....          | .....                           | .....                             | .....                                        | .....                                     | .....                          |
| Intan Jaya     | 1,318                           | 0,981                             | 0,3220825                                    | 0,084501                                  | 0,06375                        |
| Deiyai         | 0,886                           | 0,981                             | 0,5379175                                    | 0,084501                                  | 0,040625                       |
| Kota Jayapura  | 0,196                           | 0,3045                            | 1,5804175                                    | 0,084501                                  | 0,683125                       |

*Centroid* Lahat dinyatakan tabel 5

Tabel 5. Jarak Terhadap *Centroid* Lahat

| Kabupaten/Kota | Harapan Lama Sekolah (Lebak) | Rata-Rata Lama Sekolah (Lebak) | Pengeluaran Per Kapita Tiap Bulan (Lebak) | Rata-Rata Pendapatan Per Bulan (Lebak) | Angka Harapan Hidup (Lebak) |
|----------------|------------------------------|--------------------------------|-------------------------------------------|----------------------------------------|-----------------------------|
| Simeulue       | 0,436                        | 0,4815                         | 0,3008325                                 | 0,116001                               | 0,315625                    |
| Aceh Singkil   | 0,45                         | 0,315                          | 0,050835                                  | 0,116001                               | 0,04375                     |
| .....          | .....                        | .....                          | .....                                     | .....                                  | .....                       |
| Intan Jaya     | 0,882                        | 0,4995                         | 0,622915                                  | 0,0315                                 | 0,251875                    |
| Deiyai         | 0,45                         | 0,4995                         | 0,83875                                   | 0,0315                                 | 0,275                       |
| Kota Jayapura  | 0,632                        | 0,786                          | 1,279585                                  | 0,0315                                 | 0,3675                      |

*Centroid* Keerom dinyatakan tabel 6

Tabel 6. Jarak Terhadap *Centroid* Keerom

| Kabupaten/Kota | Harapan Lama Sekolah (Keerom) | Rata-Rata Lama Sekolah (Keerom) | Pengeluaran Per Kapita Tiap Bulan (Keerom) | Rata-Rata Pendapatan Per Bulan (Keerom) | Angka Harapan Hidup (Keerom) |
|----------------|-------------------------------|---------------------------------|--------------------------------------------|-----------------------------------------|------------------------------|
| Simeulue       | 0,364                         | 0,186                           | 0,3872925                                  | 0,084501                                | 0,20375                      |
| Aceh Singkil   | 0,378                         | 0,0195                          | 0,035625                                   | 0,084501                                | 0,068125                     |
| .....          | .....                         | .....                           | .....                                      | .....                                   | .....                        |
| Intan Jaya     | 0,954                         | 0,795                           | 0,709375                                   | 0                                       | 0,14                         |
| Deiyai         | 0,522                         | 0,795                           | 0,92521                                    | 0                                       | 0,163125                     |
| Kota Jayapura  | 0,56                          | 0,4905                          | 1,193125                                   | 0                                       | 0,479375                     |

c. Mengelompokkan hasil berdasarkan jarak

Hasil yang diperoleh dari perhitungan di atas akan dimasukkan ke dalam setiap kluster. Hasil dari perhitungan terdapat pada tabel 7.

Tabel 7. Hasil Kluster Terhadap *Centroid*

| Kabupaten/Kota | Cluster |
|----------------|---------|
| Simeulue       | 0       |
| Aceh Singkil   | 1       |
| .....          | .....   |
| Intan Jaya     | 1       |
| Deiyai         | 1       |
| Kota Jayapura  | 2       |

Dimana 0 dinyatakan sebagai cluster rendah, 1 sebagai cluster sedang dan 2 sebagai cluster tinggi.

d. Menentukan *centroid* baru

*Centroid* baru di tentukan berdasarkan jumlah keseluruhan nilai anggota kluster dibagi dengan jumlah anggota kluster. Hasil centroid baru yang dihasilkan adalah seperti pada tabel 8.

Tabel 8. *Centroid* Baru

| Harapan Lama Sekolah | Rata-Rata Lama Sekolah | Pengeluaran Per Kapita Tiap Bulan | Rata-Rata Pendapatan Per Bulan | Angka Harapan Hidup | Cluster   |
|----------------------|------------------------|-----------------------------------|--------------------------------|---------------------|-----------|
| 2,594971831          | 1,214334507            | 1,840218627                       | 4,216496814                    | 8,394713908         | Cluster 1 |
| 2,451944954          | 1,129252294            | 1,901630505                       | 5,137462196                    | 8,794288991         | Cluster 2 |
| 2,722395437          | 1,412851711            | 2,703226378                       | 5,008223503                    | 8,984831274         | Cluster 3 |

e. Menghitung ke iterasi berikutnya atau ke *centroid* baru

Setelah centroid baru di temukan, langkah berikutnya adalah menghitung kembali jarak tiap titik data ke centroid baru. Langkah-langkah di atas di lakukan hingga tidak terjadi perubahan kluster atau centroid tidak mengalami perubahan yang signifikan. Dalam penelitian ini perhitungan berhenti pada iterasi 39. Langkah berikutnya adalah menghitung jarak dari titik tiap data ke setiap centroid. didapatkan hasil seperti dalam tabel 9.

Tabel 9. Jarak Data Terhadap Tiap *Centorid*.

| Kabupaten/Kota | Jarak ke Centroid 1 | Jarak ke Centroid 2 | Jarak ke Centroid 3 |
|----------------|---------------------|---------------------|---------------------|
| Simeulue       | 0,904               | 2,103               | 1,316               |
| Aceh Singkil   | 0,781               | 1,802               | 0,896               |
| .....          | .....               | .....               | .....               |
| Intan Jaya     | 1,571               | 2,697               | 2,139               |
| Deiyai         | 1,452               | 2,694               | 2,096               |
| Kota Jayapura  | 1,759               | 1,421               | 0,891               |

### 3.4. Hasil *K-Means Clustering*

Berdasarkan skema *clustering* yang telah diterapkan, diperoleh hasil sebagai berikut

a. Tabel Hasil *clustering*

Jumlah kabupaten/kota tiap provinsi yang tergabung dalam kluster dinyatakan dalam tabel 10 berikut:

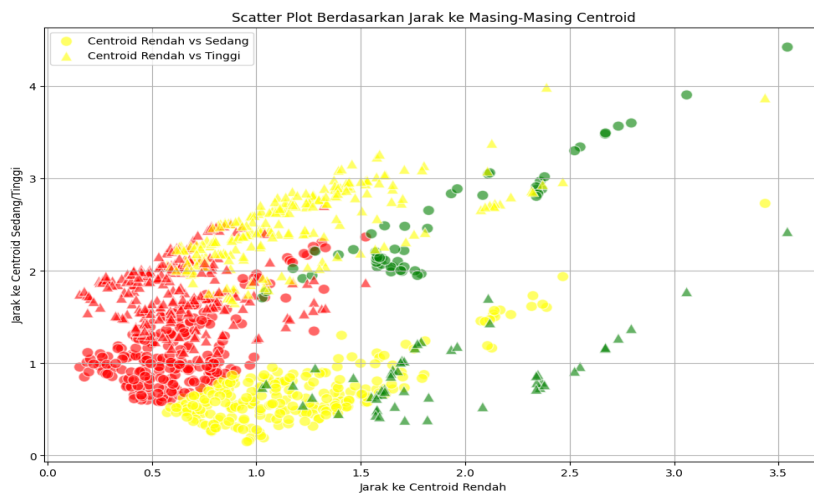
Tabel 10. Jumlah Kabupaten/Kota Tiap Kluster

| Provinsi       | Jumlah_Rendah | Jumlah_Sedang | Jumlah_Tinggi | Jumlah_Kabupaten |
|----------------|---------------|---------------|---------------|------------------|
| ACEH           | 13            | 10            | 0             | 23               |
| BALI           | 8             | 0             | 1             | 9                |
| BANTEN         | 1             | 2             | 5             | 8                |
| BENGKULU       | 0             | 9             | 1             | 10               |
| D I YOGYAKARTA | 4             | 0             | 1             | 5                |
| DKI JAKARTA    | 0             | 0             | 6             | 6                |
| GORONTALO      | 1             | 5             | 0             | 6                |
| JAMBI          | 6             | 5             | 0             | 11               |

|                      |    |    |    |    |
|----------------------|----|----|----|----|
| JAWA BARAT           | 0  | 0  | 27 | 27 |
| JAWA TENGAH          | 32 | 2  | 1  | 35 |
| JAWA TIMUR           | 28 | 9  | 1  | 38 |
| KALIMANTAN BARAT     | 2  | 12 | 0  | 14 |
| KALIMANTAN SELATAN   | 11 | 2  | 0  | 13 |
| KALIMANTAN TENGAH    | 10 | 4  | 0  | 14 |
| KALIMANTAN TIMUR     | 9  | 0  | 1  | 10 |
| KALIMANTAN UTARA     | 3  | 2  | 0  | 5  |
| KEP. BANGKA BELITUNG | 6  | 0  | 1  | 7  |
| KEPULAUAN RIAU       | 5  | 0  | 2  | 7  |
| LAMPUNG              | 4  | 9  | 2  | 15 |
| MALUKU               | 1  | 10 | 0  | 11 |
| MALUKU UTARA         | 1  | 9  | 0  | 10 |
| NUSA TENGGARA BARAT  | 3  | 7  | 0  | 10 |
| NUSA TENGGARA TIMUR  | 1  | 21 | 0  | 22 |
| PAPUA                | 5  | 24 | 0  | 29 |
| PAPUA BARAT          | 2  | 11 | 0  | 13 |
| RIAU                 | 10 | 2  | 0  | 12 |
| SULAWESI BARAT       | 0  | 6  | 0  | 6  |
| SULAWESI SELATAN     | 20 | 4  | 0  | 24 |
| SULAWESI TENGAH      | 3  | 10 | 0  | 13 |
| SULAWESI TENGGARA    | 3  | 14 | 0  | 17 |
| SULAWESI UTARA       | 9  | 3  | 3  | 15 |
| SUMATERA BARAT       | 9  | 10 | 0  | 19 |
| SUMATERA SELATAN     | 2  | 13 | 2  | 17 |
| SUMATERA UTARA       | 14 | 19 | 0  | 33 |

Pada tabel 10 di atas, hasil klasterisasi menunjukkan distribusi jumlah kabupaten/kota dalam masing-masing cluster berdasarkan provinsi. Misalnya, di Provinsi Aceh, terdapat 13 kabupaten/kota dalam cluster rendah, 10 kabupaten/kota dalam cluster sedang, dan tidak ada kabupaten/kota dalam cluster tinggi, dengan total 23 kabupaten/kota. Di Provinsi Bali, 8 kabupaten/kota termasuk dalam cluster rendah, 1 kabupaten/kota dalam cluster tinggi, dan tidak ada kabupaten/kota dalam cluster sedang, dari total 9 kabupaten/kota. Sementara itu, Provinsi Jawa Barat memiliki 27 kabupaten/kota dalam cluster tinggi, tanpa adanya kabupaten/kota dalam cluster rendah atau sedang, dari total 27 kabupaten/kota yang ada. Begitu seterusnya untuk setiap provinsi lainnya.

b. Dinyatakan dalam diagram skater-plot



Gambar 2. Diagram Skater-Plot Berdasarkan Klaster  
Berdasarkan *scatter plot* tersebut, titik berwarna merah mewakili kabupaten/kota yang

tergolong dalam cluster rendah, kuning untuk cluster sedang, dan hijau untuk cluster tinggi. Terlihat banyak kabupaten/kota dalam cluster rendah yang mulai beranjak naik menjadi cluster sedang, namun terdapat pula kabupaten/kota di cluster sedang yang berpotensi turun kembali ke cluster rendah. Selain itu, beberapa kabupaten/kota di cluster sedang mulai naik ke cluster tinggi, meskipun jumlah kabupaten/kota yang termasuk cluster tinggi masih sangat sedikit. Titik kabupaten/kota yang semakin dekat ke titik 0 pada sumbu x dan y mengindikasikan bahwa kabupaten/kota tersebut tergolong dalam cluster rendah. Jika titik berada di rentang  $x=0,5$  hingga  $x=1,5$  dan  $y=1,0$  hingga  $y=3,0$ , maka kabupaten/kota tersebut masuk dalam cluster sedang, sedangkan titik dengan  $x \geq 1,5$  dan  $y \geq 3$  termasuk ke dalam cluster tinggi.

### 3.5. Pengujian Hasil *K-Means Clustering* dengan *Davies-Bouldin Index*

Pengujian *clustering* dengan menggunakan *Davies-Bouldin Index (DBI)* menghasilkan skor sebesar 1.17. Angka ini menunjukkan kualitas *clustering* yang dievaluasi berdasarkan tingkat kerapatan di dalam *cluster* serta jarak antar *cluster*. DBI memiliki rentang nilai mulai dari 0 hingga tak terhingga, di mana nilai yang lebih kecil menandakan *clustering* yang lebih optimal dengan *cluster* yang lebih rapat dan terpisah dengan baik. Hasil skor 1.17 mengindikasikan bahwa *clustering* yang dilakukan belum cukup baik. Hal tersebut merujuk pada penelitian yang dilakukan oleh [21] dalam penelitian tersebut hasil DBI terhadap k-means dinyatakan cukup baik karena hasil DBI telah mencapai angka 0.9762331449809145. meskipun masih ada peluang untuk memperbaiki pemisahan antar *cluster* atau mengurangi penyebaran dalam *cluster*. Skor ini juga dapat digunakan sebagai referensi untuk membandingkan performa *clustering* dengan pengaturan atau metode lain..

## 4. KESIMPULAN DAN SARAN

Berdasarkan hasil temuan sebelumnya, dapat disimpulkan bahwa metode K-Means *Clustering* dapat digunakan untuk membagi setiap kabupaten/kota di setiap provinsi di Indonesia ke dalam tiga klaster berdasarkan variabel: Rata-rata Pendapatan Per Bulan, Pengeluaran Per Kapita Per Bulan, Rata-rata Lama Sekolah, Harapan Lama Sekolah, Angka Harapan Hidup Perempuan, dan Angka Harapan Hidup Laki-laki. Hasil kualitas DBI pengelompokan adalah 1,17, dan terdapat 39 iterasi. Hasil tersebut belum cukup baik, mengingat penelitian [21] serta pada dalam penelitian [22]. menyatakan DBI optimal dari klastering berkisar 0,9 hingga 0,8 atau semakin dekat dengan 0, maka klastering semakin bagus.

Temuan pengelompokan menunjukkan bahwa pengelompokan kabupaten/kota di Indonesia didominasi oleh klaster rendah dan sedang. Secara rinci, jumlah kabupaten/kota yang termasuk dalam klaster rendah mencapai 210, klaster sedang sebanyak 218, dan klaster tinggi berjumlah 55. Klaster rendah banyak ditemukan di provinsi Sumatera Utara, Bali, Jawa Tengah, DI Yogyakarta, Kalimantan Timur, Aceh, Kalimantan Selatan, Kalimantan Tengah, Jawa Timur, Kepulauan Riau, Jambi, Kepulauan Bangka Belitung, Riau, Sulawesi Selatan, dan Kalimantan Utara. Sementara itu, klaster sedang dominan di provinsi Sulawesi Barat, Papua Barat, Kalimantan Barat, Maluku, Sulawesi Tengah, Gorontalo, Lampung, Nusa Tenggara Barat, Papua, Sulawesi Tenggara, Maluku Utara, Kalimantan Tengah, dan Bengkulu. Adapun klaster tinggi banyak ditemui di provinsi Banten, DKI Jakarta, dan Jawa Barat.

Mengingat hasil DBI pada penelitian ini masih 1,17, saran untuk penelitian berikutnya agar dapat mengoptimalkan hasil dari DBI agar klaster yang terbentuk lebih maksimal dengan menggunakan metode klastering lain seperti *agglomerative clustering* atau dengan melakukan peningkatan pada *k-means* dengan memperbaiki penentuan centroid awal. Langkah lain yang dapat di ambil seperti menggunakan *elbow* atau *Silhouette* untuk melihat berapa jumlah klaster maksimalnya. Dengan mengetahui jumlah klaster maksimalnya, maka data yang terpartisi akan memiliki tingkat kemiripan yang lebih tinggi antar sesama klaster.

## DAFTAR PUSTAKA

- [1] M. F. M. Muhammad Taufan Ashshiddiqi, Ichsan Firmansyah, Khairu Syifa Ahyani, Laila Febriyuni Eka Putri, "Strategi Program IPM di Desa Pakuaon Kecamatan Sukaresmi Kabupaten Cianjur," *J. Inov. Penelit.*, vol. 2, no. 4, pp. 1153–1163, 2021.
- [2] N. Alvariqati, V. Ansori, A. Muhaimin, A. Terza, P. Studi, and S. Data, "Pembentukan Klaster terhadap Indeks Pembangunan Manusia di Wilayah Jawa Timur," *J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 5, no. 1, pp. 514–519, 2024.
- [3] A. Hakiki, Yulmardi, and Zulfanetti, "Estimasi Model Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jambi," Aug. 2020. [Online]. Available: [www.journal.uta45jakarta.ac.id](http://www.journal.uta45jakarta.ac.id)
- [4] B. Rozak, D. Febriawan, and F. N. Hasan, "Implementasi Business Intelligence untuk Visualisasi Laju Indeks Pembangunan Manusia Kota Cirebon Menggunakan Google Collab," *Sainteks*, vol. 21, no. 1, pp. 33-45, 2024.
- [5] Rahmatil R and Wijayanto A, "Analisis Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering (Studi Kasus Indeks Pembangunan Manusia tahun 2019)," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 73-80, 2021.
- [6] A. M. Sikana and A. W. Wijayanto, "Comparison Analysis of Human Development Index Grouping in Indonesia in 2019 using Partitioning and Hierarchical Clustering Methods," *J. Ilmu Komput.*, vol. 14, no. 2, pp. 66–78, 2021.
- [7] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, "Implementasi K-Means Clustering Dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, Dan Model RFM (Studi Kasus: Lantikya Store Jombang)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2099-2112, 2024.
- [8] E. Widodo, S. N. Mashita, and Y. G. Prasetyowati, "Perbandingan Metode Average Linkage, Complete Linkage, dan Ward'S pada Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia," *Fakt. Exacta*, vol. 13, no. 2, pp. 81-87, 2020.
- [9] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 849-855, 2022.
- [10] A. Maulidin, *et al.*, "Implementasi K-Means untuk Clustering Kepuasan Mahasiswa Teknik Informatika Terhadap Layanan Akademik," *Jurnal Genta Mulia*, vol. 15, no. 2, pp. 124-133, 2024.
- [11] J. Hutagalung, and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, pp. 1187-1194, 2021.
- [12] R. S. Wicaksana, D. Heksaputra, T. A. Syah, and F. F. Nur'aini, "Pendekatan K-Means Clustering Metode Elbow Pada Analisis Motivasi Pengunjung Festival Halal JHF#2," *J. Ilmu Ekonomi Islam*, vol. 9, no. 3, pp. 4162-4176, 2023.
- [13] L. S. Riza, R. A. Rosdiyana, A. Wahyudin, and A. R. Pérez, "The k-means algorithm for generating sets of items in educational assessment," *Indonesian Jurnal Science Technology*, vol. 6, no. 1, pp. 93–100, 2021.
- [14] N. Tri, S. Saptadi, P. Chyan, and V. P. Taga, "Using K-Means Algorithm to Investigate Community Behavior in Treating Waste toward Smart City," *International Journal on Advanced Science Engineering and Information Technology*, vol. 11, no. 4, pp. 1455-1462, 2021.
- [15] F. D. Rahman, M. I. Zulfa, and A. Taryana, "Clustering Dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode K-Means Dan Random Forest," *Jurnal SINTA Sistem Informasi dan Teknologi Komputasi*, vol. 1, no. 2, pp. 90-97, 2024.
- [16] T. Terttiaavini, "A Hybrid Approach Using K-Means Clustering and the SAW Method for Evaluating and Determining the Priority of SMEs in Palembang City," *Journal of Intelligent System and Computation*, vol. 6, no. 1, pp. 46–53, 2024.
- [17] M. I. Zuhendra, R. Hidayat, and H. Hendrawaty, "Penerapan Data Mining Untuk

- Klasterisasi Tingkat Kemiskinan Berdasarkan Data Terpadu Kesejahteraan Sosial (Dtk),” *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7, no. 1, pp. 32–41, 2024.
- [18] R. Anggara, S. Defit, and B. Hendrik, “Implementasi K-Means Clustering Dalam Analisa Soal Ujian CBT Universitas Baiturrahmah,” *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 5, no. 2, pp. 577–586, 2024.
- [19] S. Suraya, M. Sholeh, and D. Andayati, “Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa,” *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 6, no. 1, pp. 51–60, 2023.
- [20] Y. Imam T. Umagapi, Basirung Umaternate, Hazriani, “Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa,” *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, vol. 7, no. 1, pp. 303–308, 2023.
- [21] F. Fathurrahman, S. Harini, and R. Kusumawati, “Evaluasi Clustering K-Means Dan K-Medoid Pada Persebaran Covid-19 Di Indonesia Dengan Metode Davies-Bouldin Index (Dbi),” *Jurnal Mnemon.*, vol. 6, no. 2, pp. 117–128, 2023.
- [22] S. Ramadhani, D. Azzahra, and T. Z, “Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student’s Thesis,” *Digital Zone Jurnal Teknologi Infomasi dan Komunikasi*, vol. 13, no. 1, pp. 24–33, 2022.