



## SKEWNESS

Jurnal Statistika, Aktuaria dan Sains Data

Volume 1, No.1, April 2024

### Perbandingan Metode *K-Means* dan *Two Step Cluster* dalam Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah

Alim Jaizul Wahid<sup>1</sup>, Supriyanto<sup>2\*</sup>,

<sup>1</sup> Program Studi Matematika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jenderal Soedirman, Purwokerto, Indonesia

<sup>2</sup> Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jenderal Soedirman, Purwokerto, Indonesia

E-mail korespondensi: [supriyanto2505@unsoed.ac.id](mailto:supriyanto2505@unsoed.ac.id)\*

**Abstrak.** Nilai Indeks Pembangunan Manusia (IPM) diperoleh dari hasil gabungan ketiga dimensi dasar IPM. Namun, tidak ditunjukkan indikator yang dominan terhadap tinggi rendahnya nilai IPM. Akibatnya nilai IPM menjadi kurang spesifik menggambarkan kondisi yang ada di masyarakat, salah satunya yang terjadi di Provinsi Jawa Tengah. Tujuan penelitian ini adalah menganalisis perbandingan hasil pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022 dan menentukan metode terbaik dilihat dari nilai *Silhouette Index* yang terbesar. Metode *clustering* yang digunakan adalah metode *K-Means* dan *Two Step Cluster*. Hasil analisis diperoleh pengelompokan metode *K-Means* dengan jumlah cluster sebanyak 4 cluster menghasilkan nilai *Silhouette Index* sebesar 0,3763 sehingga termasuk kategori yang lemah, sedangkan pengelompokan metode *Two Step Cluster* diperoleh jumlah cluster sebanyak 2 cluster dan menghasilkan nilai *Silhouette Index* sebesar 0,5763 sehingga termasuk kategori yang kuat. Dengan demikian metode *Two Step Cluster* merupakan metode terbaik dalam pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022 dengan mengelompokkan kabupaten/kota menjadi 2 cluster. Cluster 1 termasuk pengelompokan kabupaten/kota dengan tingkat indikator dari IPM yang tinggi dan cluster 2 termasuk pengelompokan kabupaten/kota dengan tingkat indikator dari IPM yang sangat tinggi.

**Kata kunci:** IPM, *K-Means*, *Two Step Cluster*, *Silhouette Index*

## 1 PENDAHULUAN

United Nations Development Programme [1], memperkenalkan konsep Indeks Pembangunan Manusia (IPM) melalui laporan tahunan yang dipublikasikan dalam *Human Development Report* (HDR). Setiawan dkk. [2], menjelaskan bahwa IPM merupakan suatu pendekatan yang digunakan untuk mengukur tingkat keberhasilan pembangunan manusia. Nilai Indeks Pembangunan Manusia diperoleh dari hasil gabungan tiga dimensi dasar yaitu Umur Panjang dan Hidup Sehat, Pengetahuan, dan Standar Hidup Layak. Namun, tidak ditunjukkan indikator yang dominan terhadap tinggi rendahnya nilai IPM. Akibatnya nilai IPM menjadi kurang spesifik menggambarkan kondisi yang ada di masyarakat, salah satunya yang terjadi di Provinsi Jawa Tengah.

Menurut Nursetyohadi [3], IPM Provinsi Jawa Tengah terus mengalami peningkatan dari periode 2012 sampai 2022 dari 67,21 menjadi 72,79. Meskipun pada

tahun 2020 sempat mengalami perlambatan peningkatan dikarenakan pandemi COVID-19, nilai IPM Provinsi Jawa Tengah mulai mengalami pemulihan di tahun 2022 dengan pertumbuhan sebesar 0,87 % dibandingkan dengan tahun 2021. Namun, kondisi IPM secara kewilayahan kabupaten/kota di Provinsi Jawa Tengah masih terdapat perbedaan kualitas pembangunan yang cukup signifikan. Oleh karena itu, perlu dilakukan pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022 sehingga dapat diketahui dengan jelas kondisi yang ada di masyarakat.

Salah satu metode yang dapat digunakan mengelompokkan objek adalah metode *clustering*. Pada penelitian ini metode *clustering* yang digunakan adalah metode *K-Means* dan *Two Step Cluster*. Kedua metode tersebut dipilih karena memiliki kelebihan yang dijadikan untuk melengkapi kekurangan satu sama lain. Selanjutnya perlu ditentukan metode terbaik dengan melakukan perbandingan hasil pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022 agar dapat diperoleh hasil pengelompokan yang lebih spesifik menggambarkan kondisi yang ada di masyarakat. Metode evaluasi yang dapat digunakan yaitu nilai *Silhouette Index*.

Berdasarkan uraian di atas, tujuan dari penelitian ini adalah menganalisis hasil pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022. Selanjutnya menganalisis hasil perbandingan nilai *Silhouette Index* dari metode *K-Means* dan *Two Step Cluster* untuk mengetahui metode terbaik dilihat dari nilai *Silhouette Index* yang terbesar. Harapannya penelitian ini dapat dijadikan sebagai bahan pertimbangan oleh instansi terkait sebelum menentukan kebijakan tentang pembangunan manusia.

## **2 METODOLOGI**

### **2.1 Data Penelitian**

Penelitian ini menggunakan data sekunder yaitu data indikator dari Indeks Pembangunan Manusia kabupaten/kota di Provinsi Jawa Tengah tahun 2022 sebanyak 35 data yang diperoleh dari laman resmi Badan Pusat Statistik <https://www.bps.go.id>. Variabel yang digunakan adalah Umur Harapan Hidup saat lahir (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Riil per Kapita disesuaikan (PRK).

## 2.2 Prosedur Penelitian

Proses analisis data dilakukan menggunakan bantuan *software* IBM SPSS *Statistics 25* dan *Microsoft Excel 2016*. Prosedur penelitian ini adalah sebagai berikut:

1. Perumusan masalah

Perumusan masalah merupakan proses mengidentifikasi dengan jelas data dan variabel yang akan digunakan dalam penelitian.

2. Pengumpulan data dan input data

Melakukan pengumpulan data indikator dari IPM kabupaten/kota di Provinsi Jawa Tengah tahun 2022. Selanjutnya data tersebut di input ke dalam *Microsoft Excel 2016*.

3. Melakukan standarisasi data

Standarisasi data dilakukan untuk menyetarakan satuan data dengan menggunakan nilai *z-score*.

4. Melakukan pengujian asumsi

Terdapat dua asumsi yang harus dipenuhi sebelum melakukan proses *clustering*, adapun asumsi tersebut yaitu:

a. Sampel representatif / mewakili populasi

Dilakukan pengujian menggunakan uji *Kaiser-Meyer-Olkin* (KMO) untuk mengetahui suatu sampel mewakili populasi atau tidak. Keputusan mewakili populasi jika memiliki nilai KMO di antara  $0,5 \leq KMO \leq 1$  [4].

b. Non multikolinieritas

Ada tidaknya multikolinieritas dapat dilihat dari nilai *Variance Inflation Factor* (VIF). Keputusan terjadi multikolinieritas apabila nilai  $(VIF)_j \geq 10$  [5].

5. Ukuran jarak *Euclidean*

Ukuran jarak yang digunakan dalam penelitian ini adalah jarak *Euclidean* karena memiliki karakteristik yang sesuai dengan metode yang digunakan dalam penelitian ini. Proses pengukuran jarak *Euclidean* dilakukan menggunakan rumus [6]:

$$d(i, C_p) = \sqrt{\sum_{j=1}^m (x_{ij} - C_{pj})^2}; i = 1, 2, \dots, n; j = 1, 2, \dots, m; p = 1, 2, \dots, t \quad (1)$$

dengan  $d(i, C_p)$  adalah jarak antara data ke- $i$  dengan *centroid* ke- $p$ ,  $x_{ij}$  adalah nilai data ke- $i$  pada variabel ke- $j$ ,  $C_{pj}$  adalah nilai *centroid* ke- $p$  pada variabel ke- $j$ ,  $n$

menyatakan banyaknya data,  $m$  menyatakan banyaknya variabel, dan  $t$  menyatakan banyaknya centroid.

6. Melakukan pengelompokan metode *K-Means*

Pada penelitian ini jumlah *cluster K-Means* yang akan dibentuk sebanyak 4 *cluster* dan ukuran jarak yang digunakan adalah jarak *Euclidean*. Algoritma dari metode *K-Means* adalah sebagai berikut [7]:

- 1) Menentukan jumlah *cluster* yang akan dibentuk,
- 2) Bangkitkan *centroid* awal secara acak,
- 3) Hitung jarak masing-masing data terhadap *centroid* dari setiap *cluster* menggunakan Persamaan (1),
- 4) Kelompokkan masing-masing data ke *centroid* terdekat,
- 5) Lakukan proses iterasi dan tentukan kembali *centroid* baru dari masing-masing *cluster* seperti pada langkah 2,
- 6) Ulangi langkah 3 sampai 5 jika posisi *centroid* baru yang terbentuk tidak sama seperti *centroid* sebelumnya.

7. Melakukan pengelompokan metode *Two Step Cluster*

Pengelompokan metode *Two Step Cluster* dilakukan menggunakan ukuran jarak *Euclidean* pada Persamaan (1). Algoritma dari metode *Two Step Cluster* adalah sebagai berikut :

- 1) Pembentukan *cluster* awal

Tahapan ini dilakukan dengan pembentukan *cluster features (CF) Tree*. Pada *CF Tree* data diamati berdasarkan ukuran jarak dan memutuskan data tersebut masuk dalam *cluster* yang telah terbentuk sebelumnya atau memulai *cluster* baru.

- 2) Pembentukan *cluster* optimal

Suatu *cluster* dikatakan optimal jika memiliki jarak antar *cluster* paling jauh dan jarak antar data dalam *cluster* paling dekat. Dalam menentukan jumlah *cluster* optimal, terdapat dua langkah yang perlu dilakukan. Langkah pertama menduga jumlah *cluster* maksimal dengan menghitung kriteria *Bayesian Information Criterion (BIC)* untuk setiap *cluster*. Rumus BIC untuk jumlah *cluster* sebanyak  $K$  *cluster* adalah sebagai berikut:

$$BIC(K) = -2 \sum_{K=1}^{N_K} x_K + m_K \log(n) \quad (2)$$

dengan  $m_K$  adalah ukuran jarak pada *cluster K* yang terbentuk dari tahap *Hierarchical Cluster*,  $x_K$  adalah data pada *cluster K*,  $N_K$  adalah jumlah seluruh *cluster*, dan  $n$  adalah jumlah data. Langkah kedua yaitu menentukan jumlah *cluster* yang optimal yaitu yang memiliki nilai BIC terkecil, nilai rasio perubahan BIC terbesar, dan rasio perubahan jarak terbesar.

8. Menghitung dan membandingkan nilai *Silhouette Index*

Nilai *Silhouette Index* merupakan metode evaluasi yang digunakan untuk menentukan metode terbaik dilihat dari kualitas *cluster* yang dihasilkan dengan kategori sebagai berikut [8]:

**Tabel 1.** Nilai *Silhouette Index*

Nilai	Kategori
$0,75 \leq SI < 1$	Sangat Kuat
$0,5 \leq SI < 0,75$	Kuat
$0,25 \leq SI < 0,5$	Lemah
$0 \leq SI < 0,25$	Sangat Lemah
$-1 \leq SI < 0$	Tidak Terstruktur

Adapun algoritma dari nilai *Silhouette Index* adalah sebagai berikut:

- 1) Menghitung rata-rata jarak dari suatu data dengan data lain yang berada pada *cluster* yang sama disimbolkan dengan  $a(i)$ . Rumus dari  $a(i)$  adalah sebagai berikut:

$$a(i) = \frac{1}{C_A - 1} \sum_{C_A=1}^{C_A-1} d(i, u) \tag{3}$$

dengan  $C_A$  menyatakan jumlah data pada *cluster A* dan  $d(i, u)$  merupakan ukuran jarak *Euclidean* pada *cluster* yang sama.

- 2) Menghitung rata-rata jarak dari suatu data dengan semua data yang berada pada *cluster* lain disimbolkan dengan  $M(i, K)$ . Rumus dari  $M(i, K)$  adalah sebagai berikut:

$$M(i, K) = \frac{1}{C_B} \sum_{C_B=1}^{C_B} d(i, v) \tag{4}$$

dengan  $C_B$  menyatakan jumlah data pada *cluster B* dan  $d(i, v)$  merupakan ukuran jarak *Euclidean* antar *cluster*. Diketahui  $K$  merupakan jumlah *cluster* yang terbentuk, maka jumlah  $M(i, K)$  sesuai dengan banyaknya *cluster* yang

terbentuk. Setelah diperoleh nilai  $M(i, K)$  kemudian dipilih nilai  $M(i, K)$  yang terkecil sebagai nilai  $b(i)$ .

$$b(i) = \min\{M(i, K)\} \tag{5}$$

- 3) Menghitung nilai *Silhouette Index* dengan mensubstitusikan hasil perhitungan pada Persamaan (3) dan Persamaan (5) ke rumus berikut:

$$SI(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{6}$$

Setelah diperoleh nilai *Silhouette Index* dari metode *K-Means* dan *Two Step Cluster* kemudian dibandingkan hasilnya untuk menyimpulkan metode terbaik dalam pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022 dilihat dari nilai *Silhouette Index* yang terbesar.

### 3 HASIL DAN PEMBAHASAN

#### 3.1 Metode K-Means

Pada penelitian ini pengelompokan dengan metode *K-Means* dibentuk menjadi 4 *cluster* menggunakan *Software IBM SPSS Statistics 25*. Tahapan selanjutnya yaitu:

- 1) Penentuan *cluster* awal

**Tabel 2.** Output Nilai *Centroid* Awal

Nilai <i>z-score</i>	<i>Centroid</i> Awal			
	$C_1$	$C_2$	$C_3$	$C_4$
<i>z-score</i> (UHH)	- 0,25935	0,32274	-2,95017	1,43201
<i>z-score</i> (HLS)	0,06926	-0,50686	-0,94166	2,62372
<i>z-score</i> (RLS)	0,67686	-0,57621	-1,41159	2,21364
<i>z-score</i> (PRK)	1,05948	-0,97081	-0,56222	2,65635

Berdasarkan **Tabel 2** diketahui *centroid* awal yang akan digunakan dalam proses iterasi.

- 2) Proses iterasi

Dari hasil *output IBM SPSS Statistics 25*, dapat diketahui bahwa proses iterasi dilakukan sebanyak 5 kali. Proses iterasi dilakukan untuk mendapatkan hasil *cluster* yang optimal.

**Tabel 3.** Pusat *Cluster* Akhir

Nilai <i>z-score</i>	Pusat <i>Cluster</i>			
	$c_1$	$c_2$	$c_3$	$c_4$
<i>z-score</i> (UHH)	0,39303	-0,35693	-2,14109	1,29198
<i>z-score</i> (HLS)	0,16419	-0,68245	-0,80035	2,20251
<i>z-score</i> (RLS)	0,22870	-0,69442	-1,03593	2,17621

Nilai <i>z-score</i>	Pusat <i>Cluster</i>			
	$c_1$	$c_2$	$c_3$	$c_4$
<i>z-score</i> (PRK)	0,26699	-0,79941	-0,54384	2,00473

Berdasarkan **Tabel 3** dapat diketahui bahwa data tersebut masih mengacu pada nilai *z-score* dengan ketentuan:

- Bernilai negatif (–) berarti bahwa data berada di bawah rata-rata total.
- Bernilai positif (+) berarti bahwa data berada di atas rata-rata total.

### 3) Interpretasi *cluster*

**Tabel 4.** Keanggotaan *Cluster K-Means*

<i>Cluster</i>	Jumlah Anggota	Anggota
3 (Rendah)	3	Kab. Wonosobo, Kab. Tegal, Kab. Brebes.
2 (Sedang)	13	Kab. Cilacap, Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Magelang, Kab. Wonogiri, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Temanggung, Kab. Batang, Kab. Pekalongan, Kab. Pemalang.
1 (Tinggi)	15	Kab. Banyumas, Kab. Purworejo, Kab. Boyolali, Kab. Klaten, Kab. Sukoharjo, Kab. Karanganyar, Kab. Sragen, Kab. Pati, Kab. Kudus, Kab. Jepara, Kab. Demak, Kab. Semarang, Kab. Kendal, Kota Pekalongan, Kota Tegal.
4 (Sangat Tinggi)	4	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang.

## 3.2 Metode *Two Step Cluster*

### 1) Pembentukan *cluster* awal

Pembentukan *cluster* awal dilakukan dengan *Software IBM SPSS Statistics 25* menggunakan kriteria BIC yang disajikan pada

**Tabel 5** berikut:

**Tabel 5.** Hasil *Cluster* Awal

Jumlah <i>cluster</i>	BIC	Nilai perubahan BIC	Rasio perubahan BIC	Rasio perubahan jarak
1	123,469			
2	115,271	-8,198	1	3,46
3	124,951	9,68	-1,181	1,377
4	144,414	19,463	-2,374	1,596

Jumlah <i>cluster</i>	BIC	Nilai perubahan BIC	Rasio perubahan BIC	Rasio perubahan jarak
5	170,84	26,426	-3,224	1,18
6	189,481	18,641	-2,274	1,351
7	216,815	27,333	-3,334	0,934
8	244,397	27,582	-3,365	1,187
9	271,496	27,099	-3,306	1,02
10	294,668	23,172	-2,827	1,307
11	321,474	26,806	-3,27	0,958
12	348,498	27,024	-3,297	1,21
13	375,6	27,102	-3,306	1,264
14	402,764	27,164	-3,314	1,269
15	430,568	27,804	-3,392	1,032

Berdasarkan

**Tabel 5** diperoleh estimasi jumlah *cluster* awal yang terbentuk menggunakan kriteria BIC. Terdapat 15 *cluster* disertai dengan nilai BIC, perubahan nilai BIC, rasio perubahan BIC, dan rasio perubahan jarak.

2) Pembentukan *Cluster* optimal

Berdasarkan

**Tabel 5** dapat ditentukan jumlah *cluster* optimal adalah yang mempunyai nilai BIC yang terkecil yaitu sebesar 115,271, rasio perubahan BIC terbesar yaitu sebesar 1, dan rasio perubahan jarak yang terbesar yaitu 3,46 adalah 2 *cluster*, sehingga jumlah *cluster* optimal pada saat 2 *cluster*.

3) Interpretasi *Cluster Two Step*

**Tabel 6.** Keanggotaan *Cluster Two Step*

<i>Cluster</i>	Jumlah Anggota	Anggota
1 (Tinggi)	31	Kab. Cilacap, Kab. Banyumas, Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Purworejo, Kab. Wonosobo, Kab. Magelang, Kab. Boyolali, Kab. Klaten, Kab. Sukoharjo, Kab. Wonogiri, Kab. Karanganyar, Kab. Sragen, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Kudus, Kab. Jepara, Kab. Demak, Kab. Semarang, Kab. Temanggung, Kab. Kendal, Kab. Batang, Kab. Pekalongan, Kab. Pemalang, Kab. Tegal, Kab. Brebes, Kota Pekalongan, Kota Tegal.
2 (Sangat Tinggi)	4	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang.

### 3.3 Perbandingan Nilai *Silhouette Index*

Tahapan awal hasil pengelompokan metode *K-Means* dan *Two Step Cluster* dihitung nilai  $a(i)$  atau nilai rata-rata jarak dari suatu data dengan data lain pada *cluster* yang sama menggunakan Persamaan (3). Tahapan selanjutnya hasil pengelompokan kedua metode *clustering* tersebut dihitung nilai  $M(i, K)$  atau rata-rata jarak data dengan semua data yang berada pada *cluster* yang berbeda menggunakan Persamaan (4). Setelah semua nilai  $M(i, K)$  masing-masing *cluster* diketahui, kemudian dipilih nilai  $M(i, K)$  yang terkecil sebagai nilai  $b(i)$  menggunakan Persamaan (5).

Tahapan terakhir setelah nilai  $a(i)$  dan  $b(i)$  masing-masing *cluster* dari metode *K-Means* dan *Two Step Cluster* diketahui, maka selanjutnya untuk menentukan nilai *Silhouette Index*nya dilakukan dengan menyubstitusikan nilai  $a(i)$  dan  $b(i)$  yang telah diperoleh ke Persamaan (6). Setelah itu kualitas hasil *cluster* metode *K-Means* dan *Two Step Cluster* dapat diketahui dengan cara menghitung rata-rata dari nilai *Silhouette Index* masing-masing metode tersebut kemudian interpretasikan ke dalam Tabel 1. Hasil dari perhitungan rata-rata nilai *Silhouette Index* beserta kategorinya disajikan pada berikut.

**Tabel 7.** Rata-rata Nilai *Silhouette Index*

Metode <i>Clustering</i>	Rata-rata Nilai SI	Kategori
<i>K-Means</i>	0,3763	Lemah
<i>Two Step Cluster</i>	0,5763	Kuat

Berdasarkan **Tabel 7**, dapat diketahui kualitas hasil pengelompokan metode *Two Step Cluster* lebih baik dibandingkan dengan metode *K-Means*. Karena metode *Two Step Cluster* termasuk kategori yang kuat sedangkan metode *K-Means* termasuk kategori yang lemah.

## 4 KESIMPULAN DAN SARAN

Berdasarkan hasil analisis pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022 diperoleh metode *K-Means* terbentuk menjadi 4 *cluster* dan metode *Two Step Cluster* terbentuk menjadi 2 *cluster*. Hasil pengelompokan metode *K-Means* diperoleh *cluster* 1 beranggotakan 13 kabupaten dan 2 kota, *cluster* 2 beranggotakan 13 kabupaten, *cluster* 3 beranggotakan 3 kabupaten, dan *cluster* 4 beranggotakan 4 kota. Hasil pengelompokan metode *Two Step Cluster* diperoleh *cluster* 1 beranggotakan 29 kabupaten dan 2 kota serta *cluster* 2 beranggotakan 4 kota. Pada hasil perbandingan nilai *Silhouette Index* diperoleh metode terbaik dalam

pengelompokan kabupaten/koat di Provinsi Jawa Tengah berdasarkan indikator dari IPM tahun 2022 adalah metode *Two Step Cluster*.

Pada penelitian ini hanya menggunakan nilai *Silhouette Index* sebagai metode validasi. Sementara pada penggunaan data lain dapat berpeluang diperoleh nilai *Silhouette Index* yang sama sehingga sulit menentukan metode terbaik. Oleh karena itu, untuk mengatasi hal tersebut pada penelitian selanjutnya dapat menggunakan beberapa metode evaluasi lainnya secara bersamaan dengan nilai *Silhouette Index*.

## DAFTAR PUSTAKA

- [1] United Nations Development Programme., *Human development report 1990*, Edisi Pertama, United Nations Development Programme (UNDP)., 1990.
- [2] Setiawan, I. N., Kristiani, S. N., Nurarifin, Delyana, S., Setyawati, N., dan Arsyi, F. A., *Indeks Pembangunan Manusia 2022*, Edisi Pertama, Badan Pusat Statistik, Jakarta, 2023.
- [3] Nursetyohadi, D., *Berita Resmi Statistik (Indeks Pembangunan Manusia Jawa Tengah 2022)*, Edisi Pertama, Badan Pusat Statistik Provinsi Jawa Tengah, Semarang, 2022.
- [4] Suhaeni, C., Kurnia, A., dan Ristiyanti., *Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhierarchy, K-Means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil)*, Jurnal Media Infotama, 14(1) (2018), 31–38.
- [5] Asmar, K., *Clustering Data Indeks Pembangunan Manusia di Indonesia menggunakan Self-Organizing Maps*, Jurnal Ilmu Siber (JIS), 1(3) (2023), 89–92.
- [6] Mongi, C. E., *Penggunaan Analisis Two Step Clustering untuk Data Campuran*, Jurnal Matematika dan Aplikasi, 4(1) (2015), 9–19.
- [7] Talakua, M. W., Leleury, Z. A., dan Talluta, A. W., *Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014*, Jurnal Ilmu Matematika dan Terapan, 11(2) (2017), 119–128.
- [8] Siagian, R., Pahala Sirait, P. S., dan Halima, A., *E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model*, Journal of Informatics and Telecommunication Engineering, 5(1) (2021), 21–30.