

**EVALUASI PERFORMA ALGORITMA K-MEANS DAN
HIERARCHICAL CLUSTERING PADA PENGELOMPOKAN
DATA INDEKS PEMBANGUNAN MANUSIA (IPM)
PADA DATA BPS SUMATERA UTARA**

SKRIPSI

DISUSUN OLEH

BAYHAQI MAULANA

NPM. 2009020125



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2026**

**EVALUASI PERFORMA ALGORITMA K-MEANS DAN
HIERARCHICAL CLUSTERING PADA PENGELOMPOKAN
DATA INDEKS PEMBANGUNAN MANUSIA (IPM)
PADA DATA BPS SUMATERA UTARA**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Teknologi Informasi. pada Fakultas Ilmu
Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

BAYHAQI MAULANA

NPM. 2009020125

**PROGRAM STUDI TEKNOLOGI INFROMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

LEMBAR PENGESAHAN

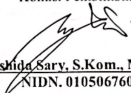
Judul Skripsi : EVALUASI PERFORMA ALGORITMA K-MEANS DAN
HIERARCHICAL CLUSTERING PADA
PENGELOMPOKAN DATA INDEKS PEMBANGUNAN
MANUSIA (IPM) PADA DATA BPS SUMATERA UTARA

Nama Mahasiswa : BAYHAQI MAULANA

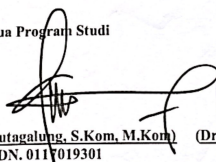
NPM : 2009020125

Program Studi : Teknologi Informasi

Menyetujui
Komisi Pembimbing


(Yoshida Sary, S.Kom., M.Kom.)
NIDN. 0105067603

Ketua Program Studi


(Fatma Sari Hutagalung, S.Kom, M.Kom)
NIDN. 0117019301

Dekan


(Dr. Al-Khulwif Rizki, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**EVALUASI PERFORMA ALGORITMA K-MEANS DAN
HIERARCHICAL CLUSTERING PADA PENGELOMPOKAN
DATA INDEKS PEMBANGUNAN MANUSIA (IPM)
PADA DATA BPS SUMATERA UTARA**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 16 Sep 2026

Yang membuat pernyataan



BAYHAQI MAULANA

NPM. 2009020125

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: BAYHAQI MAULANA
NPM	: 2009020125
Program Studi	: Teknologi Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

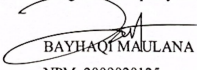
**EVALUASI PERFORMA ALGORITMA K-MEANS DAN
HIERARCHICAL CLUSTERING PADA PENGELOMPOKAN
DATA INDEKS PEMBANGUNAN MANUSIA (IPM)
PADA DATA BPS SUMATERA UTARA**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 16 Sep 2026

Yang membuat pernyataan


BAYHAQI MAULANA
NPM. 2009020125

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : BAYHAQI MAULANA
Tempat dan Tanggal Lahir : KUTACANE, 07 Juli 2001
Alamat Rumah : DESA KAMPUNG BARU
Telepon/Faks/HP : 081328623522
E-mail : Bayhaqimaulana07@gmail.com
Instansi Tempat Kerja :
Alamat Kantor :

DATA PENDIDIKAN

SD	: Negeri, Percontohan	TAMAT: 2013
SMP	: Negeri, Perisai	TAMAT: 2016
SMA	: Negeri, 12 Banda Aceh	TAMAT: 2019

KATA PENGANTAR



Puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan judul “*evaluasi performa algoritma k-means dan hierarchical clustering pada pengelompokan data indeks pembangunan manusia (ipm) pada data bps sumatera utara*” dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman.

Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Akrim, M.Pd. Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom. selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi.
4. Bapak Mhd Basri, S. Si., M.Kom. selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.
5. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom. Ketua Program Studi Teknologi Informasi.
6. Bapak Okvi Nugroho, S.Kom., M.Kom., Sekretaris Program Studi Teknologi Informasi
7. Ibu Yoshida Sary, S.Kom., M.Kom., Selaku Pembimbing skripsi yang senantiasa sabar dalam membimbing, mengarahkan, dan membantu penulis dalam menyusun skripsi ini hingga dapat diselesaikan dengan baik.
8. Seluruh Bapak/Ibu Dosen di Program Studi Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara yang telah banyak memberikan ilmu Teknologi Informasi kepada penulis.

9. Cinta pertama dan panutan penulis, ayahanda Ismadi, S.E., M.M. dan Ibunda Dasimah Netti. Terimakasih atas segala pengorbanan dan tulus kasih yang telah di berikan. Beliau tak kenal lelah mendoakan serta memberikan perhatian dan dukungan hingga penulis mampu menyelesaikan studinya sampai meraih gelar sarjana. Semoga Ayah dan Mama sehat, panjang umur dan bahagia selalu.
10. Terimakasih yang tulus penulis sampaikan kepada saudaraku adinda Khozi Shaid Al-munawar. Kehadiranmu menjadi salah satu alasan terbesar bagi penulis untuk tetap bertahan dan terus berjuang hingga akhirnya mampu menyelesaikan studi ini. Penulis menyadari bahwa tanpa dukungan, perhatian, dan keyakinan yang terus kamu berikan, mungkin penulis tidak akan mampu menyelesaikan pendidikan ini tepat pada semester ke-12. Terima kasih karena telah menjadi tempat untuk berbagi cerita, memberikan kekuatan di saat sulit, dan selalu meyakinkan bahwa setiap perjuangan akan menemukan akhirnya..
11. Kepada seseorang yang pernah bersama penulis dan penulis juga tidak bisa sebutkan namanya. Terimakasih untuk patah hati yang diberikan saat awal penyusunan tugas akhir ini memberikan cukup motivasi untuk terus maju dan berproses menjadi pribadi yang mengerti arti pengalaman, pendewasaan, sabar dan arti kehilangan sebagai bentuk proses penempatan menghadapi dinamika hidup. Terimakasih telah menjadi bagian menyenangkan sekaligus menyakitkan dari pendewasaan ini.
12. *Last but not least*, Kepada diri saya sendiri Bayhaqi Maulana. Terimakasih sudah bertahan sampai detik ini, sudah berusaha menahan sabar, ego, tetap semangat, dan tidak putus asa atas pencapaian dalam menyelesaikan tugas akhir ini.
13. Terimakasih untuk semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu dalam penyelesaian skripsi ini.

EVALUASI PERFORMA ALGORITMA K-MEANS DAN HIERARCHICAL CLUSTERING PADA PENGELOMPOKAN DATA INDEKS PEMBANGUNAN MANUSIA (IPM) PADA DATA BPS SUMATERA UTARA

ABSTRAK

Provinsi Sumatera Utara menghadapi disparitas pembangunan manusia yang signifikan antarwilayah, sehingga diperlukan pemetaan kelompok kabupaten/kota berdasarkan karakteristik pertumbuhan Indeks Pembangunan Manusia (IPM) untuk mendukung perumusan kebijakan yang lebih tepat sasaran. Penelitian ini bertujuan mengevaluasi dan membandingkan performa algoritma K-Means dan Hierarchical Clustering dalam mengelompokkan 33 kabupaten/kota di Provinsi Sumatera Utara berdasarkan data IPM periode 2020–2025. Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS), yang melalui tahap preprocessing dan normalisasi Min-Max Scaling sebelum diproses menggunakan kedua algoritma clustering. Jumlah cluster optimal ditentukan menggunakan Metode Elbow, dan kualitas hasil clustering dievaluasi menggunakan Silhouette Score, Sum of Squared Errors (SSE), dan Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa kedua algoritma menghasilkan tiga cluster dengan komposisi anggota yang identik, yaitu Cluster Tinggi (10 wilayah), Cluster Sedang (19 wilayah), dan Cluster Rendah (4 wilayah, seluruhnya berada di Kepulauan Nias), dengan nilai Silhouette Score sebesar 0,6705 pada kedua algoritma. K-Means memperoleh nilai SSE sebesar 1,1492, sedangkan Hierarchical Clustering memperoleh Davies-Bouldin Index sebesar 0,3942. Meskipun kualitas clustering kedua algoritma setara, K-Means direkomendasikan sebagai metode yang lebih optimal karena memiliki kompleksitas komputasi yang lebih rendah dan lebih efisien untuk diterapkan pada data berukuran besar. Hasil penelitian ini diharapkan dapat menjadi salah satu bahan pertimbangan bagi pemerintah daerah dalam menyusun kebijakan pembangunan manusia yang lebih terarah sesuai karakteristik masing-masing wilayah.

Kata Kunci: Indeks Pembangunan Manusia, K-Means, Hierarchical Clustering, Silhouette Score, Sumatera Utara.

EVALUATION OF K-MEANS AND HIERARCHICAL CLUSTERING
ALGORITHM PERFORMANCE IN GROUPING HUMAN
DEVELOPMENT INDEX (HDI) DATA FROM
BPS NORTH SUMATRA

ABSTRACT

North Sumatra Province faces significant disparities in human development across its regions, necessitating the grouping of districts/cities based on the growth characteristics of the Human Development Index (HDI) to support more targeted policy formulation. This study aims to evaluate and compare the performance of the K-Means and Hierarchical Clustering algorithms in grouping 33 districts/cities in North Sumatra Province based on HDI data for the 2020–2025 period. The data used is secondary data from the Central Statistics Agency (BPS), which underwent preprocessing and Min-Max Scaling normalization before being processed using both clustering algorithms. The optimal number of clusters was determined using the Elbow Method, and the quality of the clustering results was evaluated using the Silhouette Score, Sum of Squared Errors (SSE), and Davies-Bouldin Index (DBI). The results show that both algorithms produced three clusters with identical membership composition, namely the High Cluster (10 regions), Medium Cluster (19 regions), and Low Cluster (4 regions, all located in the Nias Islands), with a Silhouette Score of 0.6705 for both algorithms. K-Means obtained an SSE value of 1.1492, while Hierarchical Clustering obtained a Davies-Bouldin Index of 0.3942. Although the clustering quality of both algorithms is equivalent, K-Means is recommended as the more optimal method due to its lower computational complexity and greater efficiency for large-scale data. The results of this study are expected to serve as a consideration for local governments in formulating more targeted human development policies tailored to the characteristics of each region.

Keywords: Human Development Index, K-Means, Hierarchical Clustering, Silhouette Score, North Sumatra.

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
RIWAYAT HIDUP	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
ABSTRACT.....	viii
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian.....	6
BAB II. LANDASAN TEORI.....	7
2.1 Indeks Pembangunan Manusia (IPM)	7
2.2 Data Mining	7
2.3 Clustering	10
2.4 Metode K-Means	12
2.5 Hierarchical Clustering	14
2.6 Metode Pengujian Silhouette Coefficient	16
2.7 Kajian Penelitian yang Relevan.....	17
2.8 Kerangka Berpikir	17
2.9 Penelitian Terdahulu.....	19
BAB III. METODE PENELITIAN	22
3.1 Jenis Penelitian	22
3.2 Lokasi dan Waktu Penelitian	22
3.3 Jenis dan Sumber Data	25
3.4 Teknik Pengumpulan Data	26
3.5 Teknik Analisis Data	28
3.6 Variabel Penelitian	30
3.7 Tahapan Penelitian	31
3.8 Implementasi Algoritma K-Means	36
3.9 Implementasi Algoritma Hierarchical Clustering.....	42
3.10 Evaluasi Performa Algoritma	46
3.11 Perancangan Sistem.....	50
3.12 Perangkat Penelitian	61
3.13 Class Diagram.....	64
3.14 Rancangan Antarmuka	66
3.15 Alur Penelitian	73
3.16 Contoh Perhitungan Manual	76
BAB IV. HASIL DAN PEMBAHASAN	84
4.1 Hasil Pengumpulan Data	84
4.2 Hasil Preprocessing Data	86

4.3 Hasil Transformasi Data (Normalisasi Min-Max Scaling)	87
4.4 Penentuan Jumlah Cluster Optimal	88
4.5 Hasil Implementasi Algoritma K-Means	89
4.6 Hasil Implementasi Algoritma Hierarchical Clustering	92
4.7 Evaluasi Performa Algoritma	94
4.8 Interpretasi dan Analisis Hasil Clustering	96
4.9 Pembahasan.....	98
4.10 Implementasi Sistem	100
4.11 Tampilan Antarmuka Sistem Web (Seluruh Halaman)	101
BAB V. PENUTUP	110
5.1 Kesimpulan	110
5.2 Saran	111
DAFTAR PUSTAKA	113
LAMPIRAN	134

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	19
Tabel 3.1. Jadwal Penelitian	23
Tabel 3.2. Teknik Analisis Data	28
Tabel 3.3. Metrik Evaluasi Clustering	29
Tabel 3.4. Data Penelitian.....	30
Tabel 3.5. Variabel Penelitian	30
Tabel 3.6. Interpretasi Nilai Silhouette Score.....	47
Tabel 3.7. Interpretasi Nilai Davies-Bouldin Index	49
Tabel 3.8. Contoh Perbandingan Hasil Evaluasi Algoritma	49
Tabel 3.9. Spesifikasi Perangkat Keras.....	61
Tabel 3.10. Perangkat Lunak yang Digunakan	62
Tabel 3.11. Data Sampel untuk Perhitungan Manual	76
Tabel 3.12. Hasil Normalisasi Min-Max Manual.....	77
Tabel 3.13. Matriks Jarak Euclidean Antarwilayah	78
Tabel 3.14. Iterasi 1 - Jarak ke Centroid Awal	79
Tabel 3.15. Iterasi 2 - Jarak ke Centroid yang Diperbarui.....	79
Tabel 3.16. Tahapan Penggabungan Cluster (Ward Linkage) Manual	80
Tabel 3.17. Perhitungan Silhouette Score Manual per Wilayah	82
Tabel 3.18. Nilai Si per Cluster	82
Tabel 4.1. Data IPM Kabupaten/Kota Provinsi Sumatera Utara.....	99
Tabel 4.2. Rekapitulasi Hasil Preprocessing Data	87
Tabel 4.3. Hasil Normalisasi Data (Contoh)	88
Tabel 4.4. Hasil Pengelompokan Kabupaten/Kota Menggunakan K-Means ...	90
Tabel 4.5. Hasil Pengelompokan Hierarchical Clustering	93
Tabel 4.6. Perbandingan K-Means dan Hierarchical Clustering.....	95
Tabel 4.7. Karakteristik Cluster Berdasarkan Hasil K-Means	96

DAFTAR GAMBAR

Gambar 3.1. Grafik Metode Elbow	39
Gambar 3.2. Dendrogram Hierarchical Clustering	44
Gambar 3.3. Flowchart Penelitian	53
Gambar 3.4. Use Case Penelitian	54
Gambar 3.5. Activity Diagram Proses K-Means Clustering	54
Gambar 3.6. Activity Diagram Proses Hierarchical Clustering	55
Gambar 3.7. Activity Diagram Proses Pengelompokan Data	56
Gambar 3.8. Activity Diagram Evaluasi Performa	57
Gambar 3.9. Activity Diagram Perbandingan Hasil	58
Gambar 3.10. Sequence Diagram	59
Gambar 3.11. Class Diagram.....	65
Gambar 3.12. Rancangan Halaman Dashboard.....	66
Gambar 3.13. Rancangan Halaman Dataset	67
Gambar 3.14. Rancangan Halaman Import Data	68
Gambar 3.15. Rancangan Halaman Preprocessing	69
Gambar 3.16. Rancangan Halaman K-Means	70
Gambar 3.17. Rancangan Halaman Hierarchical	70
Gambar 3.18. Rancangan Halaman Visualisasi	71
Gambar 3.19. Rancangan Halaman Evaluasi	72
Gambar 3.20. Rancangan Halaman Laporan.....	73
Gambar 3.21. Alur Penelitian	75
Gambar 4.1. Grafik Metode Elbow (SSE vs k)	88
Gambar 4.2. Scatter Plot Hasil K-Means dan Hierarchical Clustering.....	91
Gambar 4.3. Dendrogram Hierarchical Clustering	92
Gambar 4.4. Tampilan Halaman Evaluasi Performa Algoritma	94
Gambar 4.5. Tampilan Halaman Laporan Sistem Web Clustering.....	101
Gambar 4.6. Tampilan Halaman Dashboard	102
Gambar 4.7. Tampilan Halaman Dataset IPM	103
Gambar 4.8. Tampilan Halaman Import Data	104
Gambar 4.9. Tampilan Halaman Preprocessing	105
Gambar 4.10. Tampilan Halaman K-Means.....	105
Gambar 4.11. Tampilan Halaman Hierarchical.....	106
Gambar 4.12. Tampilan Halaman Hasil Clustering	107
Gambar 4.13. Tampilan Halaman Visualisasi	108
Gambar 4.14. Tampilan Halaman Evaluasi	108
Gambar 4.15. Tampilan Halaman Laporan	109

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pembangunan manusia merupakan salah satu indikator penting dalam mengukur keberhasilan pembangunan suatu wilayah. Paradigma pembangunan modern tidak hanya berorientasi pada pertumbuhan ekonomi, tetapi juga menitikberatkan pada peningkatan kualitas hidup masyarakat secara menyeluruh. Salah satu indikator yang digunakan untuk mengukur tingkat pembangunan manusia adalah Indeks Pembangunan Manusia (IPM). IPM merupakan indeks komposit yang menggambarkan capaian pembangunan manusia berdasarkan tiga dimensi dasar, yaitu kesehatan, pendidikan, dan standar hidup layak (Fadillah et al., 2025)

Di Indonesia, IPM dihitung dan dipublikasikan secara resmi oleh Badan Pusat Statistik (BPS) dengan mengacu pada metodologi yang ditetapkan oleh United Nations Development Programme (UNDP). Penghitungan IPM dilakukan berdasarkan empat indikator utama, yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), serta pengeluaran riil per kapita yang disesuaikan. IPM menjadi salah satu instrumen penting dalam mengevaluasi tingkat kesejahteraan masyarakat karena mampu merepresentasikan kualitas pembangunan manusia secara lebih komprehensif dibandingkan indikator ekonomi semata (Ramadhan & Mutaqin, 2025)

Provinsi Sumatera Utara merupakan salah satu provinsi yang menunjukkan perkembangan IPM yang cukup baik dalam beberapa tahun terakhir. Berdasarkan

data Badan Pusat Statistik Provinsi Sumatera Utara, nilai IPM meningkat dari 71,77 pada tahun 2020 menjadi 76,47 pada tahun 2025. Peningkatan tersebut menunjukkan adanya perbaikan kualitas pembangunan manusia di Provinsi Sumatera Utara. Akan tetapi, peningkatan nilai IPM pada tingkat provinsi belum tentu mencerminkan pemerataan pembangunan manusia pada seluruh kabupaten/kota di wilayah tersebut. Perbedaan kondisi geografis, tingkat pendidikan, akses pelayanan kesehatan, serta aktivitas ekonomi dapat menyebabkan terjadinya kesenjangan pembangunan manusia antarwilayah (Lawelai & Baulete, 2025)

Analisis pembangunan manusia umumnya dilakukan melalui pendekatan statistik deskriptif, seperti analisis tren tahunan dan perbandingan nilai rata-rata antarwilayah. Meskipun mampu memberikan gambaran umum mengenai perkembangan IPM, pendekatan tersebut memiliki keterbatasan dalam mengidentifikasi pola tersembunyi pada data yang bersifat multidimensi. Data IPM kabupaten/kota periode 2020–2025 merupakan data numerik yang memiliki karakteristik runtut waktu (*time series*), sehingga diperlukan metode analisis yang mampu mengelompokkan wilayah berdasarkan tingkat kemiripan karakteristik pertumbuhan IPM.

Salah satu pendekatan yang dapat digunakan untuk mengidentifikasi pola tersebut adalah teknik *clustering* dalam bidang *data mining*. *Clustering* merupakan metode pembelajaran tanpa pengawasan (*unsupervised learning*) yang bertujuan mengelompokkan objek berdasarkan tingkat kemiripan karakteristik yang dimiliki. Melalui teknik ini, wilayah yang memiliki pola pertumbuhan IPM yang serupa dapat dikelompokkan ke dalam satu kluster sehingga diperoleh informasi yang lebih

objektif mengenai kondisi pembangunan manusia di masing-masing daerah (Fadillah et al., 2025)

Metode *clustering* yang banyak digunakan adalah K-Means Clustering. Metode ini bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan kedekatan objek terhadap pusat klaster (*centroid*) dan melakukan proses iteratif hingga mencapai kondisi konvergen. K-Means memiliki keunggulan berupa proses komputasi yang relatif cepat, mudah diimplementasikan, serta mampu menangani data berukuran besar. Namun demikian, metode ini sensitif terhadap penentuan jumlah klaster dan pemilihan *centroid* awal (Oti et al., 2021).

Selain K-Means, metode Hierarchical Clustering juga banyak digunakan dalam analisis pengelompokan data. Metode ini membentuk struktur klaster secara bertahap melalui pendekatan *agglomerative* sehingga menghasilkan representasi hierarkis dalam bentuk *dendrogram*. Keunggulan metode ini terletak pada kemampuannya dalam menggambarkan hubungan kedekatan antarobjek secara bertingkat serta fleksibilitas dalam menentukan jumlah klaster. Akan tetapi, *Hierarchical Clustering* memiliki kompleksitas komputasi yang lebih tinggi dibandingkan metode K-Means (Oti et al., 2021)

Beberapa penelitian terdahulu telah menerapkan teknik clustering untuk mengelompokkan wilayah berdasarkan tingkat pembangunan manusia. Namun, penelitian yang secara khusus membandingkan kinerja metode K-Means Clustering dan Hierarchical Clustering dalam menganalisis pertumbuhan IPM kabupaten/kota di Provinsi Sumatera Utara periode 2020–2025 masih terbatas. Kondisi ini menunjukkan adanya kesenjangan penelitian (*research gap*) yang perlu dikaji lebih lanjut.

Oleh karena itu, penelitian ini dilakukan untuk menganalisis pola pertumbuhan IPM kabupaten/kota di Provinsi Sumatera Utara menggunakan metode K-Means Clustering dan Hierarchical Clustering, serta membandingkan efektivitas kedua metode tersebut dalam menghasilkan struktur klaster yang representatif berdasarkan nilai evaluasi *Silhouette Score* dan *Sum of Squared Errors* (SSE).

Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan penerapan teknik *data mining* pada analisis pembangunan manusia, serta menjadi bahan pertimbangan bagi pemerintah daerah dalam merumuskan kebijakan pembangunan yang lebih tepat sasaran berdasarkan karakteristik pertumbuhan IPM masing-masing kabupaten/kota.

1.2. Rumusan Masalah

Berdasarkan identifikasi dan batasan masalah, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana pola pertumbuhan IPM pada masing-masing Kabupaten/kota di sumatra utara periode 2020–2025?
2. Bagaimana hasil pengelompokan Kabupaten/kota berdasarkan metode K-Means terhadap data pertumbuhan IPM?
3. Bagaimana hasil pengelompokan Kabupaten/kota berdasarkan metode *Hierarchical Clustering* terhadap data pertumbuhan IPM?
4. Metode *clustering* manakah yang lebih optimal dalam menganalisis pertumbuhan IPM Sumatra Utara berdasarkan hasil evaluasi klaster?

1.3. Batasan Masalah

Agar penelitian lebih terarah dan tidak melebar, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data IPM Sumatra Utara periode 2020–2025 yang bersumber dari website resmi Badan Pusat Statistik Sumatra Utara.
2. Unit analisis penelitian adalah 33 Kabupaten/kota yang berada di wilayah sumatra utara.
3. Data yang digunakan adalah nilai IPM tahunan periode 2020–2025.
4. Evaluasi kualitas hasil clustering dilakukan menggunakan Silhouette Score dan Davies-Bouldin Index (DBI).
5. Penelitian ini tidak membahas aspek kebijakan publik secara mendalam, melainkan berfokus pada analisis pengelompokan data.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Menganalisis pola pertumbuhan IPM pada setiap Kabupaten/kota di provinsi sumatra utara periode 2020–2025.
2. Menerapkan metode K-Means dalam pengelompokan data pertumbuhan IPM.
3. Menerapkan metode Hierarchical Clustering dalam pengelompokan data pertumbuhan IPM.
4. Menganalisis perbedaan hasil kedua metode clustering untuk menentukan metode yang lebih optimal berdasarkan evaluasi kualitas klaster.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat secara teoritis dan praktis. Secara teoritis, penelitian ini diharapkan dapat memperkaya pengembangan ilmu pengetahuan di bidang data mining, khususnya mengenai penerapan dan evaluasi algoritma K-Means dan Hierarchical Clustering dalam pengelompokan data Indeks Pembangunan Manusia (IPM). Selain itu, hasil penelitian ini diharapkan dapat menjadi referensi bagi mahasiswa, akademisi, dan peneliti dalam penelitian yang berkaitan dengan metode clustering.

Diharapkan dapat memberikan gambaran mengenai pengelompokan Indeks Pembangunan Manusia (IPM) pada 33 kabupaten/kota di Provinsi Sumatera Utara menggunakan algoritma K-Means dan Hierarchical Clustering. Hasil penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemerintah daerah, Badan Pusat Statistik (BPS), dan instansi terkait dalam mendukung perencanaan serta evaluasi kebijakan pembangunan. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya mengenai evaluasi performa algoritma clustering pada data statistik dan sosial ekonomi.

BAB II

LANDASAN TEORI

2.1 Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia (IPM) merupakan indikator komposit yang digunakan untuk mengukur kualitas pembangunan manusia berdasarkan tiga dimensi utama, yaitu kesehatan, pendidikan, dan standar hidup layak. Di Indonesia, penghitungan IPM dilakukan secara resmi oleh Badan Pusat Statistik berdasarkan metodologi yang mengacu pada standar internasional.

Menurut (Nurdin et al., 2024), IPM menjadi instrumen evaluatif yang tidak hanya mencerminkan capaian pembangunan, tetapi juga berfungsi sebagai dasar perencanaan kebijakan berbasis kesejahteraan masyarakat. Sementara itu, (Permayer & Suppa, 2022) menyatakan bahwa analisis IPM secara longitudinal mampu menggambarkan dinamika pembangunan manusia dan potensi ketimpangan antar wilayah administratif.

Pertumbuhan IPM dari tahun ke tahun mencerminkan perubahan kondisi sosial-ekonomi suatu wilayah. Menurut (Bartak & Jastrzębska, 2022), analisis pertumbuhan indikator pembangunan memerlukan pendekatan kuantitatif yang mampu menangkap pola multidimensi, terutama ketika data memiliki struktur waktu (time series).

2.2 Data Mining

Data mining merupakan proses ekstraksi informasi, pola, atau pengetahuan yang tersembunyi dari kumpulan data berukuran besar dengan memanfaatkan teknik statistik, komputasi, dan *machine learning*. Data mining merupakan bagian

dari proses yang lebih luas, yaitu *Knowledge Discovery in Databases* (KDD), yang meliputi tahapan seleksi data, pembersihan data (*data cleaning*), transformasi data, pemodelan, serta evaluasi hasil (M et al., 2018).

Data mining memiliki peran penting dalam menganalisis data berskala besar dan kompleks karena mampu menemukan pola yang sulit diidentifikasi menggunakan metode statistik konvensional. Teknik ini banyak diterapkan dalam berbagai bidang, termasuk analisis sosial-ekonomi, kesehatan, pendidikan, serta pembangunan wilayah.

Data mining berkaitan erat dengan konsep *machine learning*, khususnya pembelajaran tanpa pengawasan (*unsupervised learning*), yang memungkinkan algoritma menemukan struktur atau pola alami dalam data tanpa adanya label kelas sebelumnya. Salah satu teknik utama dalam pendekatan ini adalah *clustering*, yang digunakan untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristik yang dimiliki.

Secara umum, teknik data mining dapat dikelompokkan menjadi beberapa kategori sebagai berikut.

a. Classification

Classification merupakan teknik data mining yang digunakan untuk memprediksi atau mengelompokkan data ke dalam kelas tertentu berdasarkan data yang telah memiliki label sebelumnya. Beberapa algoritma yang termasuk dalam kategori ini adalah Decision Tree, Naïve Bayes, dan Support Vector Machine.

a. Regression

Regression merupakan teknik yang digunakan untuk memprediksi nilai kontinu berdasarkan hubungan antara variabel independen dan variabel dependen.

Teknik ini banyak digunakan dalam analisis peramalan, seperti prediksi pertumbuhan ekonomi, tingkat inflasi, dan jumlah penduduk.

b. Clustering

Clustering merupakan teknik pengelompokan data berdasarkan tingkat kemiripan karakteristik antarobjek tanpa menggunakan label kelas sebelumnya. Teknik ini termasuk ke dalam metode *unsupervised learning*. Algoritma yang umum digunakan antara lain K-Means, Hierarchical Clustering, dan DBSCAN.

c. Association Rule Mining

Association Rule Mining merupakan teknik yang digunakan untuk menemukan hubungan atau keterkaitan antaritem dalam suatu kumpulan data. Algoritma yang sering digunakan dalam teknik ini adalah Apriori dan FP-Growth.

d. Anomaly Detection

Anomaly Detection digunakan untuk mendeteksi data yang memiliki karakteristik berbeda secara signifikan dibandingkan data lainnya. Teknik ini banyak diterapkan pada sistem deteksi penipuan, keamanan jaringan, dan pengawasan transaksi keuangan.

e. Sequential Pattern Mining

Sequential Pattern Mining merupakan teknik yang digunakan untuk menemukan pola urutan kejadian yang sering muncul dalam suatu periode waktu tertentu. Teknik ini banyak dimanfaatkan dalam analisis perilaku pelanggan dan pola transaksi.

Berdasarkan berbagai teknik data mining tersebut, penelitian ini menggunakan metode *clustering* karena sesuai dengan tujuan penelitian, yaitu mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan

karakteristik pertumbuhan Indeks Pembangunan Manusia (IPM) periode 2020–2025. Melalui teknik clustering, wilayah yang memiliki pola pertumbuhan IPM yang serupa dapat dikelompokkan ke dalam klaster yang sama sehingga diperoleh informasi yang lebih objektif mengenai kondisi pembangunan manusia di masing-masing daerah.

2.3 Clustering

Clustering merupakan teknik *unsupervised learning* yang bertujuan untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristik. Objek dalam satu klaster memiliki kemiripan yang tinggi, sedangkan antar klaster memiliki perbedaan yang signifikan. clustering sangat efektif digunakan dalam pemetaan wilayah karena mampu mengidentifikasi kelompok daerah dengan karakteristik pembangunan yang serupa. Dalam konteks pembangunan manusia, clustering membantu mengelompokkan wilayah berdasarkan pola pertumbuhan indikator kesejahteraan.

Secara konseptual, clustering merupakan bagian dari teknik data mining yang digunakan untuk menemukan struktur alami dalam data tanpa menggunakan label kelas sebelumnya (Thant & Aye, 2020). Clustering bekerja dengan mengukur tingkat kemiripan antar objek, yang umumnya dihitung menggunakan metode jarak seperti Euclidean Distance. Menurut (Dolnicar & Grün, 2022), tujuan utama dari clustering adalah memaksimalkan homogenitas dalam satu klaster (*intra-cluster similarity*) dan memaksimalkan perbedaan antar klaster (*inter-cluster dissimilarity*). Hal ini menjadikan clustering sangat relevan dalam analisis data sosial-ekonomi yang memiliki karakteristik kompleks dan multidimensi.

Dalam praktiknya, clustering banyak digunakan dalam berbagai bidang, seperti segmentasi pasar, analisis citra, serta pemetaan wilayah berbasis indikator pembangunan. Dalam penelitian ini, clustering digunakan untuk mengelompokkan kabupaten/kota berdasarkan karakteristik pertumbuhan IPM sehingga dapat diperoleh pola pengelompokan wilayah yang lebih objektif. Algoritma clustering secara umum dibagi menjadi beberapa kategori, yaitu:

1. *Partition-based clustering*, yaitu metode yang membagi data ke dalam sejumlah klaster tertentu, seperti K-Means.
2. *Hierarchical clustering*, yaitu metode yang membentuk struktur klaster secara bertahap dalam bentuk hierarki.
3. *Density-based clustering*, yaitu metode yang mengelompokkan data berdasarkan kepadatan, seperti DBSCAN.
4. *Model-based clustering*, yaitu metode yang menggunakan pendekatan probabilistik dalam pembentukan klaster.

Setiap metode memiliki karakteristik dan keunggulan masing-masing, tergantung pada struktur data yang dianalisis. Oleh karena itu, pemilihan metode clustering harus disesuaikan dengan tujuan penelitian dan karakteristik dataset. Dalam penelitian ini, metode yang digunakan adalah partition-based clustering (K-Means) dan hierarchical clustering, karena kedua metode tersebut merupakan teknik yang paling umum digunakan dalam analisis data numerik dan memiliki pendekatan yang berbeda dalam membentuk klaster. Perbandingan kedua metode ini diharapkan dapat memberikan hasil pengelompokan yang lebih optimal dalam analisis pertumbuhan IPM pada tingkat kabupaten/kota.

2.4 Metode K-Means

K-Means adalah algoritma clustering berbasis partisi yang membagi data ke dalam k kluster berdasarkan jarak terdekat terhadap centroid. Proses algoritma meliputi inisialisasi centroid, perhitungan jarak, pengelompokan objek, serta pembaruan centroid hingga konvergen (Aldahdooh & Ashour, 2013). Menurut (El Agha & M. Ashour, 2012), K-Means memiliki keunggulan dalam efisiensi komputasi dan kesederhanaan implementasi. Namun demikian, algoritma ini sensitif terhadap pemilihan jumlah kluster dan posisi centroid awal. Penggunaan metode evaluasi seperti Elbow Method dan Silhouette Score sangat penting untuk menghindari bias dalam penentuan jumlah kluster.

Secara konseptual, algoritma K-Means bertujuan untuk meminimalkan variasi dalam satu kluster dengan menggunakan fungsi objektif berupa *Sum of Squared Errors* (SSE) (Liu & Lu, 2015). Fungsi tersebut dapat dinyatakan sebagai berikut:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

di mana:

- a. C_i adalah kluster ke-i
- b. i adalah indeks atau penanda kluster.
- c. x adalah data dalam kluster
- d. μ_i adalah centroid kluster

- e. Nilai SSE yang semakin kecil menunjukkan bahwa jarak antara anggota klaster dengan centroid semakin dekat, sehingga kualitas pengelompokan menjadi lebih baik.

Secara umum, proses kerja algoritma K-Means terdiri atas beberapa tahapan sebagai berikut.

1. Menentukan jumlah klaster (k) :

Pada penelitian ini, jumlah klaster ditentukan menggunakan Metode Elbow berdasarkan nilai Sum of Squared Errors (SSE)..

2. Menghitung jarak data terhadap centroid :

Setiap data dihitung jaraknya terhadap seluruh centroid menggunakan metode Euclidean Distance. Data akan ditempatkan pada klaster yang memiliki jarak paling dekat.

3. Mengelompokkan data :

Setelah seluruh jarak dihitung, setiap objek akan menjadi anggota klaster dengan centroid terdekat.

4. Menghitung ulang centroid:

Centroid baru diperoleh dengan menghitung rata-rata seluruh data yang berada pada masing-masing klaster.

5. Melakukan iterasi hingga konvergen:

Proses perhitungan jarak, pengelompokan data, dan pembaruan centroid dilakukan secara berulang hingga tidak terjadi perubahan anggota klaster atau perubahan centroid sangat kecil. Kondisi ini disebut sebagai keadaan konvergen.

Menurut (Beecks et al., 2022), K-Means merupakan salah satu algoritma clustering yang paling banyak digunakan karena memiliki kompleksitas komputasi yang relatif rendah dan mampu menangani dataset berukuran besar.

Namun demikian, K-Means juga memiliki beberapa keterbatasan, yaitu:

- a. Sensitif terhadap pemilihan jumlah kluster (k)
- b. Sensitif terhadap inisialisasi centroid awal
- c. Tidak efektif pada data yang memiliki bentuk kluster non-sferis
- d. Rentan terhadap pengaruh outlier

Oleh karena itu, dalam penerapannya diperlukan metode evaluasi untuk menentukan jumlah kluster yang optimal dan menilai kualitas hasil pengelompokan. Metode seperti Elbow Method dan Silhouette Coefficient digunakan untuk mengukur seberapa baik struktur kluster yang dihasilkan. Dalam penelitian ini, metode K-Means digunakan untuk mengelompokkan kabupaten/kota berdasarkan karakteristik pertumbuhan IPM, sehingga dapat diperoleh pola pengelompokan wilayah yang memiliki kemiripan tingkat pembangunan manusia.

2.5 Hierarchical Clustering

Hierarchical Clustering membentuk struktur kluster secara bertahap melalui pendekatan agglomerative atau divisive. Metode ini menghasilkan dendrogram yang merepresentasikan hubungan antar objek secara hierarkis. Menurut (Boyko & Tkachyk, 2023), Hierarchical Clustering memberikan fleksibilitas dalam menentukan jumlah kluster karena pengguna dapat memotong dendrogram pada tingkat tertentu sesuai kebutuhan analisis. Namun, metode ini memiliki kompleksitas komputasi yang lebih tinggi dibandingkan K-Means. Dalam

analisis pembangunan wilayah, Hierarchical Clustering sering digunakan untuk melihat struktur kedekatan antar daerah secara bertingkat (Cai et al., 2014).

Secara konseptual, Hierarchical Clustering termasuk dalam teknik unsupervised learning yang membangun struktur klaster berdasarkan matriks jarak antar objek. Pendekatan ini tidak memerlukan penentuan jumlah klaster di awal, sehingga lebih fleksibel dalam eksplorasi data. Terdapat dua pendekatan utama dalam Hierarchical Clustering, yaitu:

1. *Agglomerative (Bottom-Up)*

Setiap objek awalnya dianggap sebagai satu klaster, kemudian secara bertahap digabungkan dengan klaster lain yang memiliki jarak terdekat hingga terbentuk satu klaster besar.

2. *Divisive (Top-Down)*

Seluruh data dianggap sebagai satu klaster, kemudian dibagi menjadi klaster-klaster yang lebih kecil secara bertahap.

Dalam praktiknya, pendekatan agglomerative lebih sering digunakan karena lebih sederhana dan mudah diimplementasikan. Proses penggabungan klaster dalam Hierarchical Clustering bergantung pada metode linkage yang digunakan. Beberapa metode linkage yang umum digunakan antara lain:

1. *Single Linkage* , berdasarkan jarak minimum antar objek
2. *Complete Linkage* , berdasarkan jarak maksimum antar objek
3. *Average Linkage* , berdasarkan rata-rata jarak
4. *Ward's Method* , meminimalkan variansi dalam klaster

Menurut (Martín-Fernández et al., 2020), pemilihan metode linkage sangat memengaruhi bentuk dan kualitas klaster yang dihasilkan. Dendrogram yang

dihasilkan dalam Hierarchical Clustering memberikan visualisasi hubungan antar data, di mana tinggi percabangan menunjukkan tingkat jarak atau ketidaksamaan antar kluster. Dengan memotong dendrogram pada level tertentu, jumlah kluster yang diinginkan dapat ditentukan secara visual. Meskipun memiliki keunggulan dalam interpretasi dan fleksibilitas, Hierarchical Clustering memiliki beberapa keterbatasan, yaitu:

1. Kompleksitas komputasi yang tinggi ($O(n^2)$)
2. Sulit diterapkan pada dataset berukuran besar
3. Sensitif terhadap noise dan outlier

Namun demikian, metode ini tetap relevan dalam analisis data sosial-ekonomi, termasuk dalam penelitian ini, karena mampu menggambarkan hubungan antar wilayah secara lebih detail dan bertingkat. Dalam penelitian ini, Hierarchical Clustering digunakan sebagai metode pembandingan terhadap K-Means untuk mengelompokkan kabupaten/kota berdasarkan karakteristik pertumbuhan IPM, sehingga dapat dianalisis perbedaan struktur kluster yang dihasilkan oleh kedua metode tersebut.

2.6 Metode pengujian *Silhouette Coefficient*

Metode ini merupakan metode evaluasi cluster yang menggabungkan metode cohesian dan separation. Cohesian diukur dengan menghitung seluruh objek yang terdapat dalam sebuah cluster dan separation diukur dengan menghitung jarak rata-rata setiap objek dalam sebuah cluster dengan cluster terdekatnya (Bagirov et al., 2023). Jarak antara data dihitung dengan menggunakan rumus euclidean distance. Untuk menyediakan informasi tentang kualitas hasil clustering pada proses clustering, dapat dihitung silhouette dari masing-masing cluster bahkan

keseluruhan cluster dari hasil kerja suatu algoritma clustering. Nilai silhouette untuk keseluruhan data dengan jumlah cluster k , dapat didefinisikan sebagai $sil(k)$ yang dihitung dengan persamaan yakni rata-rata silhouette value untuk semua cluster.

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i)$$

Keterangan:

$sil(k)$: nilai silhouette semua cluster

K : banyaknya cluster k

$sil(c_i)$: rata rata nilai silhouette

2.7 Kajian Penelitian yang Relevan

K-Means dalam pemetaan pembangunan wilayah dan menemukan bahwa metode tersebut efektif dalam mengidentifikasi wilayah prioritas. Menggunakan Hierarchical Clustering untuk analisis pembangunan daerah dan menghasilkan struktur klaster yang lebih rinci melalui dendrogram. Membandingkan K-Means dan Hierarchical Clustering dan menyimpulkan bahwa karakteristik distribusi data memengaruhi performa masing-masing metode. (Fahmiah & Ningrum, 2023) menunjukkan bahwa clustering pada data pembangunan manusia mampu membantu identifikasi kesenjangan wilayah secara lebih sistematis.

Berdasarkan kajian tersebut, belum ditemukan penelitian yang secara spesifik membandingkan kedua metode clustering dalam analisis pertumbuhan IPM tingkat kabupaten/kota di Sumatra Utara periode 2020–2025. Oleh karena itu, penelitian ini memiliki unsur kebaruan dalam konteks objek dan periode analisis.

2.8 Kerangka Berpikir

Indeks Pembangunan Manusia (IPM) merupakan indikator yang digunakan untuk mengukur kualitas pembangunan manusia pada suatu wilayah berdasarkan dimensi kesehatan, pendidikan, dan standar hidup layak. Data IPM yang dipublikasikan oleh Badan Pusat Statistik menunjukkan bahwa Sumatra Utara mengalami peningkatan nilai IPM dari tahun 2020 hingga 2025. Namun demikian, peningkatan tersebut belum tentu menunjukkan pemerataan pembangunan pada seluruh kabupaten/kota.

Setiap kabupaten/kota di Sumatra Utara memiliki karakteristik sosial dan ekonomi yang berbeda, sehingga dimungkinkan terdapat variasi pola pertumbuhan IPM antar wilayah. Untuk mengidentifikasi pola tersebut secara objektif diperlukan pendekatan berbasis data mining, khususnya teknik clustering.

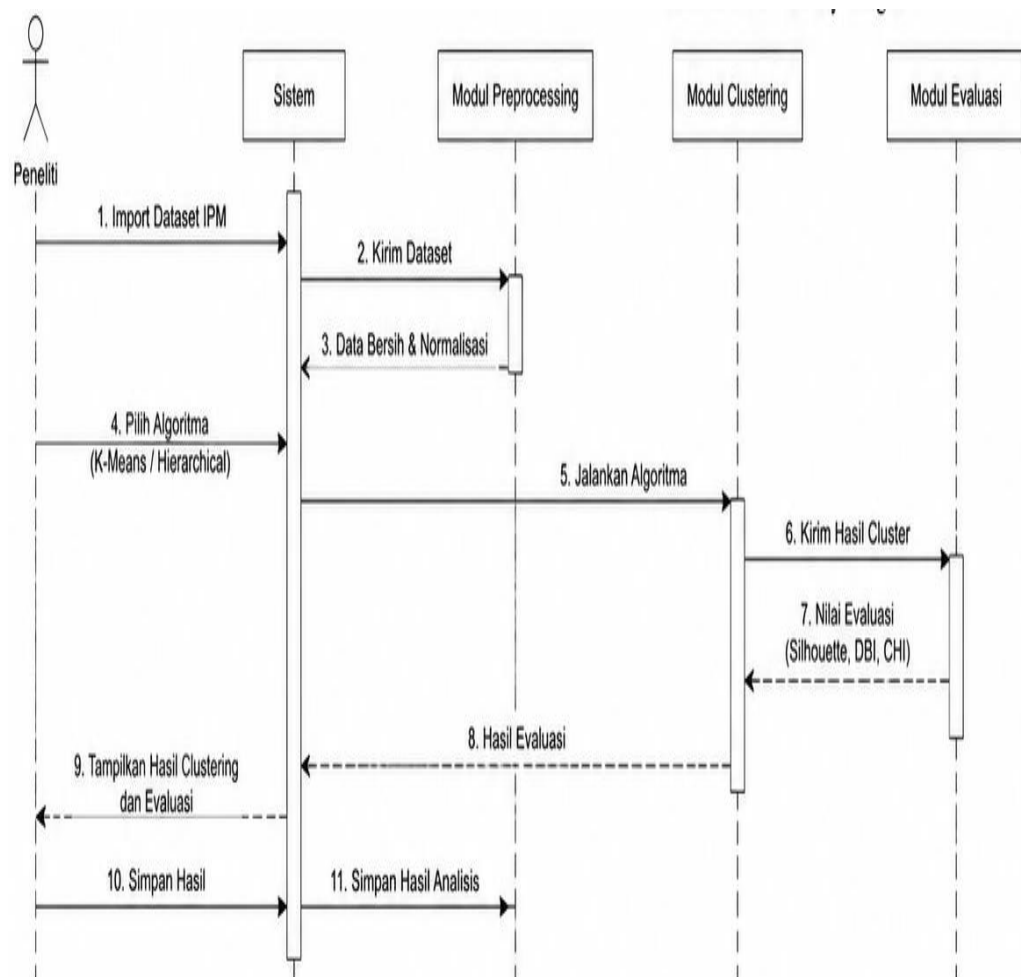
Clustering merupakan metode pengelompokan data berdasarkan tingkat kemiripan karakteristik tanpa menggunakan label kelas. Dalam penelitian ini digunakan dua metode clustering, yaitu K-Means dan Hierarchical Clustering. K-Means mengelompokkan data berdasarkan kedekatan terhadap centroid, sedangkan Hierarchical Clustering membentuk struktur pengelompokan secara bertahap melalui proses aglomeratif.

Data IPM periode 2020–2025 pada tingkat kabupaten/kota digunakan sebagai variabel penelitian. Data tersebut akan melalui tahapan preprocessing, termasuk normalisasi, sebelum dilakukan proses clustering. Selanjutnya Jumlah klaster optimal ditentukan menggunakan Metode Elbow berdasarkan nilai Sum of Squared Errors (SSE). Selanjutnya, kualitas hasil clustering dievaluasi menggunakan Silhouette Score dan Davies-Bouldin Index (DBI).

Hasil pengelompokan dari kedua metode kemudian dibandingkan berdasarkan:

1. Struktur klaster yang terbentuk
2. Konsistensi anggota klaster
3. Nilai evaluasi kualitas klaster

Perbandingan tersebut bertujuan untuk menentukan metode yang lebih optimal dalam menganalisis pertumbuhan IPM pada tingkat kabupaten/kota di Sumatera Utara. Dengan demikian, alur berpikir penelitian ini dapat dijelaskan sebagai berikut:



2.9 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

Nama	Tahun	Judul	Kesimpulan
Sikana, A. M., & Wijayanto, A. W..	2021	Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering.	K-Means adalah metode terbaik dengan klaster optimum k=6: Silhouette Score = 0,6291; CHI = 241,8875; DBI = 0,3038 (tingkat provinsi). Untuk kabupaten/kota: Silhouette = 0,5511; CHI = 1.525,4; DBI = 0,5234.
Satyahadewi, N., Sinaga, S. J., & Perdana, H...	2023	Hierarchical Cluster Analysis of Districts/Cities in North Sumatra Province Based on Human Development Index Indicators Using Pseudo.	Terbentuk 4 klaster dengan metode terbaik Ward's Linkage : Klaster 1 (Medan, Pematang Siantar, Binjai, Padang Sidempuan – IPM sangat tinggi); Klaster 4 (Nias, Nias Selatan, Nias Utara, Nias Barat – IPM sangat rendah).

Sujarwo.	2025	Klasifikasi Indeks Pembangunan Manusia Indonesia Tingkat Sumatera Utara dengan Metode K-Means.	Terbentuk 3 klaster dari 33 kab/kota Sumatera Utara: Klaster 1 (4 kab/kota, rata-rata IPM 64,29 – rendah); Klaster 2 (19 kab/kota, rata-rata IPM 71,61 – sedang); Klaster 3 (10 kab/kota, rata-rata IPM 77,24 – tinggi)
Marbun, E. R., & Chairunisah.	2024	Pengelompokan Kabupaten/Kota di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia Dengan Menggunakan Analisis Cluster Metode K-means.	Terbentuk 3 klaster : tinggi, sedang, rendah. Tren IPM Sumatera Utara meningkat dari 70,78% ke 72,40% Semua indikator IPM tahun 2020–2023 juga naik.
Saputra, N. R., & Muflih, G. Z..	2025	Pengelompokan Wilayah Indonesia Berdasarkan Komponen Indeks Pembangunan Manusia dengan Pendekatan Algoritma K-Means Clustering.	Terbentuk 3 klaster (rendah, sedang, tinggi) dengan DBI = 1,17 yang menunjukkan pemisahan klaster yang baik

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif komparatif dengan pendekatan data mining menggunakan teknik clustering. Penelitian kuantitatif digunakan karena data yang dianalisis berupa data numerik Indeks Pembangunan Manusia (IPM), sedangkan sifat komparatif bertujuan membandingkan performa algoritma K-Means Clustering dan Hierarchical Clustering dalam mengelompokkan data IPM Kabupaten/Kota di Provinsi Sumatera Utara.

Pendekatan penelitian mengacu pada metode Knowledge Discovery in Databases (KDD) yang meliputi tahapan seleksi data, preprocessing, transformasi data, penerapan algoritma clustering, serta evaluasi hasil clustering. Pendekatan ini dipilih karena mampu mengidentifikasi pola kemiripan data secara objektif sehingga dapat menghasilkan pengelompokan yang representatif.

3.2 Lokasi dan Waktu Penelitian

3.2.1 Lokasi Penelitian

Penelitian menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara yang beralamat di Jalan Asrama Nomor 179, Kelurahan Dwikora, Kecamatan Medan Helvetia, Kota Medan, Sumatera Utara. Data yang digunakan berupa data Indeks Pembangunan Manusia (IPM) Kabupaten/Kota periode 2020–2025 yang diperoleh melalui website resmi BPS Provinsi Sumatera Utara

3.2.2 Waktu Penelitian

Penelitian dilaksanakan selama Februari 2026 sampai Juli 2026, yang meliputi kegiatan studi literatur, penyusunan proposal, pengumpulan data, preprocessing data, implementasi algoritma K-Means dan Hierarchical Clustering, evaluasi hasil, serta penyusunan laporan skripsi.

Tabel Penelitian 3.1

No	Kegiatan Penelitian	Februari			Maret				April				Mei				Juni			
		2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Studi literatur																			
2	Penyusunan Proposal																			
3	Pengumpulan Data																			
4	Preprocessing Data																			
5	Implementasi K-Means																			
6	Implementasi Hierarchical Clustering																			
7	Evaluasi Dan Analisis Hasil																			

3.2.3 Objek Penelitian

Objek penelitian merupakan sasaran utama yang digunakan dalam proses analisis untuk mencapai tujuan penelitian. Pada penelitian ini, objek yang digunakan adalah data Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Sumatera Utara yang dipublikasikan oleh Badan Pusat Statistik (BPS).

Data IPM dipilih karena merupakan salah satu indikator yang digunakan untuk mengukur kualitas pembangunan manusia pada suatu wilayah. Nilai IPM

dibentuk oleh beberapa indikator utama yang mencerminkan aspek kesehatan, pendidikan, dan standar hidup masyarakat. Oleh karena itu, data IPM dapat digunakan untuk mengidentifikasi karakteristik pembangunan manusia pada masing-masing kabupaten/kota melalui proses clustering.

Variabel yang digunakan dalam penelitian ini adalah nilai Indeks Pembangunan Manusia (IPM) sebagai indikator komposit pembangunan manusia. Nilai IPM tahunan pada periode 2020–2025 (enam titik waktu) digunakan sebagai atribut dalam proses clustering, sehingga pengelompokan yang dihasilkan merefleksikan pola pertumbuhan IPM setiap kabupaten/kota dari waktu ke waktu, bukan hanya kondisi pada satu tahun tertentu.

Nilai IPM komposit dipilih sebagai atribut tunggal karena merupakan indikator ringkas yang telah merangkum tiga dimensi pembangunan manusia (kesehatan, pendidikan, dan standar hidup layak) ke dalam satu angka, sehingga cukup representatif untuk mengidentifikasi pola pengelompokan wilayah tanpa memerlukan data komponen penyusun IPM secara terpisah. Keenam nilai tahunan tersebut digunakan sebagai atribut dalam proses clustering menggunakan algoritma K-Means dan Hierarchical Clustering sehingga dapat menghasilkan kelompok kabupaten/kota yang memiliki karakteristik pertumbuhan pembangunan manusia yang relatif sama.

3.3 Jenis dan Sumber Data

3.3.1. Jenis Data

Jenis data yang digunakan dalam penelitian ini adalah data kuantitatif, yaitu data yang disajikan dalam bentuk angka sehingga dapat diolah dan dianalisis

menggunakan metode statistik maupun teknik *data mining*. Data penelitian berupa nilai Indeks Pembangunan Manusia (IPM) pada 33 kabupaten/kota di Provinsi Sumatera Utara selama periode tahun 2020–2025.

Berdasarkan dimensi waktunya, data yang digunakan termasuk data runtut waktu (*time series*), karena memuat nilai IPM pada beberapa periode pengamatan, yaitu dari tahun 2020 hingga tahun 2025. Data *time series* digunakan untuk menggambarkan perkembangan nilai IPM pada setiap kabupaten/kota selama periode penelitian sehingga dapat dianalisis pola pertumbuhannya.

Selanjutnya, data tersebut digunakan sebagai masukan dalam proses *clustering* menggunakan algoritma K-Means dan Hierarchical Clustering untuk mengelompokkan kabupaten/kota berdasarkan karakteristik pertumbuhan Indeks Pembangunan Manusia (IPM).

Dengan demikian, berdasarkan karakteristiknya, data yang digunakan dalam penelitian ini terdiri atas:

1. Data kuantitatif.
2. Data runtut waktu periode 2020–2025.

3.3.2. Sumber Data

Sumber data dalam penelitian ini merupakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara periode 2020–2025. Data tersebut diperoleh melalui publikasi statistik yang diterbitkan oleh BPS Provinsi Sumatera Utara dan tersedia dalam bentuk laporan tahunan maupun basis data statistik pada website resmi BPS Provinsi Sumatera Utara.

Data Indeks Pembangunan Manusia (IPM) yang dipublikasikan oleh BPS memiliki tingkat validitas dan reliabilitas yang tinggi karena disusun berdasarkan

metodologi statistik yang telah terstandarisasi secara nasional. Oleh karena itu, data tersebut layak digunakan sebagai sumber data dalam penelitian ini.

Seluruh data yang digunakan merupakan data IPM kabupaten/kota di Provinsi Sumatera Utara periode 2020–2025 yang diperoleh dari publikasi resmi BPS dan selanjutnya diolah menggunakan metode *K-Means Clustering* dan *Hierarchical Clustering* sesuai dengan tahapan penelitian yang telah ditetapkan.

3.4 Teknik Pengumpulan Data

Teknik pengumpulan data merupakan tahapan yang dilakukan untuk memperoleh data yang diperlukan dalam penelitian. Data yang diperoleh harus memiliki kualitas yang baik agar proses analisis menghasilkan informasi yang akurat. Dalam penelitian ini digunakan dua teknik pengumpulan data, yaitu dokumentasi dan studi literatur.

3.4.1 Dokumentasi

Teknik dokumentasi dilakukan dengan mengumpulkan data sekunder yang dipublikasikan oleh Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data diperoleh melalui publikasi statistik maupun data digital yang tersedia pada situs resmi BPS.

Data yang dikumpulkan berupa nilai Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Sumatera Utara periode 2020–2025. Seluruh data tersebut kemudian disusun dalam bentuk tabel dan dipersiapkan untuk proses pengolahan menggunakan sistem berbasis web yang dibangun dengan JavaScript.

Teknik dokumentasi dipilih karena data yang diperoleh berasal dari sumber resmi pemerintah sehingga memiliki tingkat keakuratan, validitas, dan reliabilitas yang tinggi.

3.4.2 Studi Literatur

Studi literatur dilakukan untuk memperoleh landasan teori yang mendukung pelaksanaan penelitian. Referensi diperoleh dari berbagai sumber ilmiah seperti buku, jurnal nasional, jurnal internasional, prosiding seminar, serta artikel ilmiah yang berkaitan dengan topik penelitian.

Kajian literatur difokuskan pada beberapa materi, yaitu:

- konsep Data Mining;
- metode Knowledge Discovery in Databases (KDD);
- algoritma K-Means;
- algoritma Hierarchical Clustering;
- evaluasi kualitas clustering menggunakan Silhouette Score;
- evaluasi menggunakan Davies-Bouldin Index;
- penelitian terdahulu mengenai pengelompokan data Indeks Pembangunan Manusia (IPM).

Hasil studi literatur digunakan sebagai dasar dalam menentukan metode penelitian, merancang proses analisis, memilih algoritma clustering yang digunakan, serta membandingkan hasil penelitian dengan penelitian-penelitian sebelumnya.

3.5 Teknik Analisi Data

Teknik analisis data pada penelitian ini dilakukan menggunakan pendekatan Knowledge Discovery in Databases (KDD). Analisis dimulai dengan pengumpulan data IPM dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara, kemudian dilakukan proses pemilihan atribut (*data selection*), pembersihan data (*data*

preprocessing), dan normalisasi data (*data transformation*). Selanjutnya, data dikelompokkan menggunakan algoritma K-Means dan Hierarchical Clustering. Hasil pengelompokan dievaluasi menggunakan Silhouette Score dan Davies-Bouldin Index (DBI). Tahap terakhir adalah membandingkan performa kedua algoritma dan menginterpretasikan hasil clustering untuk memperoleh kesimpulan penelitian.

Tabel 3.2 Teknik Analisis Data.

No	Tahapan Analisis	Kegiatan yang Dilakukan	Output
1	Pengumpulan Data	Mengumpulkan data Indeks Pembangunan Manusia (IPM) Kabupaten/Kota Provinsi Sumatera Utara dari Badan Pusat Statistik (BPS).	Dataset IPM
2	Data Selection	Memilih atribut yang digunakan dalam penelitian, yaitu nilai IPM tahunan periode 2020–2025.	Data yang relevan
3	Data Preprocessing	Memeriksa missing value, data duplikat, tipe data, serta memperbaiki data yang tidak sesuai.	Data bersih
4	Data Transformation	Melakukan normalisasi data menggunakan metode Min-Max Scaling atau StandardScaler agar seluruh variabel memiliki skala yang sama.	Data terstandarisasi
5	Penentuan Jumlah Cluster	Menentukan jumlah cluster optimal menggunakan Metode Elbow pada algoritma K-Means.	Nilai cluster optimal (k)

No	Tahapan Analisis	Kegiatan yang Dilakukan	Output
6	Implementasi K-Means	Mengelompokkan data menggunakan algoritma K-Means berdasarkan jumlah cluster yang telah ditentukan.	Hasil cluster K-Means
7	Implementasi Hierarchical Clustering	Mengelompokkan data menggunakan metode Agglomerative Hierarchical Clustering dengan Ward Linkage.	Hasil cluster Hierarchical
8	Evaluasi Performa	Menghitung nilai Silhouette Score dan Davies-Bouldin Index (DBI).	Nilai evaluasi masing-masing algoritma
9	Perbandingan Algoritma	Membandingkan hasil evaluasi K-Means dan Hierarchical Clustering untuk menentukan algoritma dengan performa terbaik.	Algoritma terbaik
10	Interpretasi Hasil	Menganalisis karakteristik setiap cluster dan menarik kesimpulan berdasarkan hasil pengelompokan.	Kesimpulan penelitian

Tabel 3.3 Metrik Evaluasi Clustering

Metrik Evaluasi	Fungsi	Kriteria Penilaian
Silhouette Score	Mengukur tingkat kemiripan data dalam cluster dibandingkan dengan cluster lain.	Semakin mendekati 1 , semakin baik.
Davies-Bouldin Index (DBI)	Mengukur rasio penyebaran dalam cluster terhadap jarak antarcluster.	Semakin kecil nilainya, semakin baik.

Tabel 3.4 Data Penelitian

Tahun	Kabupaten/Kota	UHH (Tahun)	Harapan Lama Sekolah (Tahun)	Rata-rata Lama Sekolah (Tahun)	Pengeluaran per Kapita Disesuaikan (Ribu Rupiah)	IPM
2020	Nias	69,75	12,57	5,36	6.898	61,93
	Mandailing Natal	62,60	13,86	8,52	9.684	66,79
	Tapanuli Selatan	64,91	13,24	9,28	11.236	70,12
2021	Nias	71,29	12,84	5,64	6.995	63,37
	Mandailing Natal	71,00	13,87	8,11	9.771	71,31
	Tapanuli Selatan	71,77	13,72	9,26	11.304	75,18
2022	Nias	71,52	13,04	5,88	7.196	64,30
	Mandailing Natal	71,35	13,85	8,73	10.201	72,17
	Tapanuli Selatan	71,35	13,37	9,34	11.727	79,96
2023	Nias	71,74	13,30	6,14	7.301	65,15
	Mandailing Natal	71,72	13,86	8,84	10.251	72,65
	Tapanuli Selatan	71,61	13,58	9,51	11.829	74,58
2024	Nias	71,89	13,81	6,40	7.698	66,13
	Mandailing Natal	71,88	13,87	9,11	10.644	73,44
	Tapanuli Selatan	71,77	13,72	9,56	12.213	75,18
2025	Nias	72,04	13,32	6,41	8.115	66,79
	Mandailing Natal	72,06	14,10	9,17	11.100	74,26
	Tapanuli Selatan	71,97	13,90	9,60	12.557	75,78

3.6 Variabel Penelitian

Variabel penelitian merupakan atribut atau karakteristik yang digunakan sebagai dasar dalam proses analisis dan pengelompokan data. Pada penelitian ini, variabel yang digunakan adalah nilai Indeks Pembangunan Manusia (IPM) tahunan yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Variabel tersebut bertipe numerik sehingga dapat langsung digunakan dalam proses clustering setelah melalui tahap preprocessing dan normalisasi data.

Variabel yang digunakan dalam penelitian ini disajikan pada Tabel 3.2.

Tabel 3.5 Variabel Penelitian.

No	Variabel	Jenis Data	Keterangan
1	Indeks Pembangunan Manusia (IPM)	Numerik	Nilai indeks yang menggambarkan tingkat pembangunan manusia pada setiap kabupaten/kota.

3.7 Tahapan Penelitian

Tahapan penelitian merupakan rangkaian proses yang dilakukan secara sistematis untuk mencapai tujuan penelitian. Pada penelitian ini digunakan metode Knowledge Discovery in Databases (KDD) sebagai kerangka kerja dalam mengolah data hingga menghasilkan informasi yang bermanfaat. Metode KDD dipilih karena menyediakan tahapan yang terstruktur, mulai dari pemilihan data, pembersihan data, transformasi data, penerapan algoritma data mining, hingga evaluasi hasil.

Melalui penerapan metode KDD, data Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Sumatera Utara dapat diolah secara sistematis sehingga menghasilkan pengelompokan wilayah berdasarkan karakteristik pola pertumbuhan nilai IPM. Hasil pengelompokan dari algoritma K-Means dan Hierarchical Clustering selanjutnya dievaluasi untuk mengetahui algoritma yang memberikan kualitas cluster terbaik.

Adapun tahapan penelitian menggunakan metode KDD pada penelitian ini terdiri atas Data Selection, Data Preprocessing, Data Transformation, Data Mining, dan Evaluation.

3.7.1 Data Selection

Data Selection merupakan tahap awal dalam metode KDD yang bertujuan memilih data yang relevan dengan tujuan penelitian. Pada tahap ini dilakukan proses identifikasi dan pemilihan data Indeks Pembangunan Manusia (IPM) yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara.

Data yang digunakan mencakup seluruh kabupaten dan kota di Provinsi Sumatera Utara sesuai periode penelitian yang telah ditentukan. Dari data yang

tersedia, dipilih atribut yang berhubungan langsung dengan pembentukan nilai IPM, sedangkan atribut yang tidak memiliki pengaruh terhadap proses clustering, seperti kode wilayah atau informasi administratif lainnya, tidak digunakan dalam proses analisis.

Atribut yang digunakan dalam penelitian ini adalah nilai Indeks Pembangunan Manusia (IPM) tahunan pada periode 2020–2025 (enam atribut, satu untuk setiap tahun), yang mencerminkan pola pertumbuhan pembangunan manusia setiap kabupaten/kota dari waktu ke waktu.

3.7.2 Data Preprocessing

Data Preprocessing merupakan proses persiapan data sebelum dilakukan analisis menggunakan algoritma clustering. Tahap ini bertujuan meningkatkan kualitas data sehingga hasil pengelompokan menjadi lebih akurat dan dapat dipercaya.

Beberapa proses yang dilakukan pada tahap preprocessing adalah sebagai berikut.

1. Pemeriksaan Missing Value

Tahap pertama dilakukan pemeriksaan terhadap seluruh atribut untuk mengetahui apakah terdapat data yang kosong atau tidak memiliki nilai. Apabila ditemukan missing value, maka dilakukan penanganan sesuai kondisi data, seperti penghapusan data yang tidak lengkap atau pengisian nilai menggunakan metode yang sesuai agar tidak memengaruhi hasil analisis.

2. Pemeriksaan Data Duplikat

Selanjutnya dilakukan pemeriksaan terhadap kemungkinan adanya data yang tercatat lebih dari satu kali. Data yang teridentifikasi sebagai duplikat akan

dihapus sehingga setiap kabupaten/kota hanya memiliki satu data yang mewakili kondisi sebenarnya.

3. Pemeriksaan Tipe Data

Setiap atribut diperiksa untuk memastikan bahwa seluruh variabel numerik memiliki format data yang benar sehingga dapat diproses menggunakan algoritma clustering. Apabila terdapat ketidaksesuaian tipe data, maka dilakukan konversi ke format numerik.

Melalui proses preprocessing, data menjadi lebih bersih, konsisten, dan siap digunakan pada tahap transformasi data.

3.7.3 Data Transformation

Data Transformation merupakan proses mengubah data ke dalam bentuk yang sesuai untuk proses clustering. Pada penelitian ini setiap nilai IPM antartahun (2020–2025) memiliki rentang yang bervariasi antarwilayah, misalnya sebagian wilayah berada pada kisaran 60-an sedangkan wilayah lain pada kisaran 80-an. Perbedaan skala tersebut dapat menyebabkan variabel dengan nilai yang lebih besar mendominasi proses pengelompokan.

Untuk mengatasi perbedaan skala tersebut dilakukan proses normalisasi data menggunakan metode Min-Max Scaling. Metode ini mengubah seluruh nilai atribut ke dalam rentang 0 sampai 1 sehingga setiap variabel memiliki kontribusi yang seimbang dalam proses clustering.

Rumus normalisasi Min-Max adalah sebagai berikut.

$$X' = \frac{(X - X_{\min})}{(X_{\max} - X_{\min})}$$

Keterangan:

- (X') = nilai hasil normalisasi.
- (X) = nilai data awal.
- $(X_{\{min\}})$ = nilai minimum pada atribut.
- $(X_{\{max\}})$ = nilai maksimum pada atribut.

Hasil transformasi berupa dataset yang telah dinormalisasi dan siap digunakan sebagai masukan bagi algoritma K-Means maupun Hierarchical Clustering.

3.7.4 Data Mining

Tahap Data Mining merupakan proses utama dalam penelitian yang bertujuan menemukan pola atau informasi tersembunyi dari data melalui penerapan algoritma clustering.

Pada penelitian ini digunakan dua algoritma clustering, yaitu K-Means dan Hierarchical Clustering. Kedua algoritma diterapkan pada dataset yang sama sehingga hasil pengelompokan dapat dibandingkan secara objektif.

Algoritma K-Means mengelompokkan data berdasarkan kedekatan setiap objek terhadap centroid sehingga menghasilkan sejumlah cluster sesuai nilai (k) yang telah ditentukan. Sebaliknya, Hierarchical Clustering membentuk struktur pengelompokan secara bertingkat melalui proses penggabungan cluster yang memiliki tingkat kemiripan paling tinggi.

Implementasi kedua algoritma dilakukan menggunakan sistem berbasis web yang dibangun dengan bahasa pemrograman JavaScript, memanfaatkan pustaka *Chart.js* dan *PapaParse* untuk mempermudah proses pembacaan data, pembentukan cluster, serta visualisasi hasil.

3.7.5 Evaluation

Tahap Evaluation dilakukan untuk menilai kualitas hasil clustering yang dihasilkan oleh algoritma K-Means dan Hierarchical Clustering. Evaluasi bertujuan mengetahui algoritma yang memberikan kualitas pengelompokan terbaik berdasarkan tingkat kedekatan data di dalam cluster dan tingkat pemisahan antarcluster.

Pada penelitian ini digunakan dua metrik evaluasi, yaitu Silhouette Score dan Davies-Bouldin Index (DBI).

Silhouette Score digunakan untuk mengukur tingkat kesesuaian suatu objek terhadap cluster tempat objek tersebut berada. Nilai Silhouette Score berada pada rentang -1 hingga 1. Semakin mendekati nilai 1 menunjukkan bahwa kualitas cluster semakin baik.

Davies-Bouldin Index digunakan untuk mengukur tingkat kesamaan antarcluster. Semakin kecil nilai DBI menunjukkan bahwa cluster yang terbentuk semakin baik karena memiliki jarak antarcluster yang semakin besar dan tingkat penyebaran data di dalam cluster yang semakin kecil.

Hasil evaluasi dari kedua algoritma kemudian dibandingkan untuk menentukan algoritma yang memiliki performa terbaik dalam mengelompokkan data Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Sumatera Utara.

3.7.6 Interpretasi Hasil

Tahap terakhir dalam metode KDD adalah interpretasi hasil. Pada tahap ini dilakukan analisis terhadap hasil pengelompokan yang diperoleh dari algoritma K-

Means dan Hierarchical Clustering. Setiap cluster dianalisis berdasarkan karakteristik nilai indikator IPM yang dimilikinya, sehingga dapat diketahui kelompok wilayah dengan tingkat pembangunan manusia yang relatif tinggi, sedang, maupun rendah.

Selain membandingkan kualitas cluster berdasarkan metrik evaluasi, interpretasi hasil juga dilakukan untuk menjelaskan makna dari setiap kelompok yang terbentuk serta memberikan gambaran mengenai kondisi pembangunan manusia di Provinsi Sumatera Utara. Hasil interpretasi ini diharapkan dapat menjadi informasi yang bermanfaat bagi pemerintah daerah, peneliti, maupun pihak terkait dalam memahami pola pembangunan manusia berdasarkan data statistik yang tersedia.

3.8 Implementasi Algoritma K-Means

Algoritma **K-Means** merupakan salah satu metode clustering non-hierarki (*partitioning clustering*) yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik antarobjek. Algoritma ini bekerja dengan cara membagi data ke dalam sejumlah cluster yang telah ditentukan sebelumnya. Setiap data akan ditempatkan pada cluster yang memiliki jarak paling dekat terhadap titik pusat cluster (*centroid*).

Pada penelitian ini, algoritma K-Means digunakan untuk mengelompokkan data Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Sumatera Utara berdasarkan nilai IPM tahunan periode 2020–2025 sebagai atribut. Seluruh proses implementasi dilakukan menggunakan sistem berbasis web yang dibangun dengan bahasa pemrograman JavaScript.

Tahapan implementasi algoritma K-Means pada penelitian ini meliputi penentuan jumlah cluster, inisialisasi centroid, perhitungan jarak antar data dengan centroid, pembentukan cluster, pembaruan centroid, serta proses iterasi hingga diperoleh hasil clustering yang stabil.

3.8.1 Penentuan Jumlah Cluster (k)

Tahap pertama dalam algoritma K-Means adalah menentukan jumlah cluster yang akan dibentuk. Penentuan jumlah cluster sangat berpengaruh terhadap kualitas hasil pengelompokan. Apabila jumlah cluster terlalu sedikit, maka data yang memiliki karakteristik berbeda dapat tergabung dalam satu kelompok. Sebaliknya, apabila jumlah cluster terlalu banyak, maka hasil pengelompokan menjadi terlalu rinci sehingga sulit untuk diinterpretasikan.

Pada penelitian ini, jumlah cluster ditentukan menggunakan Metode Elbow. Metode ini dilakukan dengan menghitung nilai Within Cluster Sum of Squares (WCSS) pada beberapa nilai (k), kemudian menampilkan hasilnya dalam bentuk grafik. Nilai (k) yang dipilih adalah titik yang membentuk pola menyerupai siku (*elbow*), yaitu titik ketika penurunan nilai WCSS mulai melambat. Titik tersebut menunjukkan jumlah cluster yang dianggap paling optimal karena mampu memberikan keseimbangan antara kualitas pengelompokan dan kompleksitas model.

3.8.2 Inisialisasi Centroid

Setelah jumlah cluster ditentukan, langkah berikutnya adalah melakukan inisialisasi centroid. Centroid merupakan titik pusat dari setiap cluster yang akan digunakan sebagai acuan dalam proses pengelompokan data.

Pada tahap awal, algoritma memilih sejumlah centroid secara acak sesuai dengan jumlah cluster yang telah ditentukan. Setiap centroid memiliki nilai untuk seluruh atribut yang digunakan dalam penelitian. Posisi centroid ini akan terus diperbarui selama proses iterasi hingga mencapai kondisi yang stabil.

3.8.3 Perhitungan Jarak Menggunakan Euclidean Distance

Setelah centroid ditentukan, setiap data dihitung jaraknya terhadap seluruh centroid menggunakan metode Euclidean Distance. Perhitungan jarak bertujuan mengetahui centroid yang memiliki kedekatan paling tinggi dengan suatu data.

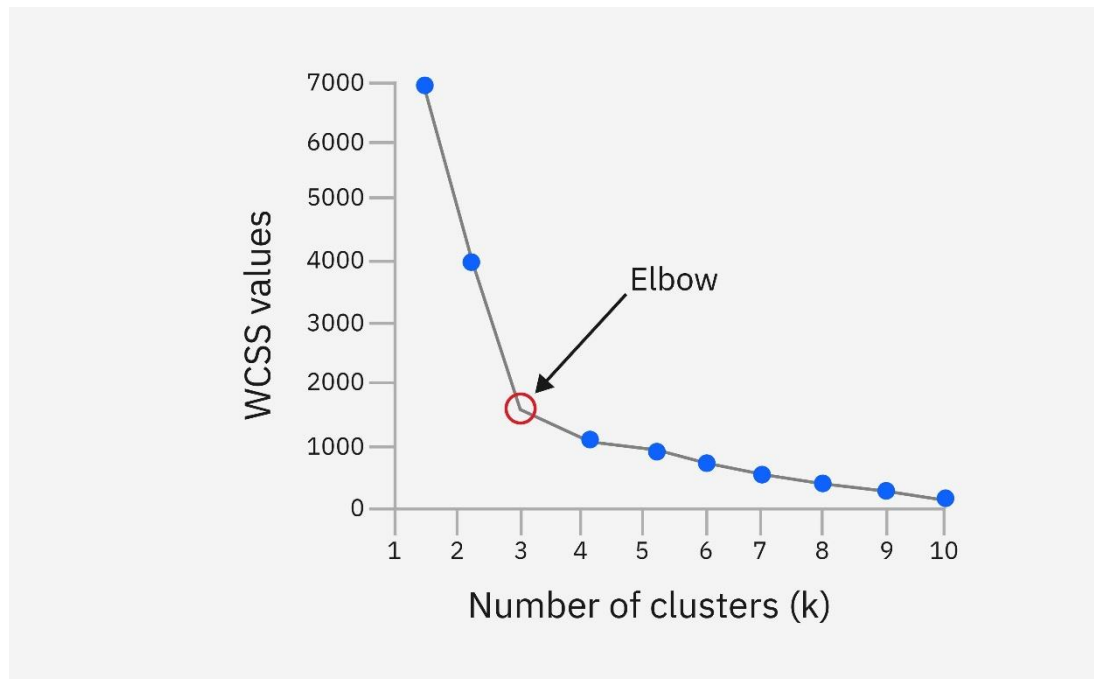
Rumus Euclidean Distance yang digunakan adalah:

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

Keterangan:

- $d(x, y)$ = jarak antara dua data.
- (x_i) = nilai atribut ke-(i) pada data pertama.
- (y_i) = nilai atribut ke-(i) pada centroid.
- (n) = jumlah atribut yang digunakan.

Semakin kecil nilai jarak yang diperoleh, maka semakin tinggi tingkat kemiripan antara data dengan centroid tersebut.



Gambar 3.1

Gambar 3.1 memperlihatkan hasil penerapan Metode Elbow untuk menentukan jumlah cluster terbaik pada algoritma K-Means. Berdasarkan grafik, nilai WCSS menurun tajam hingga $k = 3$, kemudian penurunannya menjadi lebih landai. Kondisi tersebut menunjukkan bahwa $k = 3$ merupakan jumlah cluster yang optimal karena mampu memberikan keseimbangan antara kualitas pengelompokan dan efisiensi model. Oleh karena itu, penelitian ini menggunakan 3 cluster sebagai parameter pada proses K-Means.

3.8.4 Pembentukan Cluster

Berdasarkan hasil perhitungan Euclidean Distance, setiap data ditempatkan ke dalam cluster yang memiliki jarak paling dekat terhadap centroid. Proses ini dilakukan untuk seluruh data hingga semua kabupaten/kota memiliki anggota cluster masing-masing.

Hasil dari tahap ini berupa kumpulan data yang telah terbagi ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik indikator IPM.

3.8.5 Pembaruan Centroid

Setelah seluruh data dikelompokkan, dilakukan proses pembaruan centroid. Nilai centroid baru dihitung berdasarkan rata-rata seluruh data yang berada pada masing-masing cluster.

Perhitungan centroid baru dilakukan menggunakan persamaan:

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i$$

Keterangan:

- (C_j) = centroid pada cluster ke-(j).
- (n_j) = jumlah data pada cluster ke-(j).
- (x_i) = nilai setiap data dalam cluster.

Centroid hasil perhitungan ini kemudian digunakan kembali sebagai titik pusat pada iterasi berikutnya.

3.8.6 Proses Iterasi

Setelah centroid diperbarui, algoritma kembali menghitung jarak seluruh data terhadap centroid yang baru. Proses pengelompokan dan pembaruan centroid dilakukan secara berulang hingga tidak terjadi perubahan anggota cluster atau perubahan nilai centroid sudah sangat kecil. Kondisi tersebut disebut sebagai konvergen, yaitu keadaan ketika hasil clustering telah stabil dan proses iterasi dapat dihentikan.

Pada penelitian ini, proses iterasi dilakukan secara otomatis oleh sistem berbasis web sehingga perhitungan dilakukan dengan cepat dan efisien.

3.8.7 Hasil Implementasi K-Means

Hasil implementasi algoritma K-Means berupa sejumlah cluster yang berisi kabupaten/kota di Provinsi Sumatera Utara yang memiliki karakteristik IPM yang serupa. Setiap cluster menggambarkan kelompok wilayah dengan tingkat pembangunan manusia yang relatif sama berdasarkan indikator kesehatan, pendidikan, dan standar hidup.

Hasil clustering menggunakan algoritma K-Means dibandingkan dengan hasil yang diperoleh dari algoritma Hierarchical Clustering. Perbandingan tersebut dilakukan menggunakan metrik evaluasi, yaitu Silhouette Score dan Davies-Bouldin Index, sehingga dapat diketahui algoritma yang memberikan kualitas pengelompokan terbaik.

3.8.8 Alur Implementasi Algoritma K-Means

Tahapan implementasi algoritma K-Means pada penelitian ini dapat dirangkum sebagai berikut.

1. Mengunggah dataset IPM ke dalam sistem.
2. Melakukan normalisasi data hasil preprocessing.
3. Menentukan jumlah cluster menggunakan Metode Elbow.
4. Menginisialisasi centroid awal.
5. Menghitung jarak setiap data terhadap seluruh centroid menggunakan Euclidean Distance.
6. Menempatkan setiap data pada cluster dengan jarak terdekat.
7. Menghitung centroid baru berdasarkan rata-rata anggota setiap cluster.
8. Mengulangi proses perhitungan jarak dan pembaruan centroid hingga konvergen.

9. Menyimpan hasil clustering untuk proses evaluasi dan perbandingan dengan algoritma Hierarchical Clustering.

3.9 Implementasi Algoritma Hierarchical Clustering

Algoritma Hierarchical Clustering merupakan salah satu metode clustering yang mengelompokkan data berdasarkan tingkat kemiripan antarobjek secara bertahap sehingga membentuk struktur hierarki. Berbeda dengan algoritma K-Means yang mengharuskan jumlah cluster ditentukan sejak awal, Hierarchical Clustering membentuk hubungan antarobjek dalam bentuk pohon hierarki (*hierarchical tree*) yang disebut dendrogram. Melalui dendrogram tersebut, jumlah cluster dapat ditentukan sesuai dengan kebutuhan analisis.

Pada penelitian ini digunakan metode Agglomerative Hierarchical Clustering, metode *bottom-up* yang dimulai dengan menganggap setiap data sebagai satu cluster tersendiri. Selanjutnya, dua cluster yang memiliki tingkat kemiripan paling tinggi akan digabungkan secara bertahap hingga seluruh data berada dalam satu hierarki. Implementasi algoritma dilakukan menggunakan sistem berbasis web yang dibangun dengan bahasa pemrograman JavaScript.

3.9.1 Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering merupakan metode clustering yang bekerja dengan menggabungkan cluster secara bertahap berdasarkan tingkat kedekatan antarcluster. Pada tahap awal, setiap kabupaten/kota dianggap sebagai satu cluster yang berdiri sendiri. Selanjutnya dihitung jarak antarcluster, kemudian dua cluster dengan jarak paling dekat digabung menjadi satu cluster baru.

Proses penggabungan dilakukan secara berulang hingga seluruh data tergabung dalam satu struktur hierarki. Selama proses tersebut, hubungan

antarcluster direpresentasikan dalam bentuk dendrogram sehingga memudahkan proses analisis dan penentuan jumlah cluster.

Tahapan utama pada metode Agglomerative Hierarchical Clustering adalah sebagai berikut.

1. Menganggap setiap data sebagai satu cluster.
2. Menghitung jarak antarcluster.
3. Menggabungkan dua cluster yang memiliki jarak paling dekat.
4. Menghitung kembali jarak antarcluster setelah proses penggabungan.
5. Mengulangi proses hingga seluruh data membentuk satu struktur hierarki.

Metode ini tidak bergantung pada pemilihan centroid sehingga hasil clustering lebih stabil terhadap posisi awal data dibandingkan algoritma K-Means.

3.9.2 Metode Linkage

Setelah jarak antarobjek diperoleh, langkah berikutnya adalah menentukan metode linkage yang digunakan untuk menghitung jarak antarcluster. Linkage merupakan metode yang digunakan untuk menentukan pasangan cluster yang akan digabungkan pada setiap iterasi.

Pada penelitian ini digunakan Ward Linkage. Metode Ward dipilih karena mampu menghasilkan cluster yang relatif homogen dengan meminimalkan variasi di dalam setiap cluster (*within-cluster variance*). Selain itu, metode Ward sangat sesuai diterapkan pada data numerik yang telah dinormalisasi.

Dalam setiap proses penggabungan, Ward Linkage memilih pasangan cluster yang menghasilkan kenaikan variasi terkecil sehingga cluster yang terbentuk memiliki tingkat kesamaan yang tinggi.

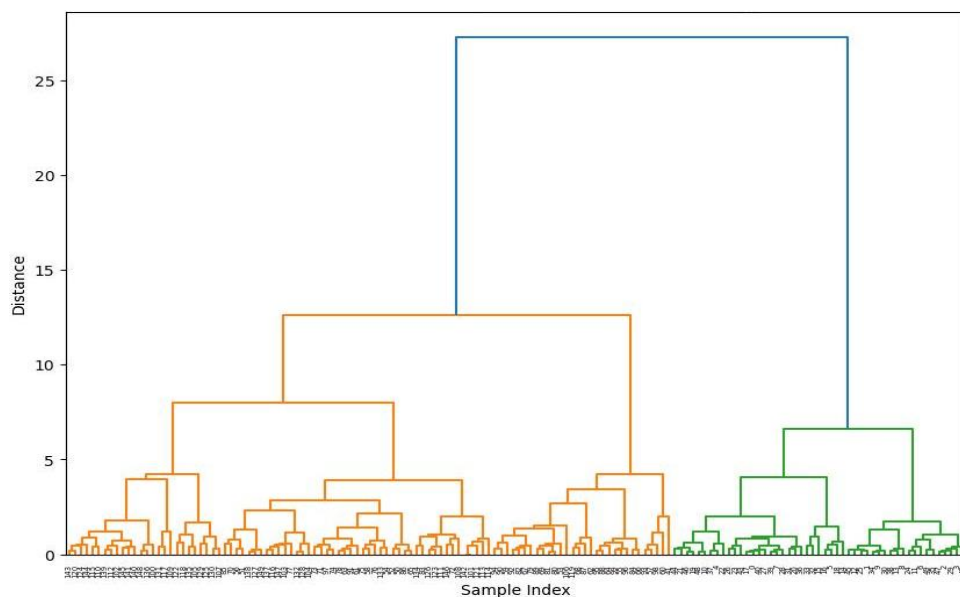
3.9.3 Pembentukan Dendrogram

Hasil dari proses Agglomerative Hierarchical Clustering divisualisasikan dalam bentuk dendrogram. Dendrogram merupakan diagram berbentuk pohon yang menunjukkan proses penggabungan cluster dari awal hingga seluruh data menjadi satu cluster.

Pada bagian bawah dendrogram terdapat seluruh objek penelitian yang mewakili kabupaten/kota di Provinsi Sumatera Utara. Sementara itu, garis-garis vertikal menunjukkan proses penggabungan antarcluster berdasarkan tingkat kemiripan.

Semakin rendah posisi penggabungan pada dendrogram, semakin tinggi tingkat kemiripan antarobjek. Sebaliknya, apabila penggabungan terjadi pada posisi yang lebih tinggi, maka kedua cluster memiliki tingkat perbedaan yang lebih besar.

Visualisasi dendrogram membantu peneliti memahami hubungan antarcluster sekaligus mempermudah interpretasi hasil clustering.



Gambar 3.2 Dendrogram Hierarchical Clustering.

3.9.4 Penentuan Jumlah Cluster

Jumlah cluster pada Hierarchical Clustering ditentukan dengan memotong dendrogram pada ketinggian tertentu (*cut-off point*). Pemotongan dilakukan pada bagian yang menunjukkan jarak penggabungan yang cukup besar sehingga cluster yang terbentuk memiliki tingkat pemisahan yang baik.

Penentuan jumlah cluster juga mempertimbangkan hasil evaluasi menggunakan metrik Silhouette Score dan Davies-Bouldin Index. Dengan demikian, jumlah cluster yang dipilih tidak hanya berdasarkan visualisasi dendrogram, tetapi juga didukung oleh hasil evaluasi kuantitatif.

3.9.5 Hasil Implementasi Hierarchical Clustering

Hasil implementasi algoritma Hierarchical Clustering berupa kelompok kabupaten/kota di Provinsi Sumatera Utara yang memiliki karakteristik Indeks Pembangunan Manusia yang relatif serupa. Setiap cluster menunjukkan kelompok wilayah dengan kemiripan pola pertumbuhan nilai IPM selama periode 2020–2025.

Hasil clustering yang diperoleh selanjutnya dibandingkan dengan hasil algoritma K-Means menggunakan metrik evaluasi yang sama. Perbandingan tersebut bertujuan untuk mengetahui algoritma yang menghasilkan kualitas cluster terbaik berdasarkan tingkat kohesi dalam cluster dan tingkat pemisahan antarcluster.

3.9.6 Alur Implementasi Hierarchical Clustering

Tahapan implementasi algoritma Hierarchical Clustering pada penelitian ini adalah sebagai berikut.

1. Mengunggah dataset IPM hasil preprocessing ke dalam sistem.
2. Menggunakan data yang telah dinormalisasi sebagai masukan algoritma.

3. Menghitung jarak antarobjek menggunakan Euclidean Distance.
4. Menganggap setiap data sebagai satu cluster.
5. Menggabungkan dua cluster yang memiliki jarak paling dekat menggunakan metode Ward Linkage.
6. Mengulangi proses penggabungan hingga terbentuk struktur hierarki.
7. Menampilkan hasil penggabungan dalam bentuk dendrogram.
8. Menentukan jumlah cluster berdasarkan hasil analisis dendrogram.
9. Menyimpan hasil clustering untuk proses evaluasi dan perbandingan dengan algoritma K-Means.

3.10 Evaluasi Performa Algoritma

Tahap evaluasi merupakan salah satu tahapan penting dalam penelitian ini karena bertujuan untuk menilai kualitas hasil pengelompokan yang dihasilkan oleh algoritma K-Means dan Hierarchical Clustering. Evaluasi dilakukan untuk mengetahui algoritma yang mampu menghasilkan cluster dengan tingkat kemiripan yang tinggi di dalam kelompok (*intra-cluster similarity*) dan tingkat perbedaan yang tinggi antar kelompok (*inter-cluster dissimilarity*).

Pada penelitian ini digunakan tiga metrik evaluasi, yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Ketiga metrik tersebut dipilih karena merupakan metode evaluasi internal yang umum digunakan dalam penelitian clustering. Evaluasi dilakukan terhadap hasil pengelompokan dari kedua algoritma menggunakan dataset yang sama sehingga hasil yang diperoleh dapat dibandingkan secara objektif.

3.10.1 Silhouette Score

Silhouette Score merupakan metode evaluasi internal yang digunakan untuk mengukur seberapa baik suatu data berada pada cluster tempat data tersebut dikelompokkan dibandingkan dengan cluster lainnya. Nilai Silhouette Score diperoleh dari perbandingan antara rata-rata jarak suatu data terhadap anggota dalam cluster yang sama dengan rata-rata jarak terhadap cluster terdekat lainnya.

Rumus Silhouette Score ditunjukkan sebagai berikut.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Keterangan:

- $s(i)$ = nilai Silhouette Score untuk data ke-(i).
- $a(i)$ = rata-rata jarak data ke-(i) dengan seluruh anggota pada cluster yang sama.
- $b(i)$ = rata-rata jarak data ke-(i) terhadap cluster terdekat.

Nilai Silhouette Score berada pada rentang -1 sampai 1. Interpretasi nilai tersebut adalah sebagai berikut.

Tabel 3.6 Interpretasi Nilai Silhouette Score

Nilai Silhouette Score	Interpretasi
0,71 – 1,00	Struktur cluster sangat baik
0,51 – 0,70	Struktur cluster baik
0,26 – 0,50	Struktur cluster cukup baik
$\leq 0,25$	Struktur cluster kurang baik

Semakin mendekati nilai 1, semakin baik kualitas cluster yang terbentuk karena menunjukkan bahwa setiap objek memiliki tingkat kemiripan yang

tinggi dengan anggota dalam cluster yang sama dan memiliki perbedaan yang jelas dengan cluster lainnya.

3.10.2 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) merupakan metode evaluasi internal yang digunakan untuk mengukur tingkat kemiripan antarcluster berdasarkan penyebaran data di dalam cluster dan jarak antarcluster. Nilai DBI dihitung menggunakan rasio antara jarak penyebaran data dalam cluster dengan jarak antar centroid dari dua cluster yang dibandingkan.

Secara umum, nilai DBI dihitung menggunakan persamaan berikut.

Rumus Davies-Bouldin Index (DBI)

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right)$$

Keterangan:

- DBI = Davies-Bouldin Index.
- k = Jumlah cluster.
- S_i = Rata-rata jarak seluruh data pada cluster ke- i terhadap centroid cluster ke- i (ukuran penyebaran cluster).
- S_j = Rata-rata jarak seluruh data pada cluster ke- j terhadap centroid cluster ke- j .
- M_{ij} = Jarak antara centroid cluster ke- i dan centroid cluster ke- j .
- $\max$ = Nilai rasio terbesar antara cluster ke- i dengan cluster lainnya.

Interpretasi nilai Davies-Bouldin Index adalah sebagai berikut.

Tabel 3.7 Interpretasi Nilai Davies-Bouldin Index

Nilai DBI	Interpretasi
Mendekati 0	Kualitas cluster sangat baik
Nilai kecil	Cluster semakin optimal
Nilai besar	Cluster kurang optimal

Semakin kecil nilai DBI yang diperoleh, semakin baik kualitas hasil clustering karena menunjukkan bahwa setiap cluster memiliki tingkat penyebaran data yang kecil dan memiliki jarak yang cukup jauh terhadap cluster lainnya.

3.10.4 Perbandingan Hasil Evaluasi

Setelah seluruh proses clustering selesai dilakukan, hasil pengelompokan dari algoritma K-Means dan Hierarchical Clustering dibandingkan menggunakan kedua metrik evaluasi tersebut. Perbandingan dilakukan untuk mengetahui algoritma yang memiliki performa terbaik dalam mengelompokkan data Indeks Pembangunan Manusia (IPM).

Contoh format hasil evaluasi ditunjukkan pada Tabel 3.4.

Tabel 3.8 Contoh Perbandingan Hasil Evaluasi Algoritma

Algoritma	Silhouette Score	Davies-Bouldin Index	SSE / Inersia	Waktu Proses
K-Means	0,6705	0,3942	1,1492	Cepat

Hierarchical Clustering	0,6705	0,3942	—	Lambat
-------------------------	--------	--------	---	--------

Berdasarkan hasil evaluasi tersebut, algoritma yang memiliki nilai Silhouette Score lebih tinggi serta nilai Davies-Bouldin Index lebih rendah dianggap memiliki performa clustering yang lebih baik. Apabila kedua algoritma menghasilkan nilai yang setara, waktu komputasi menjadi pertimbangan tambahan untuk menentukan algoritma yang lebih efisien.

3.10.5 Interpretasi Hasil Evaluasi

Tahap akhir evaluasi dilakukan dengan menginterpretasikan hasil yang diperoleh dari kedua metrik evaluasi. Interpretasi bertujuan untuk menentukan algoritma yang paling sesuai dalam mengelompokkan data Indeks Pembangunan Manusia (IPM) berdasarkan kualitas cluster yang dihasilkan.

Algoritma dengan nilai evaluasi terbaik selanjutnya digunakan sebagai dasar dalam menganalisis karakteristik setiap cluster. Setiap cluster akan diidentifikasi berdasarkan nilai rata-rata IPM sehingga dapat diketahui kelompok kabupaten/kota yang memiliki tingkat pembangunan manusia tinggi, sedang, maupun rendah.

Hasil interpretasi tersebut diharapkan dapat memberikan informasi yang berguna bagi pemerintah daerah, akademisi, maupun peneliti sebagai bahan pertimbangan dalam mengevaluasi kondisi pembangunan manusia di Provinsi Sumatera Utara serta menjadi referensi dalam penelitian selanjutnya yang menggunakan metode clustering.

3.11 Perancangan Sistem

Perancangan sistem merupakan tahapan yang dilakukan untuk memberikan gambaran mengenai alur kerja sistem yang digunakan dalam penelitian. Sistem yang dirancang bertujuan untuk mempermudah proses pengolahan data Indeks Pembangunan Manusia (IPM), mulai dari proses memasukkan data, melakukan preprocessing, menjalankan algoritma clustering, hingga menampilkan hasil evaluasi performa algoritma K-Means dan Hierarchical Clustering.

Pada penelitian ini, sistem dibangun sebagai aplikasi berbasis web menggunakan bahasa pemrograman JavaScript dengan memanfaatkan pustaka Chart.js untuk visualisasi dan PapaParse untuk pembacaan berkas CSV. Sistem menerima masukan berupa dataset IPM kabupaten/kota di Provinsi Sumatera Utara, kemudian melakukan pengolahan data sesuai tahapan Knowledge Discovery in Databases (KDD). Hasil akhir sistem berupa kelompok data (cluster), nilai evaluasi performa algoritma, serta visualisasi hasil clustering.

Perancangan sistem bertujuan agar proses penelitian dapat dilakukan secara sistematis, mudah dipahami, dan mudah direplikasi oleh peneliti lain.

3.11.1 Flowchart Penelitian

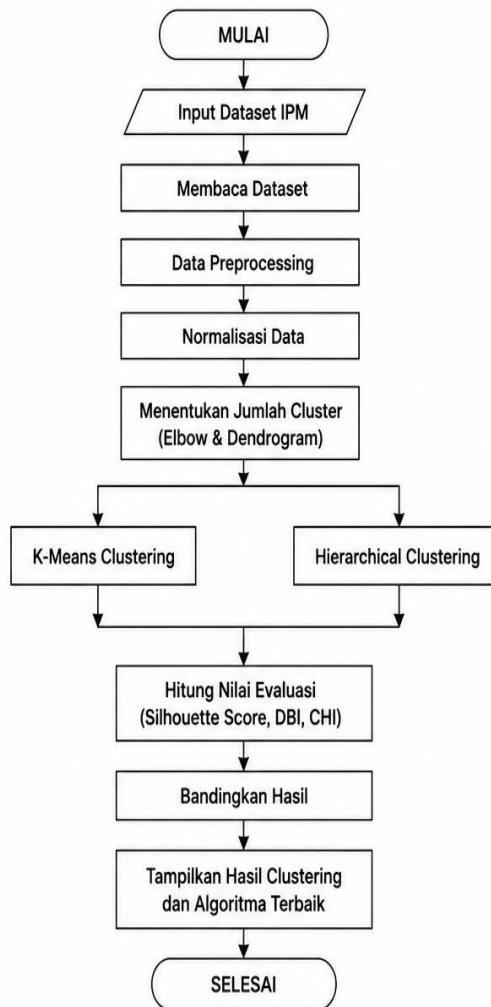
Flowchart penelitian digunakan untuk menggambarkan urutan proses yang dilakukan selama penelitian, mulai dari pengumpulan data hingga diperoleh hasil evaluasi algoritma.

Adapun alur penelitian adalah sebagai berikut.

1. Memulai proses penelitian.
2. Melakukan studi literatur mengenai Data Mining, K-Means, Hierarchical Clustering, dan evaluasi clustering.

3. Mengumpulkan data IPM dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara.
4. Melakukan preprocessing data, meliputi pemeriksaan missing value, data duplikat, dan normalisasi data.
5. Menentukan jumlah cluster menggunakan Metode Elbow.
6. Melakukan proses clustering menggunakan algoritma K-Means.
7. Melakukan proses clustering menggunakan algoritma Hierarchical Clustering.
8. Menghitung nilai evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index.
9. Membandingkan performa kedua algoritma.
10. Menarik kesimpulan penelitian.
11. Selesai.

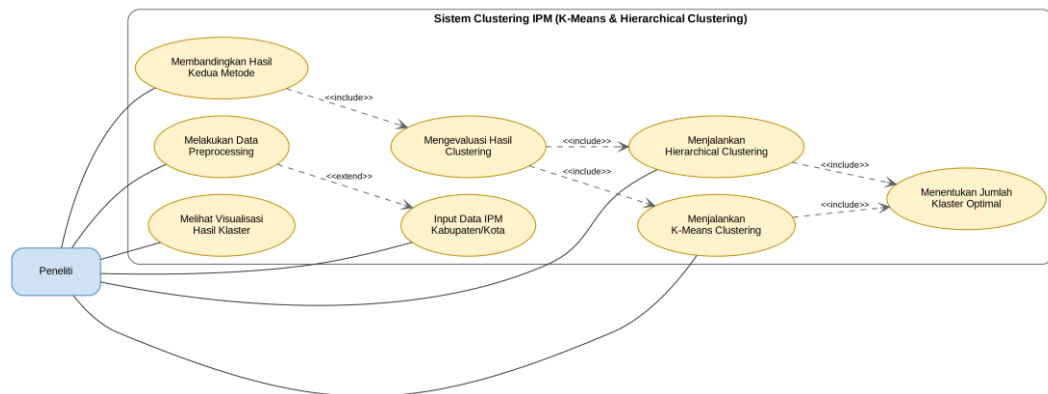
Flowchart penelitian dapat digambarkan menggunakan simbol flowchart standar pada aplikasi Draw.io atau Microsoft Visio.



Gambar 3.3 Flowchart Penelitian

3.11.2 Use Case Diagram

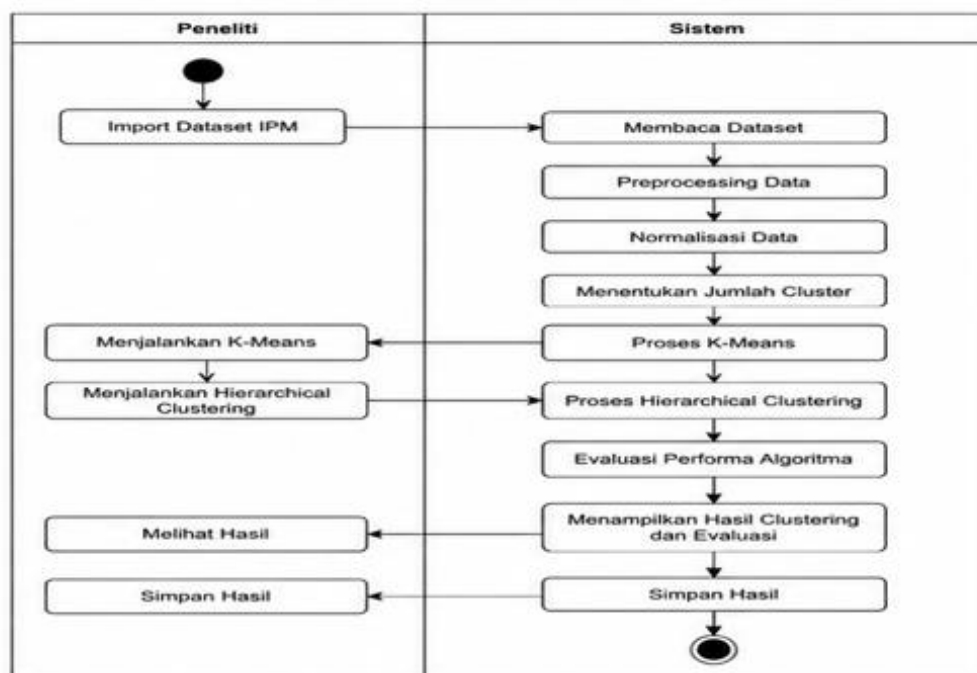
Use Case diagram menggambarkan interaksi antara peneliti sebagai pengguna (actor) dengan fungsi-fungsi utama yang disediakan oleh sistem clustering IPM, mulai dari input data, preprocessing, penentuan jumlah klaster, penerapan algoritma K-Means dan Hierarchical Clustering, evaluasi, hingga perbandingan dan visualisasi hasil.



Gambar 3.4 Use case penelitian.

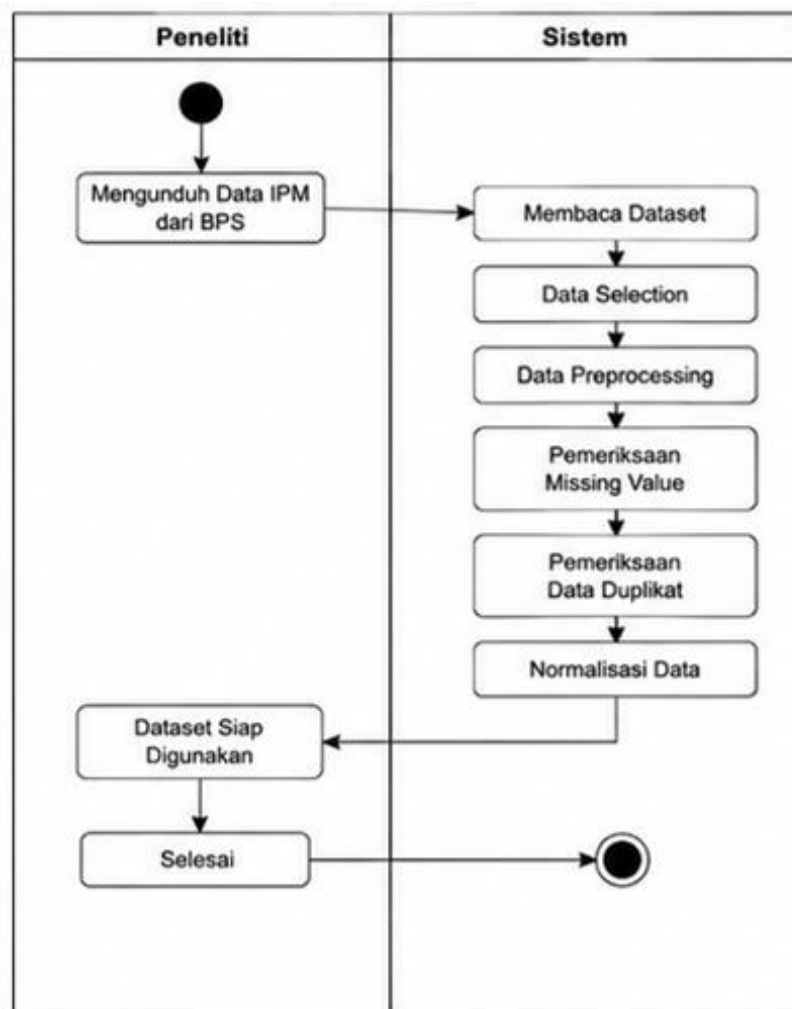
3.11.3. Activity diagram

menggambarkan alur kerja utama sistem dalam menerapkan algoritma K-Means, dimulai dari input data, tahapan preprocessing, penentuan jumlah kluster, inisialisasi centroid, perhitungan jarak, pengelompokan data, hingga proses iterasi berulang sampai kondisi konvergen tercapai.



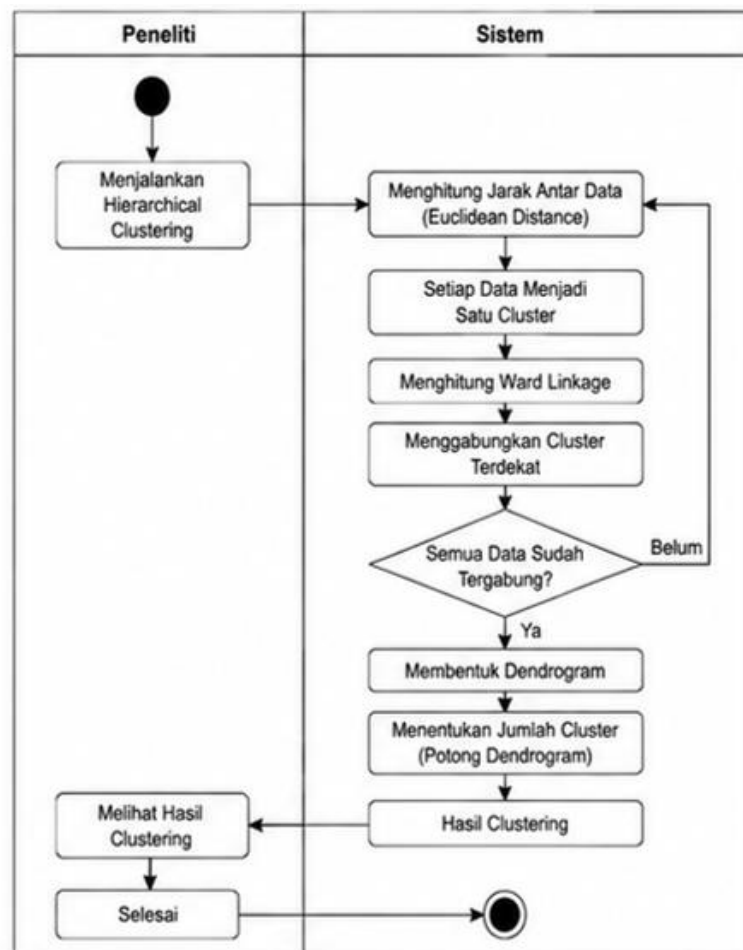
Gambar 3.5 Activity Diagram Proses K-Means Clustering

Gambar 3.5 menunjukkan alur proses pengelompokan data menggunakan algoritma K-Means. Proses dimulai dari peneliti mengimpor dataset IPM, kemudian sistem membaca data, melakukan preprocessing dan normalisasi data. Selanjutnya sistem menentukan jumlah cluster menggunakan Metode Elbow, kemudian menjalankan proses K-Means Clustering. Setelah proses selesai, sistem menampilkan hasil clustering dan peneliti dapat melihat serta menyimpan hasil analisis. Diagram ini menggambarkan alur kerja algoritma K-Means secara sistematis dari input data hingga menghasilkan kelompok data.



Gambar 3.6 Activity Data

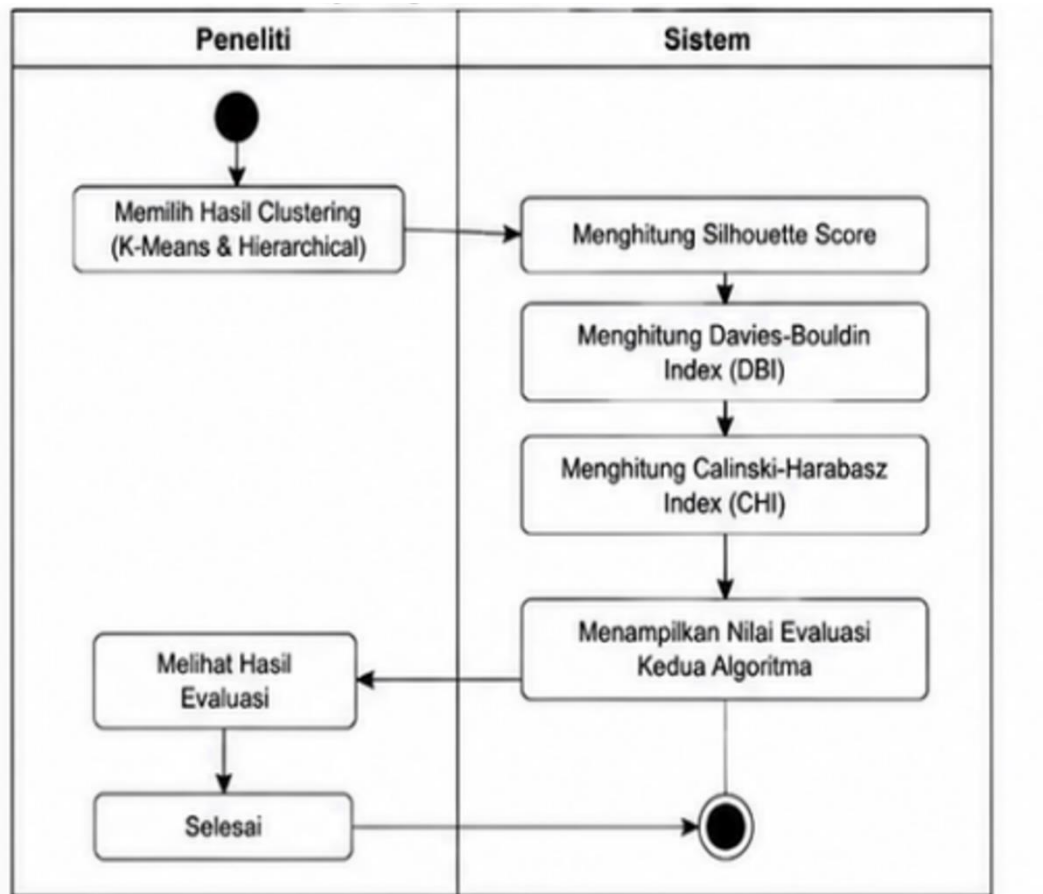
Gambar 3.6 menunjukkan alur proses pengumpulan dan persiapan data penelitian. Proses dimulai ketika peneliti mengunduh data IPM dari Badan Pusat Statistik (BPS). Selanjutnya sistem membaca dataset, melakukan data selection, preprocessing, pemeriksaan missing value, pemeriksaan data duplikat, serta normalisasi data. Setelah seluruh tahapan selesai, dataset dinyatakan siap digunakan pada proses clustering.



Gambar. 3.7 Activity hierarchical clistering

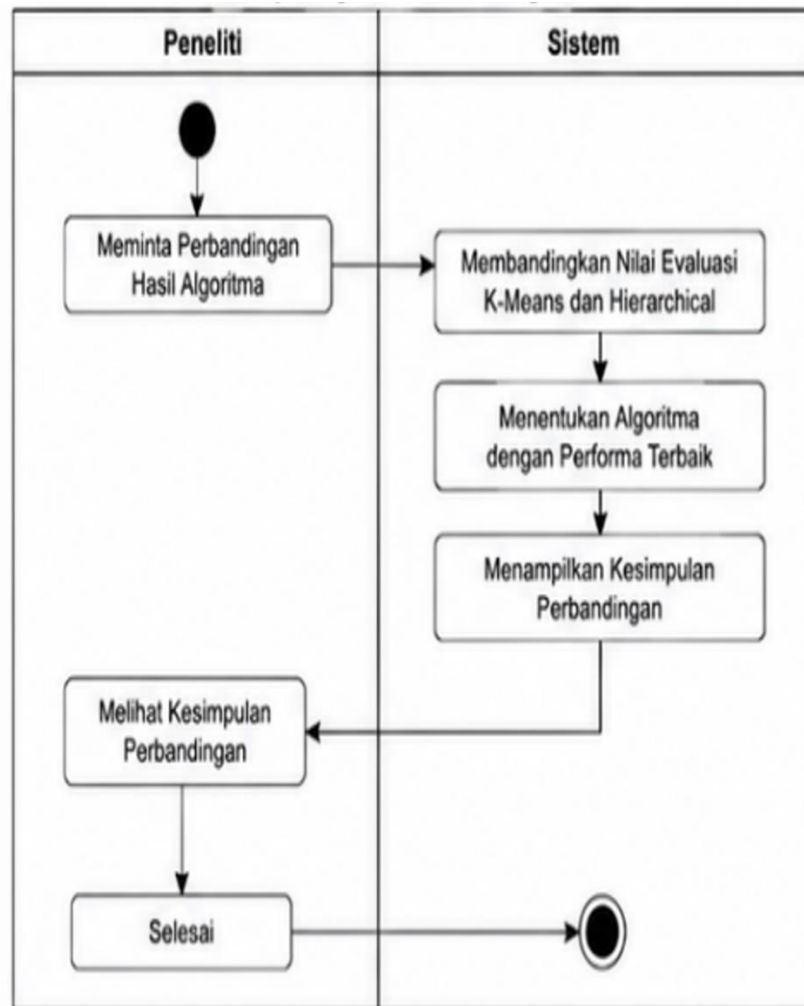
Gambar 3.7 menunjukkan alur proses pengelompokan data menggunakan algoritma Hierarchical Clustering. Proses dimulai ketika peneliti menjalankan algoritma, kemudian sistem menghitung jarak antar data, membentuk setiap data

sebagai satu cluster awal, menghitung Ward Linkage, dan menggabungkan cluster yang memiliki jarak paling dekat. Proses ini dilakukan secara berulang hingga seluruh data tergabung, kemudian sistem membentuk dendrogram, menentukan jumlah cluster, dan menampilkan hasil clustering kepada peneliti.



Gambar 3.8 Activity Diagram Evaluasi Performa

Gambar 3.8 menggambarkan proses evaluasi performa hasil clustering. Peneliti memilih hasil clustering dari algoritma K-Means dan Hierarchical Clustering, kemudian sistem menghitung nilai Silhouette Score dan Davies-Bouldin Index (DBI). Setelah seluruh nilai evaluasi diperoleh, sistem menampilkan hasil evaluasi sehingga peneliti dapat mengetahui kualitas pengelompokan dari masing-masing algoritma.



Gambar 3.9 Activity Activity Diagram Perbandingan Hasil

Gambar 3.9 menunjukkan proses perbandingan hasil evaluasi antara algoritma K-Means dan Hierarchical Clustering. Peneliti meminta sistem untuk melakukan perbandingan, kemudian sistem membandingkan nilai evaluasi dari kedua algoritma, menentukan algoritma dengan performa terbaik berdasarkan metrik evaluasi, dan menampilkan kesimpulan hasil perbandingan. Selanjutnya peneliti dapat melihat hasil perbandingan sebagai dasar dalam penarikan kesimpulan penelitian.

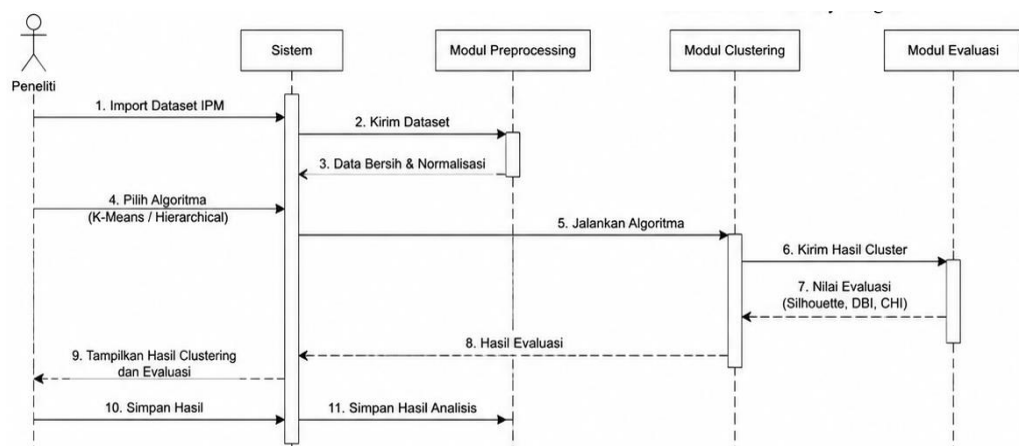
3.11.4 Sequence Diagram

Sequence Diagram menggambarkan interaksi antara peneliti dengan sistem selama proses pengolahan data.

Urutan interaksi adalah sebagai berikut.

1. Peneliti mengunggah dataset IPM.
2. Sistem menerima dataset.
3. Sistem melakukan preprocessing.
4. Sistem melakukan normalisasi data.
5. Peneliti memilih algoritma K-Means atau Hierarchical Clustering.
6. Sistem menjalankan algoritma yang dipilih.
7. Sistem menghasilkan cluster.
8. Sistem menghitung nilai evaluasi performa.
9. Sistem menampilkan hasil clustering dan evaluasi.
10. Peneliti menyimpan hasil analisis.
11. Proses selesai.

Sequence Diagram memperlihatkan urutan komunikasi antara pengguna dan sistem sehingga memudahkan pemahaman terhadap alur kerja aplikasi.



Gambar 3.10 5Sequence Diagram

3.11.5 Deskripsi Proses Sistem

Sistem yang dirancang pada penelitian ini terdiri atas empat proses utama, yaitu proses input, proses pengolahan data, proses clustering, dan proses keluaran.

1. Input Data

Sistem menerima dataset IPM kabupaten/kota di Provinsi Sumatera Utara dalam format Microsoft Excel atau CSV.

2. Pengolahan Data

Sistem melakukan preprocessing, meliputi pemeriksaan missing value, penghapusan data duplikat, serta normalisasi data agar seluruh atribut memiliki skala yang seragam.

3. Proses Clustering

Dataset yang telah diproses digunakan sebagai masukan bagi algoritma K-Means dan Hierarchical Clustering. Kedua algoritma dijalankan secara terpisah menggunakan data yang sama sehingga hasilnya dapat dibandingkan secara objektif.

4. Output Sistem

Sistem menghasilkan informasi berupa hasil pengelompokan kabupaten/kota, visualisasi cluster, nilai Silhouette Score dan Davies-Bouldin Index, serta perbandingan performa kedua algoritma.

Perancangan sistem ini menjadi pedoman dalam proses implementasi dan pengujian sehingga seluruh tahapan penelitian dapat dilakukan secara terstruktur serta menghasilkan informasi yang sesuai dengan tujuan penelitian.

3.12 Perangkat Penelitian

Perangkat penelitian merupakan sarana yang digunakan untuk mendukung proses pengolahan data, implementasi algoritma, serta evaluasi hasil penelitian. Perangkat yang digunakan terdiri atas perangkat keras (*hardware*) dan perangkat lunak (*software*). Penggunaan perangkat yang sesuai bertujuan agar proses pengolahan data dapat berjalan dengan baik, cepat, dan menghasilkan keluaran yang akurat.

3.12.1 Perangkat Keras (Hardware)

Perangkat keras merupakan komponen utama yang digunakan untuk menjalankan seluruh proses penelitian, mulai dari pengolahan data, implementasi algoritma clustering, hingga visualisasi hasil. Adapun spesifikasi perangkat keras yang digunakan dalam penelitian ini disajikan pada Tabel 3.6.

Tabel 3.9 Spesifikasi Perangkat Keras

No	Perangkat	Spesifikasi
1	Laptop/Komputer	Digunakan sebagai media utama penelitian
2	Processor	Intel Core i5 Generasi ke-10 atau setara (AMD Ryzen 5)
3	Memori (RAM)	Minimal 8 GB
4	Media Penyimpanan	SSD minimal 256 GB
5	Monitor	Resolusi minimal 1366 × 768 piksel
6	Keyboard dan Mouse	Perangkat input untuk menjalankan aplikasi

Spesifikasi tersebut dinilai telah memadai untuk menjalankan sistem berbasis web yang digunakan, termasuk proses normalisasi data, implementasi algoritma K-Means dan Hierarchical Clustering, serta evaluasi performa algoritma secara lancar melalui peramban (browser).

3.12.2 Perangkat Lunak (Software)

Perangkat lunak digunakan untuk membantu proses pengolahan data, implementasi algoritma clustering, visualisasi hasil, serta penyusunan laporan penelitian. Adapun perangkat lunak yang digunakan disajikan pada Tabel 3.7.

Tabel 3.10 Perangkat Lunak yang Digunakan

No	Perangkat Lunak	Fungsi
1	Windows 11 (atau sistem operasi yang digunakan)	Sistem operasi untuk menjalankan seluruh aplikasi penelitian
2	Google Chrome (atau peramban modern lain)	Menjalankan sistem clustering berbasis web
3	Visual Studio Code	Editor kode untuk menulis dan mengembangkan sistem (HTML, CSS, JavaScript)
4	Microsoft Excel	Melihat, mengedit, dan menyimpan dataset
5	JavaScript (ES6)	Bahasa pemrograman utama untuk seluruh logika sistem, termasuk implementasi algoritma clustering
6	PapaParse	Membaca dan mengurai (parsing) berkas CSV yang diunggah pengguna

7	Chart.js	Menampilkan grafik Elbow, dendrogram, scatter plot, dan visualisasi hasil evaluasi
8	HTML5 & CSS3	Menyusun struktur dan tampilan antarmuka sistem
9	Modul JavaScript kustom (native)	Implementasi manual algoritma K-Means, Hierarchical Clustering (Ward Linkage), Silhouette Score, dan Davies-Bouldin Index
10	Google Chrome DevTools (Opsional)	Debugging dan pengujian sistem di sisi klien (client-side)

Setiap perangkat lunak memiliki fungsi yang saling mendukung dalam proses penelitian. JavaScript digunakan sebagai bahasa pemrograman utama karena dapat dijalankan langsung di peramban tanpa memerlukan instalasi tambahan. Pustaka PapaParse digunakan untuk membaca serta mengurai berkas CSV, sedangkan seluruh algoritma—normalisasi Min-Max Scaling, K-Means, Hierarchical Clustering, serta perhitungan metrik evaluasi seperti Silhouette Score dan Davies-Bouldin Index—diimplementasikan secara manual dalam kode JavaScript.

Sementara itu, pustaka Chart.js digunakan untuk menghasilkan seluruh visualisasi hasil analisis, termasuk grafik Elbow, dendrogram, dan scatter plot, sehingga memudahkan interpretasi data secara interaktif langsung di peramban.

3.12.3 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Dataset berisi data Indeks Pembangunan Manusia (IPM) kabupaten/kota beserta indikator penyusunnya.

Atribut yang digunakan dalam penelitian ini adalah nilai Indeks Pembangunan Manusia (IPM) tahunan pada periode 2020–2025 (enam atribut, satu untuk setiap tahun).

Sebelum digunakan dalam proses clustering, dataset terlebih dahulu melalui tahapan *data preprocessing* dan *data transformation* untuk memastikan kualitas data serta menyeragamkan skala antarvariabel.

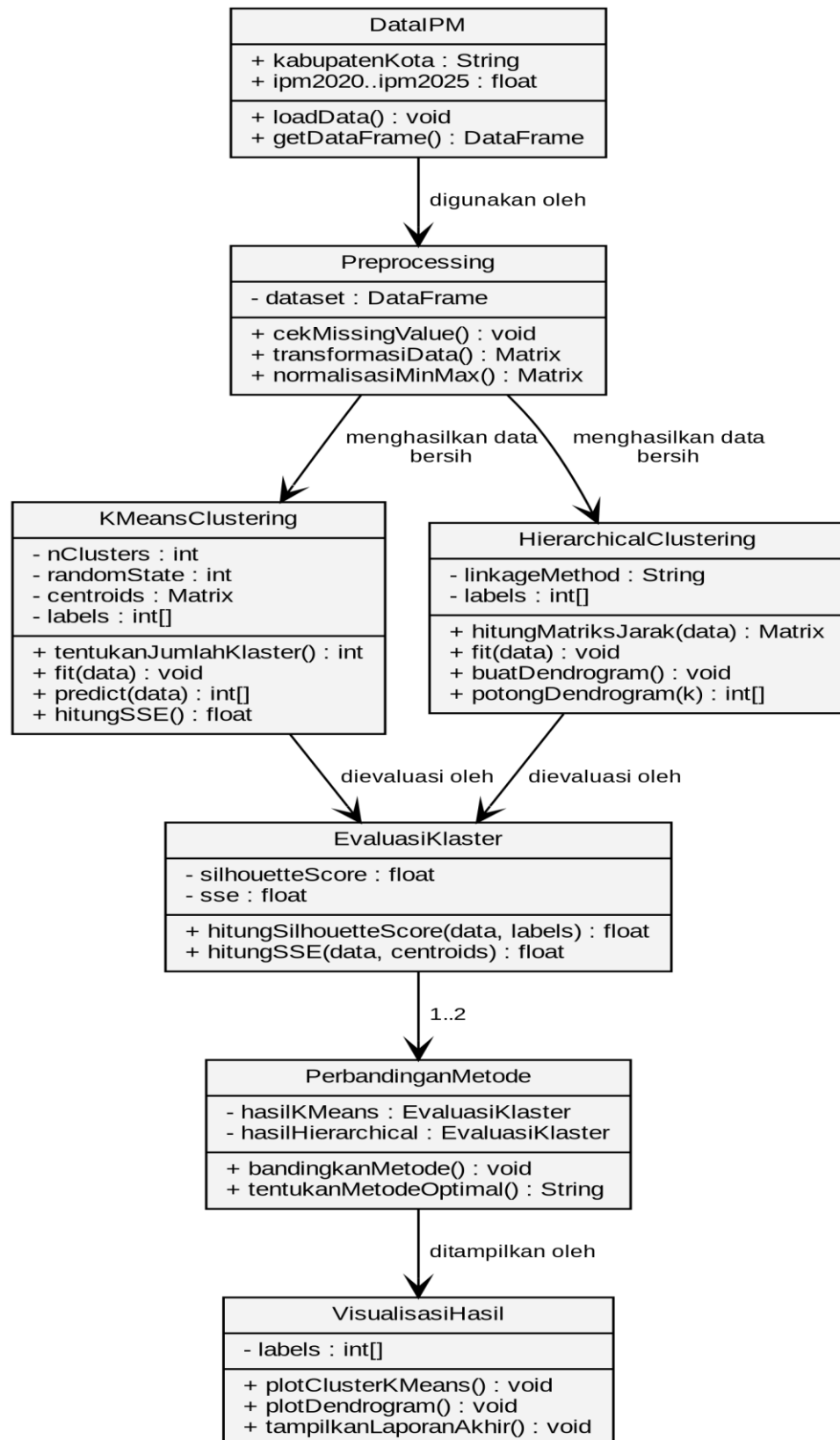
3.12.4 Lingkungan Implementasi

Implementasi algoritma dilakukan menggunakan bahasa pemrograman JavaScript dengan memanfaatkan pustaka yang telah dijelaskan sebelumnya. Seluruh proses, mulai dari membaca data, preprocessing, normalisasi, penerapan algoritma K-Means dan Hierarchical Clustering, hingga evaluasi hasil, dijalankan dalam satu sistem web terintegrasi sehingga memudahkan proses analisis dan dokumentasi penelitian.

Lingkungan implementasi yang terintegrasi juga memudahkan peneliti dalam melakukan pengujian, visualisasi hasil, serta membandingkan performa kedua algoritma secara konsisten menggunakan dataset yang sama.

3.13 Class Diagram

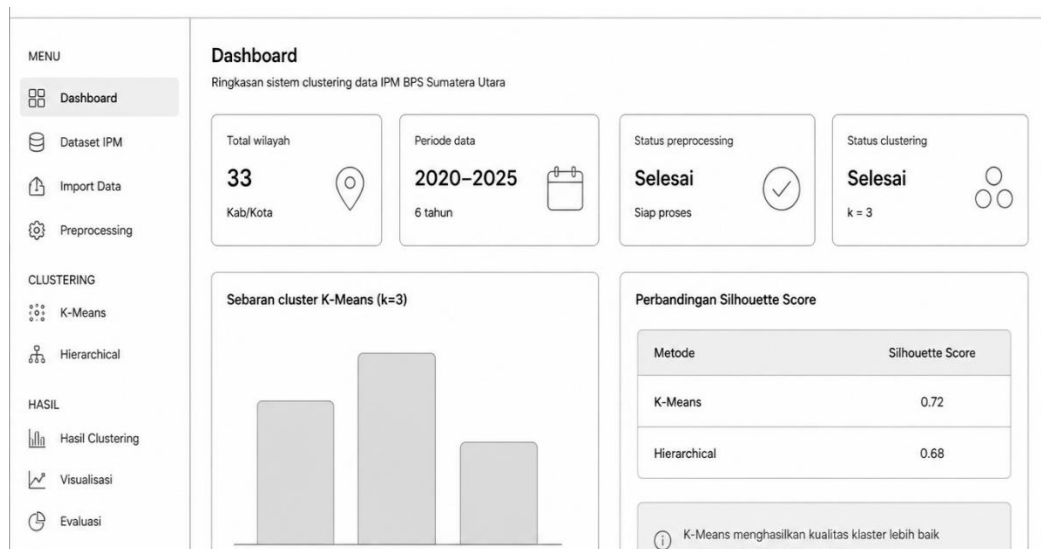
Class diagram menggambarkan struktur objek yang membangun sistem clustering IPM beserta atribut, metode, dan relasi antar kelas, meliputi kelas DataIPM, Preprocessing, KMeansClustering, HierarchicalClustering, EvaluasiKlaster, PerbandinganMetode, dan VisualisasiHasil.



Gambar 3.11 Class Diagram

3.14 Rancangan Antarmuka

1. Halaman Dashboard



Gambar 3.12 Halaman Dashboard

Pada area utama dashboard ditampilkan ringkasan informasi sistem, meliputi jumlah wilayah yang dianalisis, periode data yang digunakan (2020–2025), status preprocessing, dan status proses clustering. Selain itu, dashboard menyajikan visualisasi sebaran hasil clustering K-Means dalam bentuk diagram batang untuk memberikan gambaran jumlah kabupaten/kota pada setiap cluster. Di sisi kanan terdapat tabel perbandingan nilai Silhouette Score antara metode K-Means dan Hierarchical Clustering sebagai indikator kualitas hasil pengelompokan, disertai kesimpulan metode yang memberikan performa terbaik. Wireframe ini dirancang untuk memberikan tampilan yang sederhana, terstruktur, dan memudahkan pengguna dalam memantau hasil proses clustering secara keseluruhan.

2. Halaman Dataset

MENU

Dashboard

Dataset IPM

Import Data

Preprocessing

CLUSTERING

K-Means

Hierarchical

HASIL

Hasil Clustering

Visualisasi

Evaluasi

Laporan

Dataset IPM

Data IPM Kab/Kota Sumatera Utara periode 2020–2025 (sumber: BPS)

Cari kabupaten/kota...

Ekspor CSV

No	Kabupaten/Kota	IPM 2020	IPM 2021	IPM 2022	IPM 2023	IPM 2024	IPM 2025
1	Kota Medan	80.97	80.65	81.41	82.18	82.72	83.10
2	Kota Binjai	74.07	74.15	75.12	75.88	76.20	76.74
3	Deli Serdang	71.77	71.94	72.83	73.45	74.02	74.61
4	Nias Selatan	57.11	57.44	58.22	58.89	59.31	59.77
5	Nias Barat	55.20	55.63	56.74	57.10	57.48	57.93
...	...	...	...	...	...	...	...

Gambar 3.13 Halaman Dataset

Pada area utama ditampilkan tabel dataset IPM yang memuat nomor, nama kabupaten/kota, serta nilai IPM untuk setiap tahun dari 2020 hingga 2025. Untuk memudahkan pengelolaan data, tersedia fitur pencarian berdasarkan nama kabupaten/kota dan tombol Ekspor CSV untuk mengunduh data dalam format CSV. Di bagian bawah tabel ditampilkan informasi jumlah data yang sedang ditampilkan dari keseluruhan data yang tersedia. Wireframe ini dirancang agar pengguna dapat melihat, mencari, dan mengekspor data IPM dengan mudah sebagai tahap awal sebelum dilakukan proses preprocessing dan clustering.

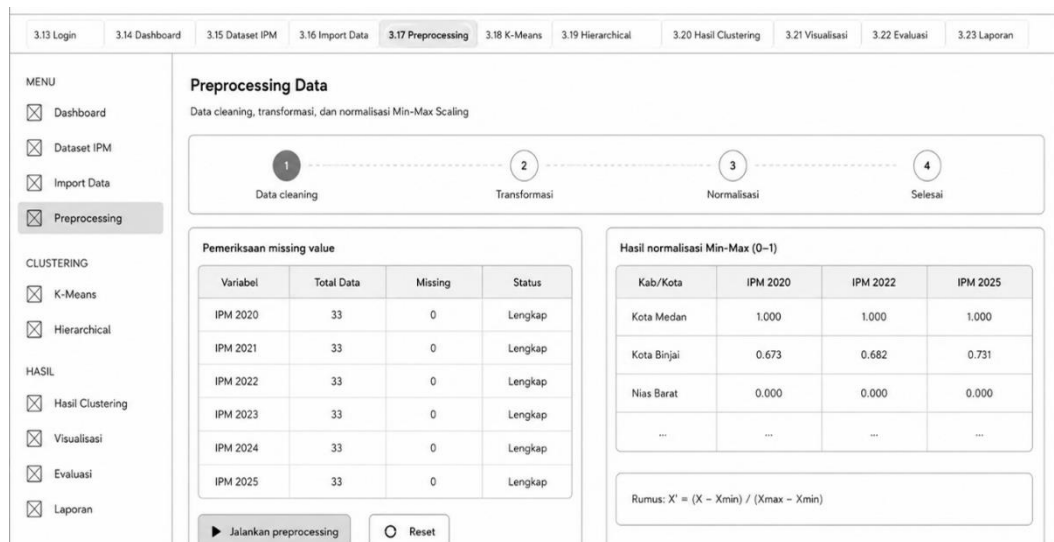
3. Halaman Import Data

No	Kabupaten/Kota	IPM 2020	IPM 2021	IPM 2022	IPM 2023	IPM 2024	IPM 2025
1	Kota Medan	80.97	80.65	81.41	82.18	82.72	83.10
2	Kota Binjai	74.07	74.15	75.12	75.88	76.20	76.74
3	Delit Serdang	71.77	71.94	72.83	73.45	74.02	74.61
4	Nias Selatan	57.11	57.44	58.22	58.89	59.31	59.77
5	Nias Barat	55.20	55.63	56.74	57.10	57.48	57.93
...	...	...	...	...	...	...	...

Gambar 3. 14 Halaman Import Data

Pada area utama disediakan form unggah berkas dengan metode drag-and-drop maupun tombol Pilih Berkas. Sistem juga menampilkan informasi mengenai format file yang didukung serta kolom yang wajib tersedia, yaitu Kabupaten/Kota dan data IPM tahun 2020–2025. Setelah file dipilih, sistem menampilkan ringkasan hasil pembacaan data, seperti jumlah baris, jumlah kolom, dan jumlah data yang hilang (*missing value*), serta pratinjau isi dataset dalam bentuk tabel. Di bagian bawah tersedia tombol Simpan ke Sistem untuk menyimpan data ke basis data dan tombol Batal untuk membatalkan proses impor. Wireframe ini dirancang agar proses pengunggahan dan validasi data menjadi lebih mudah, jelas, dan terstruktur sebelum memasuki tahap preprocessing.

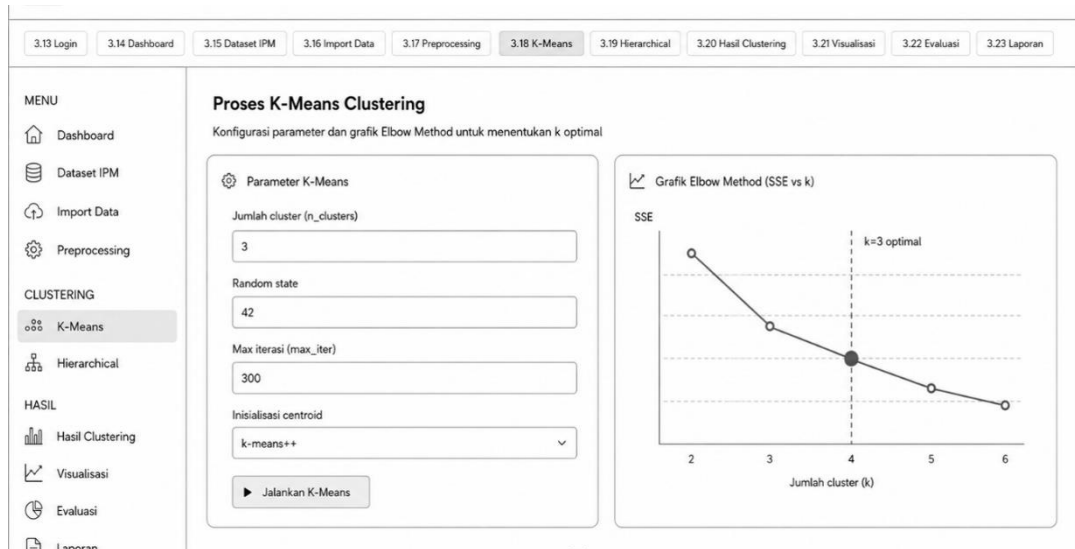
4. Halaman Preprocessing



Gambar 3.15 Halaman Preprocessing

Pada area utama terdapat indikator tahapan preprocessing yang menunjukkan urutan proses dari data cleaning hingga selesai. Selanjutnya, sistem menampilkan hasil pemeriksaan missing value dalam bentuk tabel yang berisi nama variabel, jumlah data, jumlah data yang hilang, dan status kelengkapan setiap atribut. Di sisi kanan ditampilkan hasil normalisasi Min-Max berupa nilai data yang telah berada pada rentang 0–1, disertai dengan rumus normalisasi yang digunakan, yaitu $X' = \frac{X - X_{min}}{X_{max} - X_{min}}$. Pada bagian bawah tersedia tombol Jalankan Preprocessing untuk memulai proses serta tombol Reset untuk mengembalikan kondisi awal. Wireframe ini dirancang agar pengguna dapat memverifikasi kualitas data dan memastikan seluruh data telah siap digunakan pada tahap clustering.

5. Halaman K-Means



Gambar 3.16 Halaman K-Means

Wireframe ini dirancang agar pengguna dapat mengatur parameter algoritma secara fleksibel sekaligus memperoleh rekomendasi jumlah cluster terbaik sebelum proses clustering dilakukan.

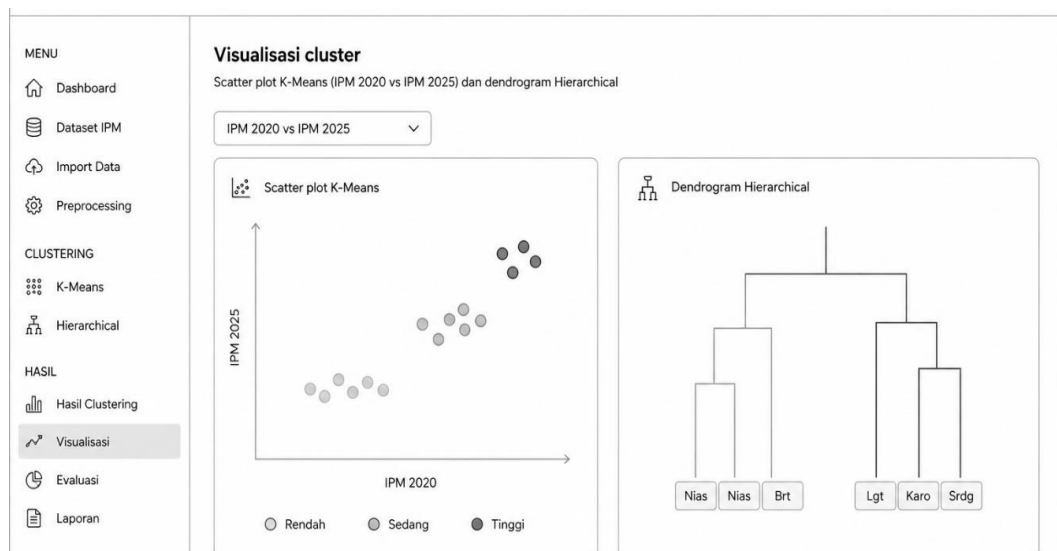
6. Halaman Hierarchical

Hasil clustering								
Label cluster per kab/kota berdasarkan nilai IPM 2020–2025								
K-Means (k=3)		Semua cluster		Ekspor				
No	Kabupaten/Kota	2020	2021	2022	2023	2024	Cluster	Kategori
1	Kota Medan	80.97	80.65	81.41	82.18	82.72	1	Tinggi
2	Kota Binjai	74.07	74.15	75.12	75.88	76.20	1	Tinggi
3	Deli Serdang	71.77	71.94	72.83	73.45	74.02	2	Sedang
4	Langkat	67.11	67.44	68.22	68.89	69.31	2	Sedang
5	Nias Selatan	57.11	57.44	58.22	58.89	59.31	3	Rendah
6	Nias Barat	55.20	55.63	56.74	57.10	57.48	3	Rendah
...	...	...	...	...	...	...	...	...

Gambar 3. 17 Halaman Hierarvhhical

Wireframe ini dirancang agar pengguna dapat mengatur parameter algoritma secara fleksibel sekaligus memperoleh rekomendasi jumlah cluster terbaik sebelum proses clustering dilakukan.

7. Halaman Visualisasi

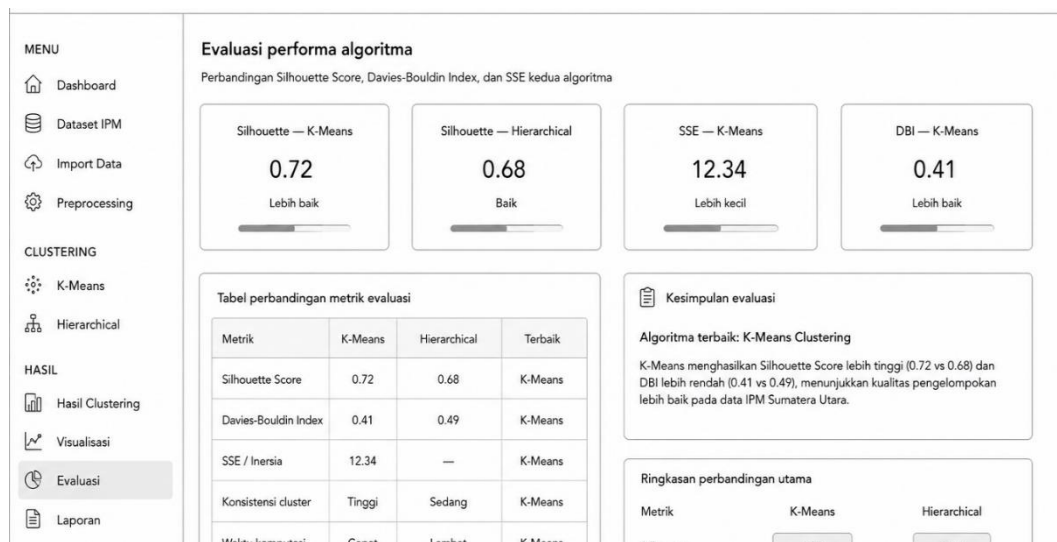


Gambar 3.18 Halaman Visualisasi

Pada area utama tersedia pilihan visualisasi yang memungkinkan pengguna menentukan atribut atau periode data yang akan ditampilkan, misalnya IPM 2020 vs IPM 2025. Selanjutnya, sistem menampilkan scatter plot hasil K-Means yang memperlihatkan persebaran data berdasarkan nilai IPM dengan warna atau simbol yang membedakan setiap cluster, yaitu Rendah, Sedang, dan Tinggi. Di sampingnya ditampilkan dendrogram Hierarchical Clustering yang menggambarkan hubungan hierarkis antar objek dan proses pembentukan cluster berdasarkan tingkat kemiripan data. Wireframe ini dirancang agar pengguna dapat membandingkan hasil visualisasi dari metode K-Means dan Hierarchical Clustering

secara mudah, sehingga interpretasi hasil pengelompokan menjadi lebih jelas dan informatif

8. Halaman Evaluasi



Gambar 3.19 Halaman Evaluasi

Pada area utama ditampilkan ringkasan nilai metrik evaluasi dalam bentuk kartu informasi, meliputi Silhouette Score, Sum of Squared Error (SSE), dan Davies-Bouldin Index (DBI) untuk mengukur kualitas hasil clustering. Di bawahnya terdapat tabel perbandingan metrik evaluasi yang menampilkan hasil masing-masing algoritma beserta metode yang memberikan performa terbaik pada setiap metrik. Selain itu, tersedia kesimpulan evaluasi yang menjelaskan algoritma dengan kinerja terbaik berdasarkan hasil pengujian, serta ringkasan perbandingan utama yang memudahkan pengguna melihat perbedaan nilai evaluasi antara K-Means dan Hierarchical Clustering secara cepat. Wireframe ini dirancang agar proses analisis dan pemilihan algoritma terbaik dapat dilakukan secara lebih mudah, objektif, dan informatif.

9. Halaman Laporan

Laporan dan export hasil
Ringkasan seluruh komponen hasil penelitian — ekspor ke PDF atau Excel

Ekspor PDF Ekspor Excel Cetak

Ringkasan komponen laporan penelitian

Komponen	Keterangan	Status
Dataset	33 kab/kota, IPM 2020–2025 (BPS Sumut)	Tersedia
Preprocessing	0 missing value, Min-Max Scaling	Selesai
K-Means (k=3)	Silhouette 0.72, DBI 0.41, SSE 12.34	Selesai
Hierarchical (k=3)	Silhouette 0.68, DBI 0.49, Ward Linkage	Selesai
Visualisasi	Scatter plot & dendrogram	Tersedia
Evaluasi	K-Means lebih unggul pada semua metrik	Selesai

Ringkasan cluster K-Means

Cluster	Kategori	Jumlah	Rata-rata IPM
1	Tinggi	7	76.40

Rekomendasi algoritma
K-Means Clustering direkomendasikan
Menghasilkan Silhouette Score 0.72 dan DBI 0.41 — lebih optimal dibanding Hierarchical Clustering pada data IPM BPS Sumatera Utara

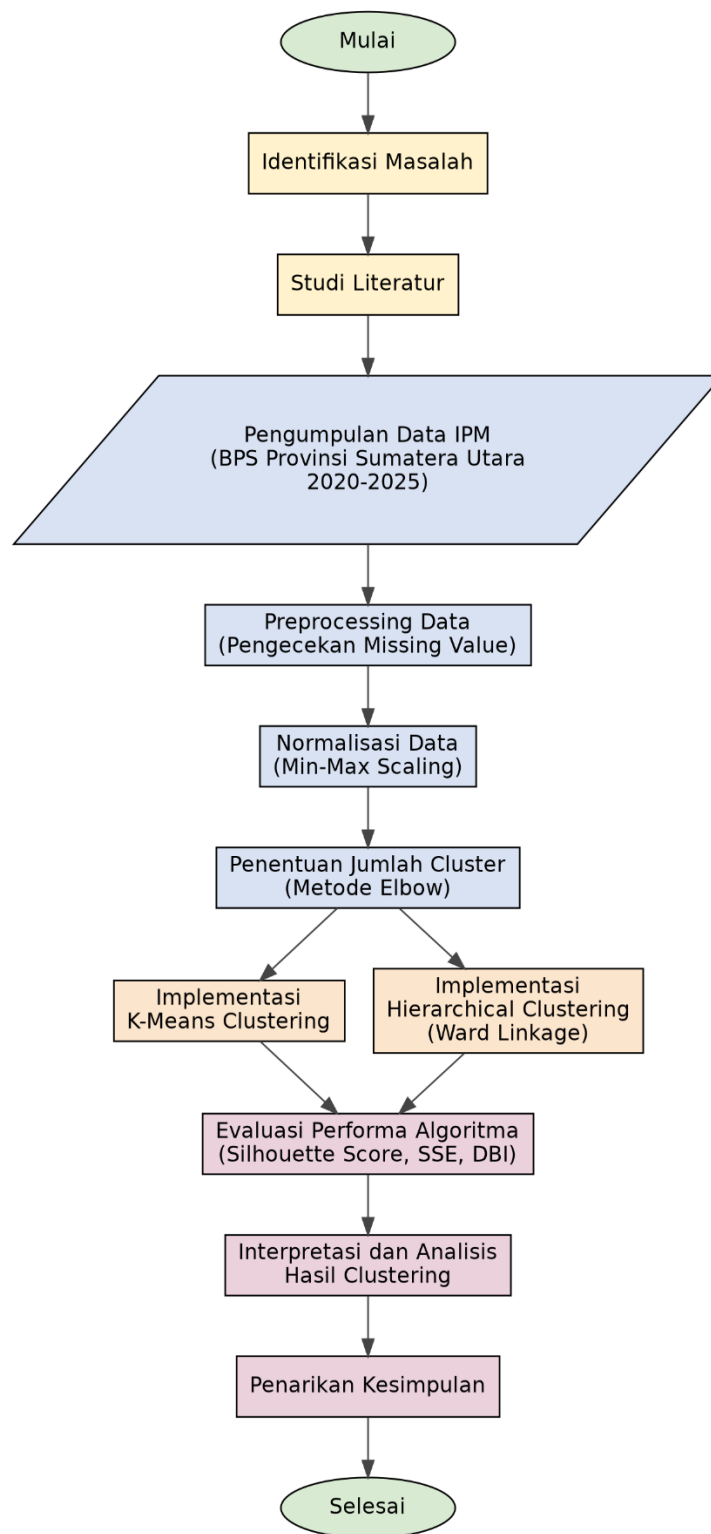
Gambar 3.20 Halaman Laporan

Pada area utama tersedia tombol Ekspor PDF, Ekspor Excel, dan Cetak yang memungkinkan pengguna menghasilkan laporan akhir dalam berbagai format. Selanjutnya, sistem menampilkan ringkasan komponen penelitian yang berisi informasi mengenai dataset, hasil preprocessing, proses clustering menggunakan K-Means dan Hierarchical Clustering, visualisasi, serta hasil evaluasi beserta status penyelesaiannya.

3.15 Alur Penelitian.

Alur penelitian ini menggambarkan keseluruhan tahapan yang dilakukan mulai dari identifikasi masalah hingga penarikan kesimpulan, sebagaimana ditunjukkan pada Gambar 3.21. Tahapan tersebut dijelaskan sebagai berikut: (1) identifikasi masalah untuk merumuskan permasalahan penelitian; (2) studi literatur untuk mengkaji teori dan penelitian terdahulu yang relevan; (3) pengumpulan data IPM 33

kabupaten/kota dari BPS Provinsi Sumatera Utara periode 2020–2025; (4) preprocessing data melalui pengecekan missing value; (5) normalisasi data menggunakan Min-Max Scaling; (6) penentuan jumlah cluster optimal menggunakan Metode Elbow; (7) implementasi algoritma K-Means dan Hierarchical Clustering secara paralel pada data yang telah dinormalisasi; (8) evaluasi performa kedua algoritma menggunakan Silhouette Score, SSE, dan Davies-Bouldin Index (DBI); (9) interpretasi dan analisis karakteristik hasil clustering; hingga (10) penarikan kesimpulan penelitian.



Gambar 3. 21 Alur Penelitian.

3.16 Contoh Perhitungan Manual

Untuk memberikan gambaran mengenai cara kerja algoritma secara matematis, berikut disajikan contoh perhitungan manual menggunakan 5 (lima) wilayah sampel yang mewakili karakteristik data penelitian, yaitu Kota Medan, Toba, Simalungun, Langkat, dan Nias Barat. Perhitungan manual ini mencakup normalisasi data, perhitungan jarak Euclidean, algoritma K-Means, algoritma Hierarchical Clustering (Ward Linkage), serta evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index (DBI). Hasil perhitungan manual ini digunakan untuk memverifikasi kebenaran logika algoritma yang diimplementasikan pada sistem.

3.16.1 Data Contoh

Data IPM kelima wilayah sampel periode 2020–2025 disajikan pada Tabel 3.11 berikut.

Tabel 3.11 Data Sampel untuk Perhitungan Manual

Wilayah	2020	2021	2022	2023	2024	2025
Kota Medan	80,98	81,58	82,13	82,61	83,23	83,74
Toba	75,16	77,03	77,58	77,83	78,44	79,17
Simalungun	73,25	74,33	74,68	75,17	75,88	76,48
Langkat	71,00	73,50	74,03	74,33	74,85	75,58
Nias Barat	61,51	63,05	63,96	64,68	65,66	66,65

3.16.2 Perhitungan Normalisasi Min-Max Manual

Normalisasi dilakukan pada setiap kolom (tahun) menggunakan rumus Min-Max Scaling: $X' = (X - X_{\min}) / (X_{\max} - X_{\min})$. Sebagai contoh, untuk kolom

tahun 2020, diperoleh $X_{\min} = 61,51$ (Nias Barat) dan $X_{\max} = 80,98$ (Kota Medan), sehingga perhitungan untuk setiap wilayah adalah sebagai berikut.

$$\text{Kota Medan: } X' = (80,98 - 61,51) / (80,98 - 61,51) = 1,0000$$

$$\text{Toba: } X' = (75,16 - 61,51) / (80,98 - 61,51) = 13,65 / 19,47 = 0,7011$$

$$\text{Simalungun: } X' = (73,25 - 61,51) / 19,47 = 0,6030$$

$$\text{Langkat: } X' = (71,00 - 61,51) / 19,47 = 0,4874$$

$$\text{Nias Barat: } X' = (61,51 - 61,51) / 19,47 = 0,0000$$

Perhitungan yang sama dilakukan untuk kolom tahun 2021 hingga 2025, sehingga diperoleh hasil normalisasi lengkap pada Tabel 3.12.

Tabel 3.12 Hasil Normalisasi Min-Max Manual

Wilayah	2020	2021	2022	2023	2024	2025
Kota Medan	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Toba	0,7011	0,7545	0,7496	0,7334	0,7274	0,7326
Simalungun	0,6030	0,6087	0,5900	0,5851	0,5817	0,5752
Langkat	0,4874	0,5640	0,5542	0,5382	0,5231	0,5225
Nias Barat	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

3.16.3 Perhitungan Jarak Euclidean Manual

Jarak antarwilayah dihitung menggunakan data yang telah dinormalisasi dengan rumus jarak Euclidean: $d(x,y) = \sqrt{(\sum (x_i - y_i)^2)}$. Sebagai contoh, jarak antara Kota Medan dan Toba dihitung sebagai berikut.

$$d(\text{Kota Medan, Toba}) = \sqrt{[(1,0000-0,7011)^2 + (1,0000-0,7545)^2 + (1,0000-0,7496)^2 + (1,0000-0,7334)^2 + (1,0000-0,7274)^2 + (1,0000-0,7326)^2]}$$

$$d(\text{Kota Medan, Toba}) = \sqrt{[0,0894 + 0,0603 + 0,0627 + 0,0711 + 0,0743 + 0,0715]} = \sqrt{0,4293} = 0,6552$$

Perhitungan yang sama dilakukan untuk seluruh pasangan wilayah, sehingga diperoleh matriks jarak lengkap pada Tabel 3.13.

Tabel 3.13 Matriks Jarak Euclidean Antarwilayah

Wilayah	Kota Medan	Toba	Simalungun	Langkat	Nias Barat
Kota Medan	0,0000	0,6552	1,0032	1,1491	2,4495
Toba	0,6552	0,0000	0,3526	0,4941	1,7962
Simalungun	1,0032	0,3526	0,0000	0,1583	1,4470
Langkat	1,1491	0,4941	0,1583	0,0000	1,3035
Nias Barat	2,4495	1,7962	1,4470	1,3035	0,0000

3.16.4 Perhitungan K-Means Manual

Perhitungan K-Means manual dilakukan dengan $k=3$, menggunakan Kota Medan, Simalungun, dan Nias Barat sebagai centroid awal (mewakili dugaan awal cluster Tinggi, Sedang, dan Rendah).

Iterasi 1 – Perhitungan jarak tiap wilayah ke centroid awal:

Tabel 3.14 Iterasi 1 – Jarak ke Centroid Awal

Wilayah	Jarak ke C1 (Medan)	Jarak ke C2 (Simalungun)	Jarak ke C3 (Nias Barat)	Cluster Terpilih
Kota Medan	0,0000	1,0032	2,4495	C1 (Tinggi)
Toba	0,6552	0,3526	1,7962	C2 (Sedang)
Simalungun	1,0032	0,0000	1,4470	C2 (Sedang)
Langkat	1,1491	0,1583	1,3035	C2 (Sedang)
Nias Barat	2,4495	1,4470	0,0000	C3 (Rendah)

Berdasarkan Tabel 3.14, setiap wilayah dimasukkan ke cluster dengan jarak terpendek. Selanjutnya, centroid diperbarui dengan menghitung rata-rata anggota tiap cluster: C1 tetap beranggotakan Kota Medan saja (centroid tidak berubah), C2 beranggotakan Toba, Simalungun, dan Langkat sehingga centroid baru dihitung sebagai rata-rata ketiganya, dan C3 tetap beranggotakan Nias Barat saja.

Iterasi 2 – Perhitungan jarak dengan centroid yang telah diperbarui:

Tabel 3.15 Iterasi 2 – Jarak ke Centroid yang Diperbarui

Wilayah	Jarak ke C1	Jarak ke C2 (baru)	Jarak ke C3	Cluster Terpilih
Kota Medan	0,0000	0,9354	2,4495	C1 (Tinggi)
Toba	0,6552	0,2812	1,7962	C2 (Sedang)
Simalungun	1,0032	0,0779	1,4470	C2 (Sedang)
Langkat	1,1491	0,2145	1,3035	C2 (Sedang)
Nias Barat	2,4495	1,5152	0,0000	C3 (Rendah)

Karena hasil pengelompokan pada iterasi 2 sama dengan iterasi 1 (tidak ada anggota cluster yang berpindah), maka proses iterasi dihentikan dan dinyatakan konvergen. Hasil akhir perhitungan manual menunjukkan Kota Medan masuk Cluster Tinggi, Toba/Simalungun/Langkat masuk Cluster Sedang, dan Nias Barat masuk Cluster Rendah — konsisten dengan pola pengelompokan yang dihasilkan oleh sistem pada data 33 kabupaten/kota secara keseluruhan (lihat subbab 4.5).

3.16.5 Perhitungan Hierarchical Clustering (Ward Linkage) Manual

Perhitungan Hierarchical Clustering manual menggunakan metode Ward Linkage, dengan rumus kenaikan variansi: $\Delta E = (n_A \cdot n_B) / (n_A + n_B) \times \|\mu_A - \mu_B\|^2$. Pada setiap tahap, dicari pasangan cluster dengan nilai ΔE terkecil untuk digabungkan. Proses penggabungan (agglomerative) dari 5 cluster individual hingga menjadi 1 cluster disajikan pada Tabel 3.16.

Tabel 3.16 Tahapan Penggabungan Cluster (Ward Linkage) Manual

Tahap	Cluster yang Digabungkan	Nilai ΔE (Ward)
1	Simalungun + Langkat	0,01252
2	Toba + (Simalungun+Langkat)	0,11864
3	Kota Medan + (Toba+Simalungun+Langkat)	0,65622
4	Nias Barat + (semua wilayah lain)	2,44632

Berdasarkan Tabel 3.16, wilayah yang pertama kali digabungkan adalah Simalungun dan Langkat karena memiliki nilai ΔE terkecil (paling mirip), diikuti oleh Toba, kemudian Kota Medan bergabung terakhir sebelum keseluruhan wilayah, dan Nias Barat menjadi wilayah paling berbeda karena bergabung paling akhir dengan nilai ΔE terbesar. Hasil penggabungan bertahap ini konsisten dengan hasil algoritma K-Means, yaitu Kota Medan cenderung terpisah sebagai kelompok Tinggi, Toba/Simalungun/Langkat cenderung berkelompok sebagai kelompok Sedang, dan Nias Barat sebagai kelompok Rendah yang paling berbeda karakteristiknya.

3.16.6 Perhitungan Evaluasi Manual (Silhouette Score dan DBI)

Evaluasi manual dilakukan terhadap hasil pengelompokan K-Means pada subbab 3.16.4, yaitu $C1=\{\text{Kota Medan}\}$, $C2=\{\text{Toba, Simalungun, Langkat}\}$, dan $C3=\{\text{Nias Barat}\}$.

a. Silhouette Score

Silhouette Score dihitung dengan rumus $s(i) = (b(i) - a(i)) / \max(a(i), b(i))$, dengan $a(i)$ adalah rata-rata jarak titik i ke anggota lain dalam cluster yang sama, dan $b(i)$ adalah rata-rata jarak terkecil titik i ke anggota cluster lain. Sebagai contoh perhitungan untuk Simalungun (anggota $C2$):

$$a(\text{Simalungun}) = \text{rata-rata jarak ke Toba dan Langkat} = (0,3526 + 0,1583) / 2 = 0,2554$$

$$b(\text{Simalungun}) = \text{rata-rata jarak terkecil ke cluster lain} = \min(\text{jarak ke } C1, \text{ jarak ke } C3) = \min(1,0032 ; 1,4470) = 1,0032$$

$$s(\text{Simalungun}) = (1,0032 - 0,2554) / \max(0,2554 ; 1,0032) = 0,7478 / 1,0032 = 0,7454$$

Perhitungan yang sama dilakukan untuk seluruh wilayah, dengan catatan bahwa untuk cluster beranggota tunggal ($C1$ dan $C3$), nilai $a(i) = 0$ berdasarkan konvensi karena tidak ada anggota pembanding dalam cluster yang sama, sehingga $s(i) = 0$. Hasil lengkap disajikan pada Tabel 3.17.

Tabel 3.17 Perhitungan Silhouette Score Manual per Wilayah

Wilayah	a(i)	b(i)	s(i)
Kota Medan	0,0000	0,9358	0,0000
Toba	0,4233	0,6552	0,3539
Simalungun	0,2554	1,0032	0,7454
Langkat	0,3262	1,1491	0,7162

Nias Barat	0,0000	1,5155	0,0000
------------	--------	--------	--------

Rata-rata Silhouette Score = $(0,0000 + 0,3539 + 0,7454 + 0,7162 + 0,0000) / 5 = 0,3631$. Nilai rata-rata pada contoh 5 wilayah ini lebih rendah dibandingkan hasil sistem pada 33 wilayah (0,6705) karena pengaruh dua cluster beranggota tunggal (C1 dan C3) yang secara konvensi bernilai $s(i)=0$, sehingga menurunkan rata-rata keseluruhan. Hal ini merupakan keterbatasan wajar dari sampel kecil dan bukan indikasi kesalahan perhitungan.

b. Davies-Bouldin Index (DBI)

DBI dihitung dengan terlebih dahulu menentukan S_i (rata-rata jarak anggota cluster ke centroidnya) untuk setiap cluster:

Tabel 3.18 Nilai S_i per Cluster

Cluster	S_i (rata-rata jarak ke centroid)
C1 (Tinggi)	0,0000
C2 (Sedang)	0,1912
C3 (Rendah)	0,0000

Selanjutnya dihitung $R_{ij} = (S_i + S_j) / M_{ij}$ untuk setiap pasangan cluster, dengan M_{ij} adalah jarak antarcentroid, kemudian diambil nilai maksimum R_{ij} untuk tiap cluster:

$$R(C1) = 0,2044 ; R(C2) = 0,2044 ; R(C3) = 0,1262$$

$$DBI = \text{rata-rata } R = (0,2044 + 0,2044 + 0,1262) / 3 = 0,1783$$

Nilai DBI sebesar 0,1783 pada contoh 5 wilayah ini menunjukkan pemisahan antarcluster yang sangat baik (semakin kecil nilai DBI, semakin baik

pemisahannya), konsisten dengan hasil evaluasi sistem pada 33 wilayah yang juga menunjukkan nilai DBI rendah (0,3942).

Berdasarkan keseluruhan perhitungan manual pada subbab 3.16, dapat disimpulkan bahwa logika algoritma K-Means, Hierarchical Clustering, dan metrik evaluasi yang diimplementasikan pada sistem web telah sesuai dengan perhitungan matematis yang seharusnya, sehingga hasil yang ditampilkan sistem pada Bab IV dapat dipertanggungjawabkan kebenarannya.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Pengumpulan Data

Penelitian ini menggunakan data sekunder Indeks Pembangunan Manusia (IPM) yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data yang digunakan mencakup nilai IPM tahunan seluruh 33 kabupaten/kota di Provinsi Sumatera Utara selama periode 2020–2025. Data diperoleh dari Buku Indeks Pembangunan Manusia Provinsi Sumatera Utara yang diterbitkan oleh BPS dan telah disusun ke dalam format tabel untuk keperluan pengolahan lebih lanjut.

Nilai IPM merupakan indeks komposit yang mencerminkan tiga dimensi utama pembangunan manusia, yaitu: (1) kesehatan yang direpresentasikan oleh Umur Harapan Hidup (UHH); (2) pendidikan yang direpresentasikan oleh Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS); serta (3) standar hidup layak yang direpresentasikan oleh Pengeluaran per Kapita Disesuaikan. Nilai IPM per tahun digunakan sebagai atribut dalam proses clustering. Data lengkap disajikan pada Tabel 4.1 berikut.

Tabel 4.1 Data IPM Kabupaten/Kota Provinsi Sumatera Utara Periode 2020–2025

No	Kabupaten/Kota	2020	2021	2022	2023	2024	2025
1	Nias	61,93	63,37	64,30	65,15	66,13	66,79
2	Mandailing Natal	66,79	71,31	72,17	72,65	73,44	74,26
3	Tapanuli Selatan	70,12	73,37	73,96	74,58	75,18	75,78
4	Tapanuli Tengah	69,23	71,48	72,18	72,77	73,34	73,97
5	Tapanuli Utara	73,47	76,06	76,43	76,86	77,48	78,09
6	Toba	75,16	77,03	77,58	77,83	78,44	79,17

No	Kabupaten/Kota	2020	2021	2022	2023	2024	2025
7	Labuhan Batu	72,01	73,16	73,96	74,70	75,20	75,79
8	Asahan	70,29	72,55	73,18	73,59	74,17	75,13
9	Simalungun	73,25	74,33	74,68	75,17	75,88	76,48
10	Dairi	71,57	73,89	74,60	75,18	75,88	76,68
11	Karo	74,43	75,81	76,37	76,88	77,49	78,10
12	Deli Serdang	75,44	76,21	76,86	77,16	77,56	78,29
13	Langkat	71,00	73,50	74,03	74,33	74,85	75,58
14	Nias Selatan	61,89	63,24	64,05	64,98	65,87	66,74
15	Humbang Hasundutan	68,87	71,28	72,05	72,49	73,33	73,93
16	Pakpak Barat	67,59	70,65	71,57	72,30	73,09	73,89
17	Samosir	70,63	71,75	72,57	72,93	73,77	74,63
18	Serdang Bedagai	70,24	72,25	72,89	73,40	74,15	74,69
19	Batu Bara	68,36	70,80	71,75	72,56	73,14	73,92
20	Paluta	69,85	71,96	72,77	73,45	74,16	75,12
21	Padang Lawas	68,25	70,48	71,41	72,16	73,10	73,90
22	Labusel	71,40	73,15	73,59	74,23	75,11	75,74
23	Labura	71,61	73,73	74,61	75,45	76,20	76,71
24	Nias Utara	62,36	63,63	64,55	65,44	66,42	67,38
25	Nias Barat	61,51	63,05	63,96	64,68	65,66	66,65
26	Kota Sibolga	73,63	75,87	76,69	77,07	77,52	78,11
27	Kota Tanjung Balai	68,65	73,78	74,73	75,42	75,95	76,70
28	Kota P. Siantar	78,75	79,17	79,71	80,46	81,17	81,71
29	Kota Tebing Tinggi	75,17	76,65	77,39	78,17	78,68	79,26
30	Kota Medan	80,98	81,58	82,13	82,61	83,23	83,74
31	Kota Binjai	75,89	76,57	77,50	78,11	78,67	79,20
32	Kota P. Sidempuan	75,22	77,19	77,73	78,10	78,61	79,19
33	Kota Gunung Sitoli	69,31	70,51	71,11	71,55	72,30	73,29

Berdasarkan Tabel 4.1, seluruh 33 kabupaten/kota mengalami peningkatan nilai IPM setiap tahunnya selama periode 2020–2025. Kota Medan secara konsisten menempati posisi tertinggi dengan IPM mencapai 83,74 pada tahun 2025, sedangkan Nias Barat mencatat nilai terendah sebesar 66,65. Kesenjangan sekitar 17 poin ini mencerminkan adanya disparitas pembangunan manusia yang signifikan antarwilayah dan menjadi alasan pentingnya dilakukan pengelompokan wilayah berdasarkan karakteristik IPM.

4.2 Hasil Preprocessing Data

Sebelum dilakukan proses clustering, data IPM terlebih dahulu melalui tahapan preprocessing. Tahapan ini meliputi pemeriksaan missing value, data duplikat, serta verifikasi tipe data. Sebagaimana ditunjukkan pada Tabel 4.2, preprocessing menghasilkan 0 missing value dan seluruh data siap untuk diproses lebih lanjut.

4.2.1 Pemeriksaan Missing Value

Hasil pemeriksaan menunjukkan bahwa seluruh data IPM 33 kabupaten/kota untuk periode 2020–2025 tersedia secara lengkap tanpa ada nilai yang hilang, sehingga tidak diperlukan penanganan imputasi data.

4.2.2 Pemeriksaan Data Duplikat

Hasil pemeriksaan menunjukkan bahwa tidak terdapat data yang tercatat lebih dari satu kali. Seluruh 33 entri bersifat unik sehingga tidak diperlukan proses penghapusan baris duplikat.

4.2.3 Pemeriksaan Tipe Data

Seluruh kolom nilai IPM bertipe numerik (float) yang merupakan format yang sesuai untuk proses analisis clustering.

Tabel 4.2 Rekapitulasi Hasil Preprocessing Data

Pemeriksaan	Hasil	Keterangan
Missing Value	0	Tidak ada data yang hilang pada seluruh atribut
Data Duplikat	0	Tidak ada baris yang terduplikasi
Tipe Data	Numerik (float)	Seluruh variabel bertipe numerik dan valid untuk clustering
Jumlah Baris	33	33 Kabupaten/Kota Provinsi Sumatera Utara
Jumlah Kolom	6	Nilai IPM tahun 2020, 2021, 2022, 2023, 2024, dan 2025

Berdasarkan Tabel 4.2, data IPM telah memenuhi standar kualitas yang diperlukan. Tidak ditemukan missing value maupun data duplikat sehingga data dapat langsung dilanjutkan ke tahap normalisasi.

4.3 Hasil Transformasi Data (Normalisasi Min-Max Scaling)

Data ditransformasi menggunakan metode Min-Max Scaling untuk menyeragamkan nilai ke skala 0 hingga 1 agar setiap atribut memberikan kontribusi yang seimbang dalam proses clustering. Rumus yang digunakan adalah:

$$X' = (X - X \text{ min}) / (X \text{ max} - X \text{ min})$$

Keterangan: X' adalah nilai normalisasi, X adalah nilai IPM asli, X min adalah nilai minimum, dan X max adalah nilai maksimum pada atribut tersebut. Tabel 4.3 menyajikan contoh hasil normalisasi menggunakan data BPS asli.

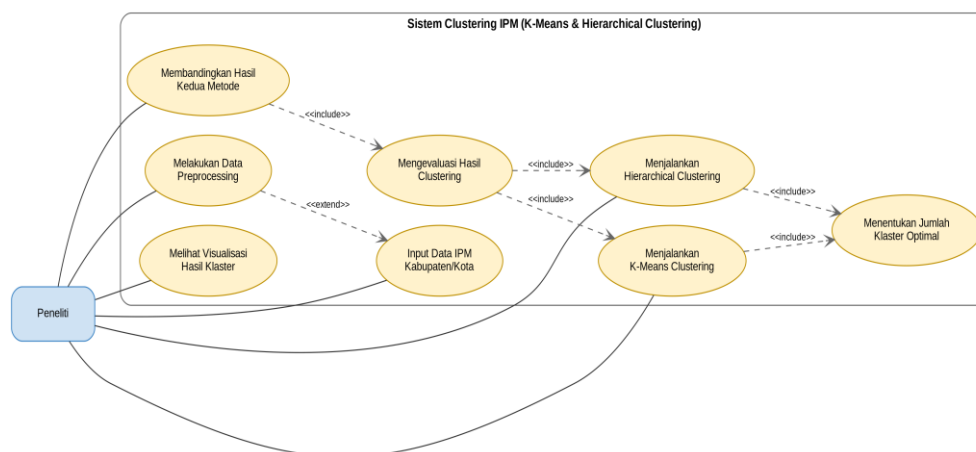
Tabel 4.3 Contoh Hasil Normalisasi Data (Min-Max Scaling)

Kabupaten/Kota	IPM 2020 (Asli)	IPM 2025 (Asli)	IPM 2020 (Normal)	IPM 2025 (Normal)
Kota Medan	80,98	83,74	1,000	1,000
Nias Barat	61,51	66,65	0,000	0,000
Kota P. Siantar	78,75	81,71	0,890	0,890
Tapanuli Utara	73,47	78,09	0,620	0,680
Nias	61,93	66,79	0,022	0,008

Kota Medan memiliki nilai normalisasi 1,000 sebagai wilayah dengan IPM tertinggi, sedangkan Nias Barat memiliki nilai mendekati 0,000 sebagai wilayah dengan IPM terendah. Data yang telah dinormalisasi selanjutnya digunakan sebagai masukan algoritma clustering.

4.4 Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster optimal dilakukan menggunakan Metode Elbow. Metode ini menghitung nilai SSE untuk berbagai nilai k, kemudian memilih nilai k di mana penurunan SSE mulai melambat membentuk pola "siku" pada grafik. Gambar 4.1 menunjukkan hasil Metode Elbow dari sistem.



Gambar 4.1 Grafik Metode Elbow (SSE vs k)

Berdasarkan hasil Metode Elbow, penurunan nilai Sum of Squared Error (SSE) berlangsung cukup tajam pada nilai $k = 2$ hingga $k = 3$, kemudian mulai melandai pada $k = 4$. Meskipun sistem menandai titik elbow pada $k = 4$, penelitian ini menggunakan $k = 3$ karena beberapa pertimbangan. Pertama, tujuan penelitian adalah mengelompokkan kabupaten/kota ke dalam tiga kategori pembangunan manusia, yaitu tinggi, sedang, dan rendah, sehingga hasil lebih mudah diinterpretasikan. Kedua, hasil evaluasi menggunakan Silhouette Score menunjukkan kualitas cluster yang baik pada $k = 3$ dengan nilai 0,6705. Ketiga, jumlah anggota pada masing-masing cluster relatif seimbang sehingga memudahkan proses analisis dan penyusunan rekomendasi kebijakan. Oleh karena itu, jumlah cluster yang digunakan pada penelitian ini ditetapkan sebanyak tiga cluster untuk kedua algoritma agar proses perbandingan dapat dilakukan secara konsisten.

4.5 Hasil Implementasi Algoritma K-Means

4.5.1 Proses Clustering K-Means

Algoritma K-Means diimplementasikan dengan parameter $k=3$, inisialisasi centroid menggunakan k-means++ (Arthur & Vassilvitskii, 2007), random state 42, dan maksimum iterasi 300. Data masukan adalah data yang telah dinormalisasi. Proses iterasi berlangsung hingga konvergen, yaitu ketika tidak terdapat perubahan anggota cluster antariterasi.

4.5.2 Hasil Clustering K-Means

Berdasarkan hasil implementasi K-Means, seluruh 33 kabupaten/kota berhasil dikelompokkan ke dalam 3 cluster sebagaimana disajikan pada Tabel 4.4 berikut.

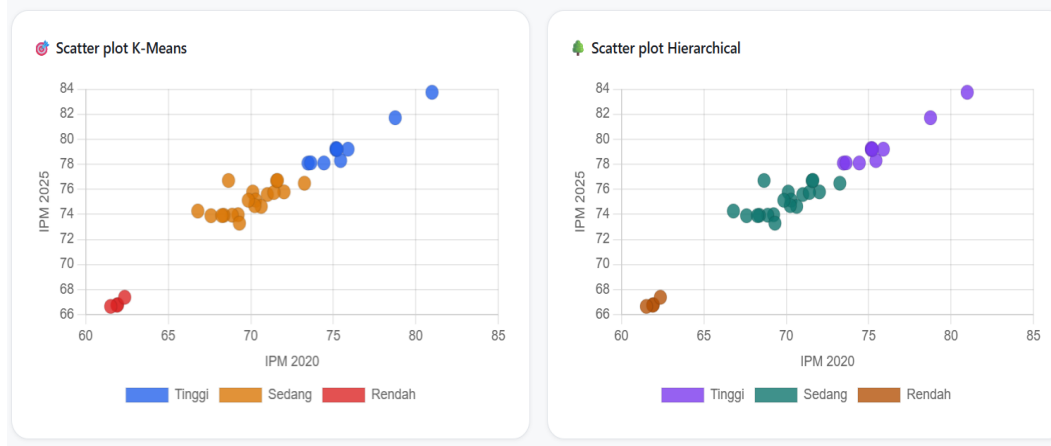
Tabel 4.4 Hasil Pengelompokan Kabupaten/Kota Menggunakan K-Means**(k=3)**

Cluster	Kategori	Jumlah	Anggota Cluster	Rata-rata IPM
1	Tinggi	10	Karo, Deli Serdang, Kota Sibolga, Kota Tebing Tinggi, Kota Medan, Kota Binjai, Kota P. Sidempuan, Kota P. Siantar, Tapanuli Utara, Toba	77,93
2	Sedang	19	Tapanuli Selatan, Tapanuli Tengah, Labuhan Batu, Asahan, Simalungun, Dairi, Langkat, Humbang Hasundutan, Pakpak Barat, Samosir, Serdang Bedagai, Batu Bara, Paluta, Padang Lawas, Labusel, Labura, Kota Tanjung Balai, Mandailing Natal, Kota Gunung Sitoli	73,05
3	Rendah	4	Nias, Nias Selatan, Nias Utara, Nias Barat	64,57

Cluster 1 (Tinggi) terdiri atas 10 wilayah dengan rata-rata IPM 77,93, didominasi kota-kota besar. Cluster 2 (Sedang) merupakan kelompok terbesar dengan 19 wilayah dan rata-rata IPM 73,05, termasuk di dalamnya Paluta (Padang Lawas Utara), Labusel (Labuhan Batu Selatan), dan Labura (Labuhan Batu Utara). Cluster 3 (Rendah) terdiri atas 4 kabupaten di Kepulauan Nias dengan rata-rata IPM 64,57. Visualisasi hasil K-Means ditunjukkan pada Gambar 4.2 berikut.

Visualisasi Cluster

Scatter plot K-Means dan Hierarchical Clustering (IPM 2020 vs IPM 2025)



Gambar 4.2 Scatter Plot Hasil K-Means dan Hierarchical Clustering (IPM 2020 vs IPM 2025)

Berdasarkan Gambar 4.2, scatter plot K-Means (kiri) menunjukkan pemisahan yang jelas antara ketiga cluster. Titik merah (Cluster Rendah) berada di kiri bawah, titik oranye (Cluster Sedang) berkelompok di tengah, dan titik biru (Cluster Tinggi) berada di kanan atas, yang mencerminkan pola peningkatan IPM yang konsisten dari 2020 ke 2025.

4.5.3 Nilai Evaluasi K-Means

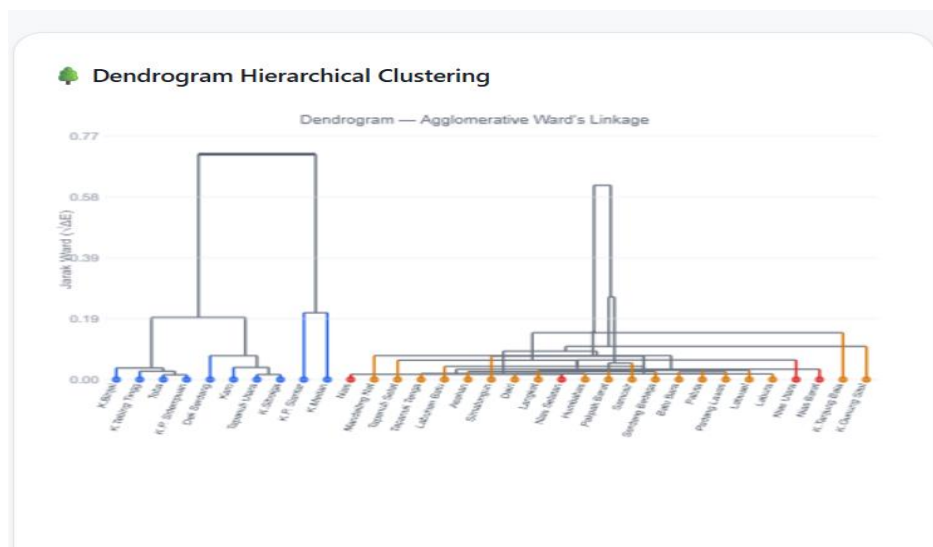
Hasil evaluasi algoritma K-Means menghasilkan Silhouette Score sebesar **0,6705** dan nilai SSE sebesar 1,1492. Nilai Silhouette Score berada pada rentang 0,51–0,70, yang menurut kategorisasi Kaufman & Rousseeuw (1990) menunjukkan bahwa struktur cluster yang terbentuk tergolong baik (reasonable structure). Nilai tersebut mengindikasikan bahwa sebagian besar objek memiliki tingkat kemiripan yang tinggi terhadap anggota dalam cluster yang sama serta memiliki pemisahan yang cukup jelas terhadap cluster lainnya. Selain itu, nilai SSE sebesar 1,1492 menunjukkan bahwa variasi data di dalam masing-masing cluster relatif kecil

sehingga anggota setiap cluster bersifat cukup homogen. Semakin kecil nilai SSE, semakin baik tingkat kekompakan cluster yang dihasilkan oleh algoritma K-Means. Selain itu, algoritma K-Means juga menghasilkan nilai Davies-Bouldin Index (DBI) sebesar 0,3942, yang mendekati nol dan mengindikasikan pemisahan antarcluster yang baik.

4.6 Hasil Implementasi Algoritma Hierarchical Clustering

4.6.1 Proses dan Dendrogram Hierarchical Clustering

Hierarchical Clustering diimplementasikan menggunakan metode Agglomerative dengan Ward Linkage (Ward, 1963). Proses dimulai dari 33 cluster individual dan dilakukan 30 tahap penggabungan ($n - k = 33 - 3 = 30$) hingga terbentuk 3 cluster final. Seluruh proses divisualisasikan dalam bentuk dendrogram pada Gambar 4.3 berikut.



Gambar 4.3 Dendrogram Hierarchical Clustering (Ward's Linkage)

Berdasarkan Gambar 4.3, dendrogram menunjukkan tiga kelompok utama yang terbentuk, ditandai dengan warna biru (Cluster Tinggi), oranye (Cluster

Sedang), dan merah (Cluster Rendah). Garis vertikal paling tinggi pada dendrogram berada pada jarak Ward sekitar 0,77, yang menunjukkan adanya pemisahan yang signifikan antara ketiga kelompok tersebut.

4.6.2 Hasil Clustering Hierarchical

Berdasarkan hasil pemotongan dendrogram, seluruh 33 kabupaten/kota dikelompokkan ke dalam 3 cluster yang disajikan pada Tabel 4.5 berikut.

Tabel 4.5 Hasil Pengelompokan Kabupaten/Kota Menggunakan Hierarchical Clustering (k=3)

Cluster	Kategori	Jumlah	Anggota Cluster	Rata-rata IPM
1	Tinggi	10	Karo, Deli Serdang, Kota Sibolga, Kota Tebing Tinggi, Kota Medan, Kota Binjai, Kota P. Sidempuan, Kota P. Siantar, Tapanuli Utara, Toba	77,93
2	Sedang	19	Tapanuli Selatan, Tapanuli Tengah, Labuhan Batu, Asahan, Simalungun, Dairi, Langkat, Humbang Hasundutan, Pakpak Barat, Samosir, Serdang Bedagai, Batu Bara, Paluta, Padang Lawas, Labusel, Labura, Kota Tanjung Balai, Mandailing Natal, Kota Gunung Sitoli	73,05
3	Rendah	4	Nias, Nias Selatan, Nias Utara, Nias Barat	64,57

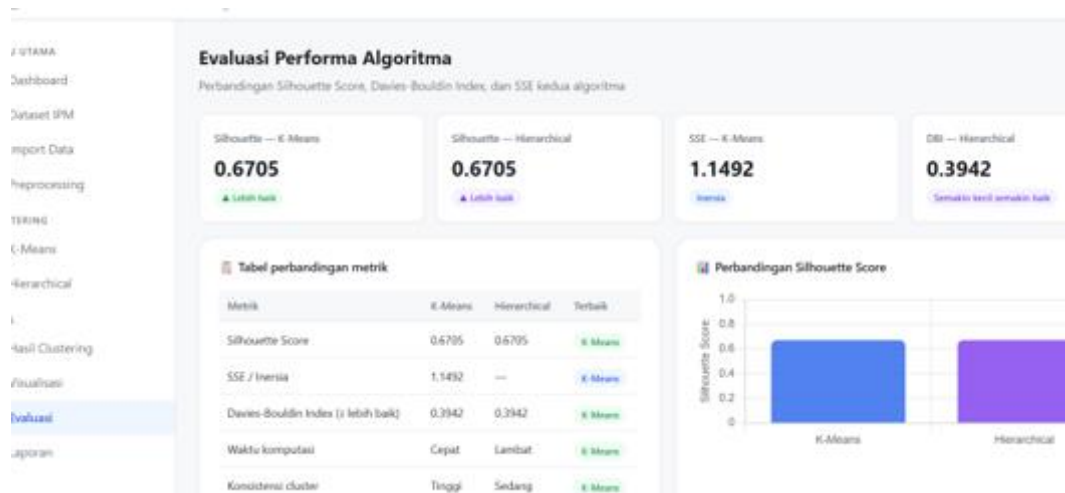
Hasil Hierarchical Clustering menghasilkan konfigurasi yang identik dengan K-Means, baik komposisi anggota maupun jumlah wilayah pada setiap cluster. Keidentikan ini memperkuat validitas hasil pengelompokan yang diperoleh.

4.6.3 Nilai Evaluasi Hierarchical Clustering

Hasil evaluasi algoritma Hierarchical Clustering menghasilkan Silhouette Score sebesar 0,6705 dan Davies–Bouldin Index (DBI) sebesar 0,3942. Nilai Silhouette Score menunjukkan bahwa kualitas cluster yang terbentuk berada pada kategori baik. Sementara itu, nilai DBI yang mendekati nol menunjukkan bahwa setiap cluster memiliki tingkat penyebaran internal yang rendah serta jarak antarcluster yang cukup jauh. Dengan demikian, hasil pengelompokan menggunakan Hierarchical Clustering dapat dikatakan memiliki kualitas yang baik karena menghasilkan cluster yang kompak sekaligus terpisah secara jelas.

4.7 Evaluasi Performa Algoritma

Evaluasi performa dilakukan untuk membandingkan kualitas hasil clustering kedua algoritma. Seluruh nilai evaluasi diperoleh langsung dari output sistem sebagaimana ditunjukkan pada Gambar 4.4 berikut.



Gambar 4.4 Tampilan Halaman Evaluasi Performa Algoritma

Tabel 4.6 Perbandingan Hasil Evaluasi Algoritma K-Means dan Hierarchical Clustering

Metrik Evaluasi	K-Means	Hierarchical	Terbaik
Silhouette Score (↑ lebih baik)	0,6705	0,6705	K-Means (skor sama, lebih efisien)
SSE / Inersia	1,1492	—	K-Means
Davies-Bouldin Index (↓ lebih baik)	0,3942	0,3942	K-Means (skor sama, lebih konsisten)
Kompleksitas Komputasi (teoritis)	$O(n)$ per iterasi	$O(n^2 \log n)$	K-Means (lebih efisien secara teoritis)
Skalabilitas untuk Data Berukuran Besar	Baik	Terbatas	K-Means

Berdasarkan Tabel 4.6 dan Gambar 4.4, kedua algoritma menghasilkan Silhouette Score dan DBI yang identik, sehingga dari sisi kualitas hasil pengelompokan keduanya setara. K-Means ditetapkan sebagai algoritma yang lebih direkomendasikan karena secara teoritis memiliki kompleksitas komputasi yang lebih rendah ($O(n)$ per iterasi dibandingkan $O(n^2 \log n)$ pada Hierarchical Clustering) sehingga lebih skalabel untuk data berukuran besar, selaras dengan kesimpulan evaluasi sistem yang menyatakan bahwa kualitas kedua algoritma setara dan K-Means direkomendasikan karena efisiensi komputasinya.

4.8 Interpretasi dan Analisis Hasil Clustering

Interpretasi dilakukan berdasarkan hasil K-Means sebagai algoritma yang direkomendasikan. Perlu dicatat bahwa proses clustering menggunakan nilai IPM seluruh periode 2020–2025 sebagai atribut (bukan hanya nilai tahun 2025 saja), sehingga kolom "Rata-rata IPM" pada tabel berikut merupakan rata-rata gabungan periode tersebut, bukan nilai IPM tahun 2025 semata. Pendekatan ini menjelaskan mengapa urutan anggota cluster tidak selalu identik dengan urutan nilai IPM 2025

saja—misalnya Tapanuli Utara tergabung dalam Cluster 1 (Tinggi) karena pola pertumbuhan IPM-nya sepanjang 2020–2025 paling mendekati karakteristik cluster tersebut, meskipun nilai IPM 2025-nya (78,09) berada di bawah beberapa anggota lain pada cluster yang sama.

Tabel 4.7 Karakteristik Cluster Berdasarkan Hasil K-Means

Karakteristik	Cluster 1 (Tinggi)	Cluster 2 (Sedang)	Cluster 3 (Rendah)
Rata-rata IPM 2020–2025	77,93	73,05	64,57
Jumlah Wilayah	10	19	4
IPM Tertinggi (2025)	83,74 (Kota Medan)	76,71 (Labura)	66,79 (Nias)
IPM Terendah (2025)	78,09 (Tapanuli Utara)	73,90 (Padang Lawas)	66,65 (Nias Barat)
Karakteristik Wilayah	Perkotaan dan kab. maju	Kabupaten daratan	Kepulauan Nias

4.8.1 Cluster 1 — Kategori IPM Tinggi

Cluster 1 terdiri atas 10 kabupaten/kota dengan rata-rata IPM 77,93. Anggotanya adalah Karo, Deli Serdang, Kota Sibolga, Kota Tebing Tinggi, Kota Medan, Kota Binjai, Kota P. Sidempuan, Kota P. Siantar, Tapanuli Utara, dan Toba. Kota Medan menjadi wilayah dengan IPM tertinggi mencapai 83,74 pada tahun 2025. Cluster ini didominasi kota-kota besar yang memiliki infrastruktur lebih baik, akses layanan pendidikan dan kesehatan yang lebih merata, serta aktivitas ekonomi yang lebih beragam.

4.8.2 Cluster 2 — Kategori IPM Sedang

Cluster 2 merupakan kelompok terbesar dengan 19 kabupaten/kota dan rata-rata IPM 73,05. Cluster ini mencakup sebagian besar kabupaten di daratan Sumatera Utara dengan basis ekonomi beragam mulai dari pertanian, perkebunan, hingga perdagangan. Nilai IPM seluruh anggota cluster ini terus meningkat dari tahun ke tahun, namun masih terdapat ruang untuk peningkatan terutama pada aspek pendidikan lanjutan dan pengeluaran per kapita.

4.8.3 Cluster 3 — Kategori IPM Rendah

Cluster 3 terdiri atas 4 kabupaten di Kepulauan Nias dengan rata-rata IPM 64,57, yaitu Kabupaten Nias (66,79), Nias Selatan (66,74), Nias Utara (67,38), dan Nias Barat (66,65). Kondisi geografis berupa wilayah kepulauan menjadi tantangan utama dalam penyediaan infrastruktur dan aksesibilitas layanan pendidikan serta kesehatan. Meskipun demikian, keempat kabupaten ini menunjukkan tren pertumbuhan IPM yang positif, dengan peningkatan rata-rata sekitar 4–5 poin selama periode 2020–2025.

4.9 Pembahasan

4.9.1 Pola Pertumbuhan IPM di Provinsi Sumatera Utara

Data BPS menunjukkan bahwa IPM rata-rata Provinsi Sumatera Utara meningkat dari 71,77 (2020) menjadi 76,47 (2025). Angka rata-rata provinsi ini merupakan angka resmi hasil perhitungan langsung BPS pada level provinsi (bukan rata-rata sederhana dari 33 kabupaten/kota pada Tabel 4.1), karena metode agregasi komponen IPM di tingkat provinsi berbeda dengan agregasi sederhana antarwilayah. Hasil clustering memperlihatkan bahwa meskipun pertumbuhan terjadi di seluruh wilayah, disparitas antarwilayah masih cukup besar. Kota Medan

yang mencapai 83,74 berbanding jauh dengan Nias Barat yang baru mencapai 66,65 pada tahun 2025. Pola clustering yang konsisten dengan kondisi geografis ini mengindikasikan bahwa faktor aksesibilitas dan ketersediaan infrastruktur memiliki hubungan erat dengan tingkat pembangunan manusia di Provinsi Sumatera Utara.

4.9.2 Perbandingan Performa Algoritma

Berdasarkan hasil penelitian, algoritma K-Means dan Hierarchical Clustering menghasilkan komposisi anggota cluster yang identik, yaitu terdiri atas 10 wilayah pada cluster tinggi, 19 wilayah pada cluster sedang, dan 4 wilayah pada cluster rendah. Selain itu, kedua algoritma juga menghasilkan nilai Silhouette Score sebesar 0,6705, yang menunjukkan kualitas cluster yang sama baiknya. Kesamaan hasil tersebut mengindikasikan bahwa struktur data IPM Provinsi Sumatera Utara memiliki pola pengelompokan yang cukup jelas sehingga dua algoritma dengan pendekatan berbeda mampu menghasilkan konfigurasi cluster yang sama.

Perbedaan utama kedua algoritma terletak pada mekanisme pembentukan cluster. K-Means membentuk cluster melalui proses iteratif dengan memperbarui centroid hingga konvergen, sedangkan Hierarchical Clustering membentuk struktur bertingkat (dendrogram) melalui proses penggabungan cluster secara bertahap menggunakan metode Ward Linkage. Meskipun hasil akhirnya identik, K-Means secara teoritis memiliki kompleksitas komputasi yang lebih rendah, yaitu $O(n)$ per iterasi dibandingkan $O(n^2 \log n)$ pada Hierarchical Clustering, sehingga lebih efisien dan lebih sesuai digunakan pada data berukuran besar. Oleh sebab itu, K-Means direkomendasikan sebagai algoritma utama dalam penelitian ini.

Temuan ini sejalan dengan penelitian Sikana & Wijayanto (2021) yang membandingkan metode partitioning dan hierarchical clustering pada data IPM

Indonesia, di mana kedua pendekatan cenderung menghasilkan pengelompokan yang relatif konsisten meskipun didasarkan pada mekanisme pembentukan cluster yang berbeda. Penelitian Rahmati & Wijayanto (2021) yang membandingkan K-Means, Fuzzy C-Means, dan Hierarchical Clustering pada data IPM juga menunjukkan bahwa evaluasi Silhouette Score dan Davies-Bouldin Index dapat menghasilkan nilai yang berdekatan antaralgoritma partitioning dan hierarchical apabila struktur data memiliki pola pengelompokan yang jelas. Kesamaan pola ini memperkuat argumen bahwa data IPM Provinsi Sumatera Utara memang memiliki struktur kelompok yang cukup tegas, sehingga hasil penelitian ini konsisten dengan temuan pada wilayah lain di Indonesia.

4.9.3 Implikasi terhadap Kebijakan Pembangunan

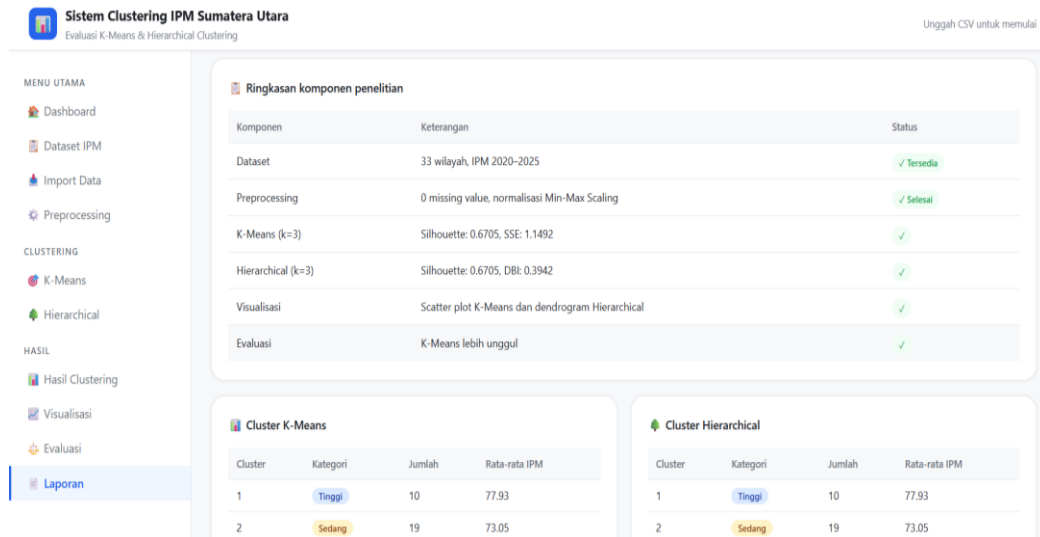
Hasil pengelompokan memberikan basis yang berguna bagi Pemerintah Provinsi Sumatera Utara dalam memprioritaskan alokasi sumber daya dan menyusun kebijakan pembangunan yang lebih tepat sasaran berdasarkan karakteristik masing-masing cluster. Cluster 3 (Kepulauan Nias) memerlukan program intervensi prioritas berupa pembangunan infrastruktur dasar serta perluasan jangkauan layanan pendidikan dan kesehatan. Cluster 2 memerlukan penguatan pada aspek diversifikasi ekonomi dan peningkatan kualitas pendidikan. Wilayah pada Cluster 1 dapat dijadikan model rujukan pembangunan dan berperan sebagai pusat pertumbuhan yang mendorong wilayah di sekitarnya melalui pemerataan investasi, transfer pengetahuan, dan kerja sama antardaerah. Pendekatan kebijakan yang terdiferensiasi berdasarkan karakteristik cluster diharapkan lebih efektif dalam mengurangi disparitas pembangunan manusia antarwilayah.

4.9.4 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam menginterpretasikan hasilnya. Pertama, penelitian ini hanya menggunakan nilai IPM komposit sebagai atribut clustering, tanpa melibatkan komponen penyusunnya secara terpisah (Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran per Kapita), sehingga belum dapat mengidentifikasi dimensi pembangunan spesifik yang paling memengaruhi pengelompokan suatu wilayah. Kedua, periode data yang digunakan terbatas pada enam tahun (2020-2025), sehingga pola tren jangka panjang di luar periode tersebut belum tercakup dalam analisis. Ketiga, penelitian ini hanya mencakup wilayah Provinsi Sumatera Utara sehingga hasil clustering dan karakteristik cluster yang terbentuk tidak dapat digeneralisasikan secara langsung ke provinsi lain yang memiliki karakteristik geografis dan sosial-ekonomi berbeda. Keterbatasan-keterbatasan ini dapat menjadi arah pengembangan bagi penelitian selanjutnya.

4.10 Implementasi Sistem

Implementasi sistem yang dibangun pada penelitian ini berhasil mengimplementasikan seluruh tahapan proses clustering secara otomatis melalui aplikasi berbasis web. Pengguna dapat mengunggah dataset IPM, melakukan preprocessing, menjalankan normalisasi menggunakan Min-Max Scaling, menentukan jumlah cluster menggunakan Metode Elbow, melakukan proses clustering menggunakan algoritma K-Means maupun Hierarchical Clustering, serta melihat hasil evaluasi performa masing-masing algoritma.



Gambar 4.5 Tampilan Halaman Laporan Sistem Web Clustering

Berdasarkan Gambar 4.5, sistem berhasil menjalankan seluruh proses dengan status "Selesai" pada seluruh komponen. K-Means menghasilkan Silhouette: 0,6705 dan SSE: 1,1492; Hierarchical menghasilkan Silhouette: 0,6705 dan DBI: 0,3942; kualitas kedua algoritma setara, dan sistem merekomendasikan K-Means karena lebih efisien secara komputasi. Tampilan antarmuka sistem telah sesuai dengan rancangan yang dijelaskan pada Bab III.

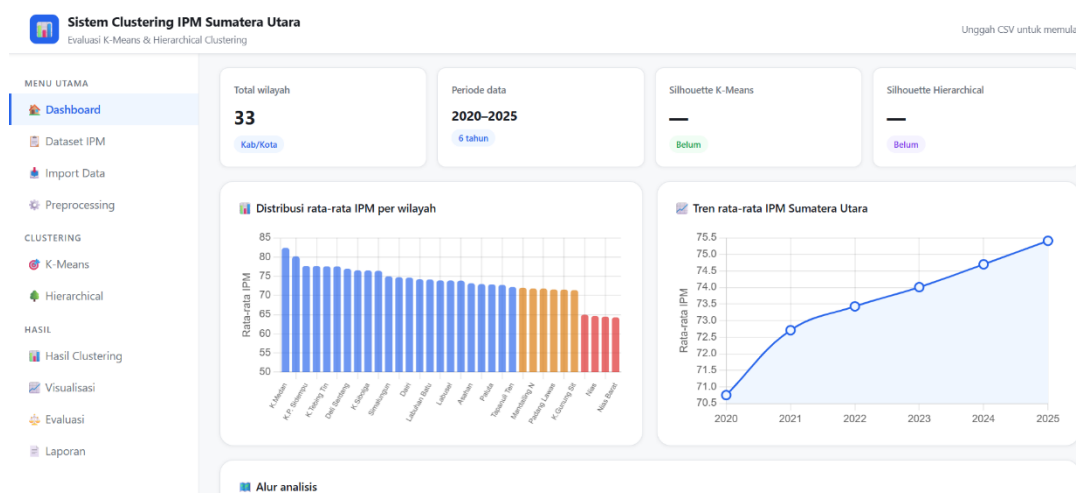
4.11 Tampilan Antarmuka Sistem Web (Seluruh Halaman)

Bagian ini menyajikan tampilan antarmuka seluruh halaman pada sistem web clustering IPM yang dikembangkan, mulai dari Dashboard hingga Laporan, sebagai dokumentasi implementasi sistem secara utuh sesuai rancangan pada Bab III.

4.11.1 Halaman Dashboard

Halaman Dashboard menampilkan ringkasan status analisis (total wilayah, periode data, Silhouette Score K-Means dan Hierarchical), grafik distribusi rata-

rata IPM per wilayah, grafik tren rata-rata IPM Sumatera Utara, serta pelacak progres tahapan analisis (Data dimuat, Preprocessing, K-Means, Hierarchical, Evaluasi). Tombol "Jalankan Semua Analisis" memungkinkan seluruh pipeline dijalankan sekaligus.



Gambar 4.6 Tampilan Halaman Dashboard

4.11.2 Halaman Dataset IPM

Halaman Dataset IPM menampilkan seluruh data IPM 33 kabupaten/kota Sumatera Utara periode 2020-2025 yang sedang aktif digunakan sistem, dilengkapi fitur pencarian wilayah dan ekspor ulang data ke format CSV untuk keperluan verifikasi.

Sistem Clustering IPM Sumatera Utara
Evaluasi K-Means & Hierarchical Clustering

Unggah CSV untuk memulai

MENU UTAMA

Dashboard

Dataset IPM

Import Data

Preprocessing

CLUSTERING

K-Means

Hierarchical

HASIL

Hasil Clustering

Visualisasi

Evaluasi

Laporan

Dataset IPM

Data IPM 33 Kabupaten/Kota Sumatera Utara 2020-2025 (Sumber: BPS)

Cari kabupaten/kota...

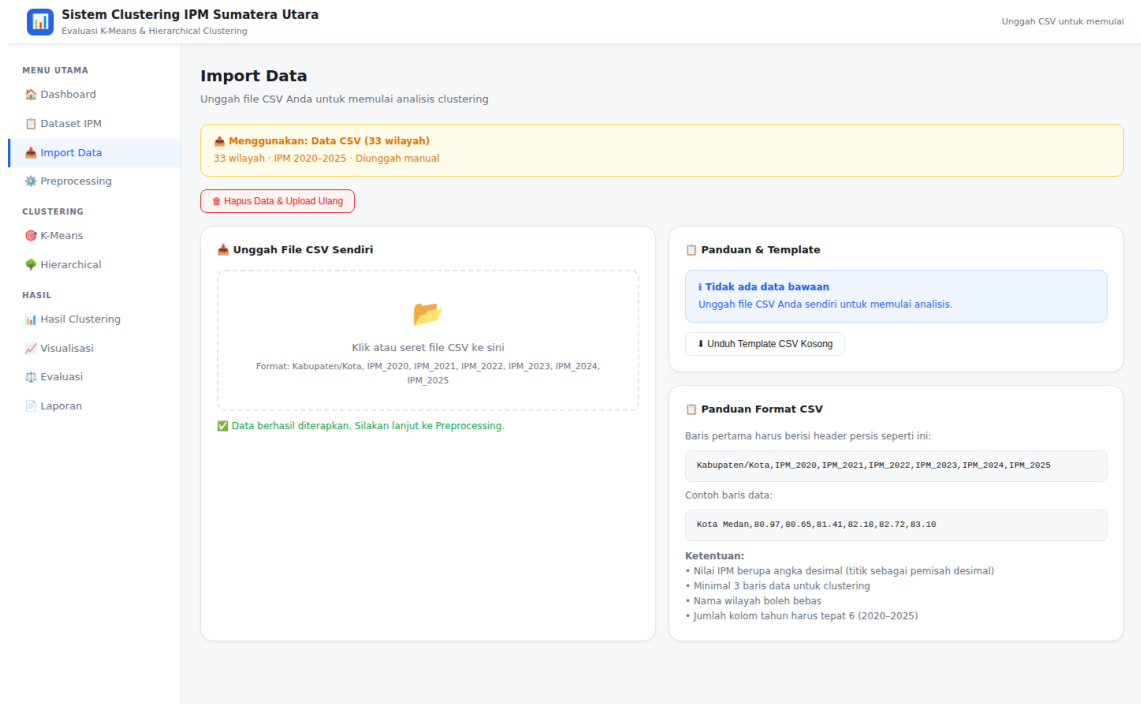
Eksport CSV

No	Kabupaten/Kota	IPM 2020	IPM 2021	IPM 2022	IPM 2023	IPM 2024	IPM 2025	Rata-rata	Tren
1	Nias	61.93	63.37	64.30	65.15	66.13	66.79	64.61	↗
2	Mandailing Natal	66.79	71.31	72.17	72.65	73.44	74.26	71.77	↗
3	Tapanuli Selatan	70.12	73.37	73.96	74.58	75.18	75.78	73.83	↗
4	Tapanuli Tengah	69.23	71.48	72.18	72.77	73.34	73.97	72.16	↗
5	Tapanuli Utara	73.47	76.06	76.43	76.86	77.48	78.09	76.40	↗
6	Toba	75.16	77.03	77.58	77.83	78.44	79.17	77.53	↗
7	Labuhan Batu	72.01	73.16	73.96	74.70	75.20	75.79	74.14	↗
8	Asahan	70.29	72.55	73.18	73.59	74.17	75.13	73.15	↗
9	Simalungun	73.25	74.33	74.68	75.17	75.88	76.48	74.97	↗
10	Dairi	71.57	73.89	74.60	75.18	75.88	76.68	74.63	↗
11	Karo	74.43	75.81	76.37	76.88	77.49	78.10	76.51	↗
12	Deli Serdang	75.44	76.21	76.86	77.16	77.56	78.29	76.92	↗
13	Langkat	71.00	73.50	74.03	74.33	74.85	75.58	73.88	↗
14	Nias Selatan	61.89	63.24	64.05	64.98	65.87	66.74	64.46	↗
15	Humbahas	68.87	71.28	72.05	72.49	73.33	73.93	71.99	↗
16	Pakpak Barat	67.59	70.65	71.57	72.30	73.09	73.89	71.52	↗
17	Samosir	70.63	71.75	72.57	72.93	73.77	74.63	72.71	↗
18	Serdang Bedagai	70.24	72.25	72.89	73.40	74.15	74.69	72.94	↗
19	Batu Bara	68.36	70.80	71.75	72.56	73.14	73.92	71.76	↗
20	Paluta	69.85	71.96	72.77	73.45	74.16	75.12	72.88	↗
21	Padang Lawas	68.25	70.48	71.41	72.16	73.10	73.90	71.55	↗
22	Labusel	71.40	73.15	73.59	74.23	75.11	75.74	73.87	↗
23	Labura	71.61	73.73	74.61	75.45	76.20	76.71	74.72	↗
24	Nias Utara	62.36	63.63	64.55	65.44	66.42	67.38	64.96	↗
25	Nias Barat	61.51	63.05	63.96	64.68	65.66	66.65	64.25	↗
26	Kota Sibolga	73.63	75.87	76.69	77.07	77.52	78.11	76.48	↗
27	Kota Tanjung Balai	68.65	73.78	74.73	75.42	75.95	76.70	74.20	↗
28	Kota P. Siantar	78.75	79.17	79.71	80.46	81.17	81.71	80.16	↗
29	Kota Tebing Tinggi	75.17	76.65	77.39	78.17	78.68	79.26	77.55	↗
30	Kota Medan	80.98	81.58	82.13	82.61	83.23	83.74	82.38	↗
31	Kota Binjai	75.89	76.57	77.50	78.11	78.67	79.20	77.66	↗
32	Kota P. Sidempuan	75.22	77.19	77.73	78.10	78.61	79.19	77.67	↗
33	Kota Gunung Sitoli	69.31	70.51	71.11	71.55	72.30	73.29	71.35	↗

Gambar 4.7 Tampilan Halaman Dataset IPM

4.11.3 Halaman Import Data

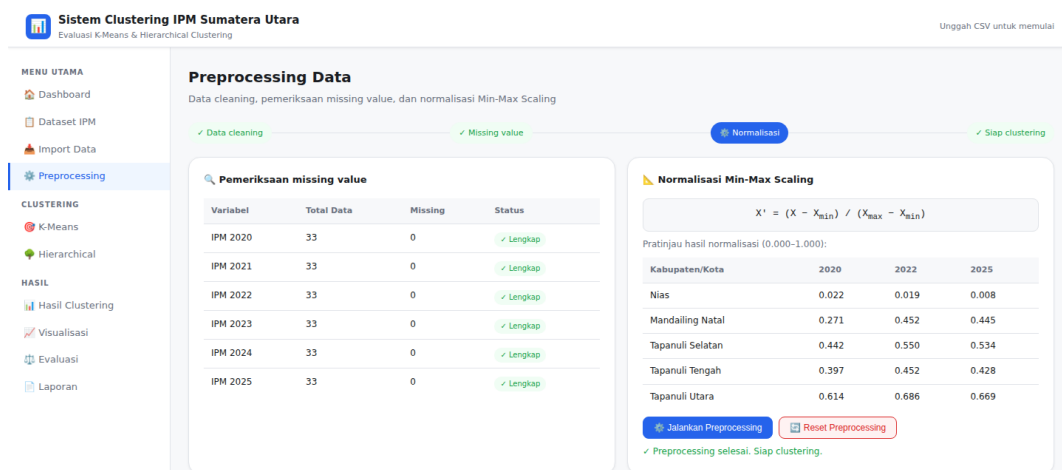
Halaman Import Data menyediakan fitur unggah data CSV melalui drag-and-drop maupun pemilihan file secara manual. Sistem melakukan validasi format kolom, menampilkan pratinjau data sebelum diterapkan, serta menyediakan template CSV kosong sebagai panduan format bagi pengguna.



Gambar 4.8 Tampilan Halaman Import Data

4.11.4 Halaman Preprocessing

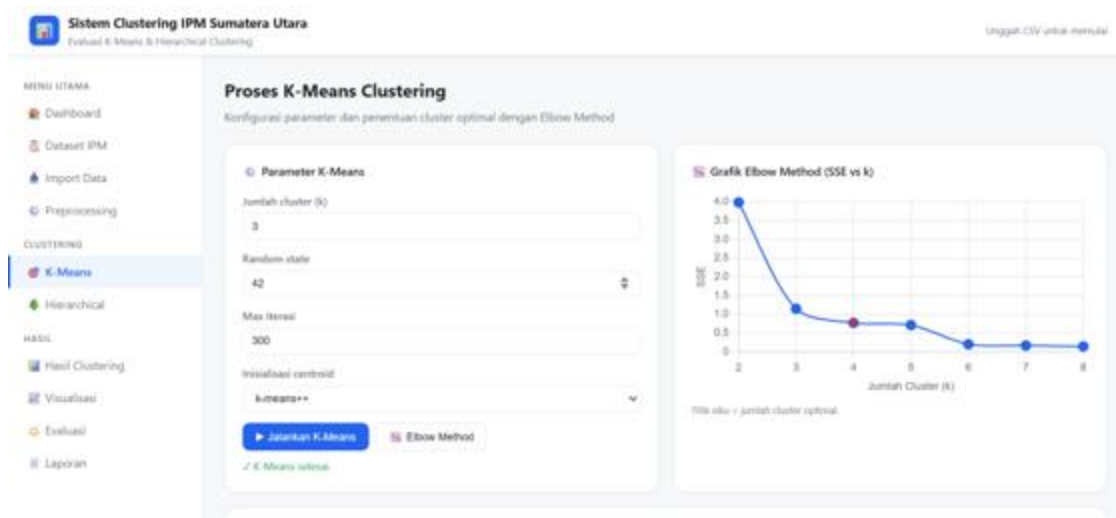
Halaman Preprocessing menampilkan hasil pemeriksaan missing value pada setiap variabel tahun serta pratinjau hasil normalisasi Min-Max Scaling. Rumus normalisasi ditampilkan langsung pada halaman ini sebagai referensi metodologis.



Gambar 4.9 Tampilan Halaman Preprocessing

4.11.5 Halaman K-Means

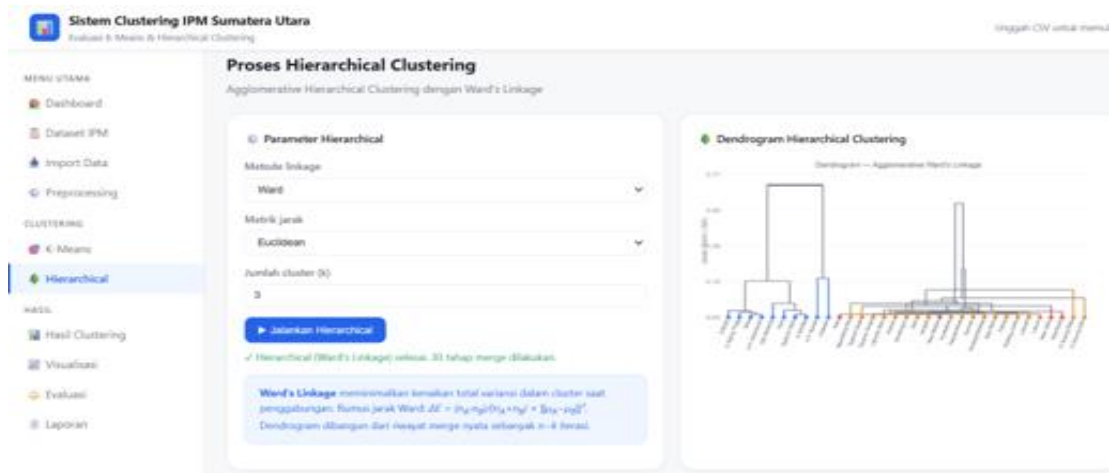
Halaman K-Means menyediakan panel konfigurasi parameter (jumlah cluster k, random state, maksimum iterasi, metode inisialisasi centroid), grafik Metode Elbow (ditampilkan sebelumnya pada Gambar 4.1), serta tabel hasil clustering yang mencakup Silhouette Score, SSE, jumlah iterasi konvergen, dan ringkasan tiap cluster.



Gambar 4.10 Tampilan Halaman K-Means

4.11.6 Halaman Hierarchical

Halaman Hierarchical menyediakan panel konfigurasi (metode linkage, metrik jarak, jumlah cluster), dendrogram hasil penggabungan cluster (ditampilkan sebelumnya pada Gambar 4.3), serta tabel hasil clustering berikut Silhouette Score dan Davies-Bouldin Index.



Gambar 4.11 Tampilan Halaman Hierarchical

4.11.7 Halaman Hasil Clustering

Halaman Hasil Clustering menyajikan label cluster untuk setiap kabupaten/kota berdasarkan algoritma yang dipilih (K-Means atau Hierarchical), dilengkapi filter kategori cluster dan fitur ekspor hasil ke format CSV.

Sistem Clustering IPM Sumatera Utara
Subsistem K-Means & Hierarchical Clustering

Hasil Clustering

Label cluster per kabupaten/kota berdasarkan nilai IPM 2020-2025

K-Means (k=2) Semua cluster Elaborasi CVR

No	Kabupaten/Kota	2020	2021	2022	2023	2024	2025	Batas-batas	Cluster	Kategori
1	Max	61.93	61.37	64.30	65.15	66.13	66.79	64.61	3	Kandah
2	Mandailing Natal	66.79	71.31	72.17	72.65	73.44	74.26	73.77	2	Sekelap
3	Tapanuli Selatan	70.12	73.37	73.96	74.58	75.18	75.78	73.83	2	Sekelap
4	Tapanuli Tengah	69.23	71.48	72.18	72.77	73.24	73.97	72.16	2	Sekelap
5	Tapanuli Utara	73.47	76.06	76.43	76.86	77.48	78.09	76.40	1	Tinggi
6	Toba	75.16	77.03	77.58	77.83	78.44	79.17	77.53	1	Tinggi
7	Labuhan Batu	72.01	73.18	73.96	74.30	75.20	75.79	74.14	2	Sekelap
8	Asahan	70.29	72.55	73.18	73.59	74.17	75.13	73.13	2	Sekelap
9	Simelungut	73.25	74.33	74.88	75.17	75.88	76.48	74.97	2	Sekelap
10	Guari	71.57	73.89	74.60	75.18	75.88	76.68	74.63	2	Sekelap
11	Karo	74.41	76.81	76.37	76.88	77.49	78.10	76.51	1	Tinggi
12	Deli Serdang	75.44	76.21	76.88	77.18	77.56	78.29	76.92	1	Tinggi
13	Langkat	71.00	73.30	74.03	74.33	74.85	75.38	73.88	2	Sekelap
14	Nias Selatan	62.89	63.28	64.05	64.98	65.87	66.74	64.66	3	Kandah
15	Humbahas	68.87	71.28	72.05	72.49	73.33	73.93	71.99	2	Sekelap
16	Pekpak Barat	67.59	70.65	71.57	72.30	73.09	73.89	71.52	2	Sekelap
17	Samosir	70.63	71.75	72.57	72.93	73.77	74.63	72.71	2	Sekelap
18	Serdang Bedagai	70.24	72.25	72.89	73.40	74.15	74.66	72.94	2	Sekelap
19	Batu Bara	68.36	70.80	71.75	72.56	73.14	73.92	71.76	2	Sekelap
20	Peluta	69.85	71.96	72.77	73.45	74.16	75.13	72.88	2	Sekelap
21	Pedang Lawas	68.25	70.48	71.41	72.16	73.10	73.90	71.55	2	Sekelap
22	Labusel	71.40	73.15	73.59	74.23	75.11	75.74	73.87	2	Sekelap
23	Labura	71.61	73.73	74.61	75.45	76.20	76.71	74.72	2	Sekelap
24	Nias Utara	62.36	63.63	64.55	65.44	66.42	67.38	64.96	3	Kandah
25	Nias Barat	61.51	63.05	63.96	64.68	65.66	66.65	64.25	3	Kandah
26	Kota Sibolga	73.63	75.87	76.69	77.87	77.52	78.11	76.48	1	Tinggi
27	Kota Tanjung Balai	68.65	71.78	74.73	75.42	75.95	76.70	74.20	2	Sekelap
28	Kota P. Siantar	78.75	79.17	79.71	80.46	81.27	81.71	80.16	1	Tinggi
29	Kota Teling Tinggi	75.17	76.65	77.39	78.17	78.68	79.26	77.55	1	Tinggi
30	Kota Medan	80.98	81.58	82.15	82.61	83.23	83.74	82.38	1	Tinggi
31	Kota Binjai	75.89	76.57	77.50	78.11	78.67	79.20	77.66	1	Tinggi
32	Kota P. Sidempuan	75.22	77.19	77.73	78.10	78.63	79.19	77.67	1	Tinggi
33	Kota Gunung Siboli	69.31	70.51	71.11	71.55	72.30	73.29	71.35	2	Sekelap

Gambar 4.12 Tampilan Halaman Hasil Clustering

4.11.8 Halaman Visualisasi

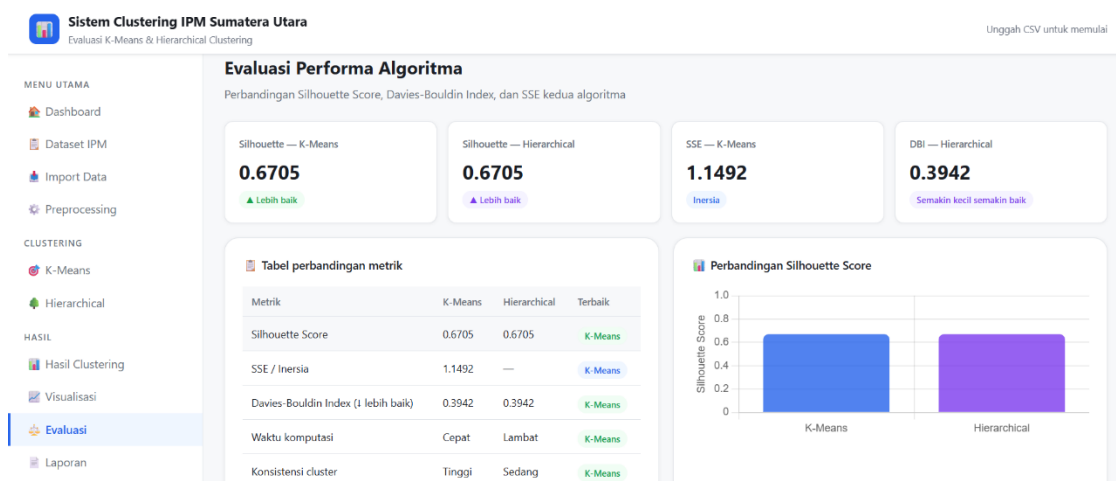
Halaman Visualisasi menampilkan scatter plot IPM 2020 terhadap IPM 2025 untuk hasil K-Means dan Hierarchical secara berdampingan, serta grafik batang perbandingan jumlah anggota tiap cluster antara kedua algoritma. Tampilan scatter plot telah disajikan sebelumnya pada Gambar 4.2. Visualisasi ini penting karena memungkinkan pengguna memahami secara intuitif seberapa jelas pemisahan antarcluster terbentuk, tanpa harus membaca dan membandingkan angka numerik satu per satu.



Gambar 4.13 Tampilan Halaman Visualisasi

4.11.9 Halaman Evaluasi

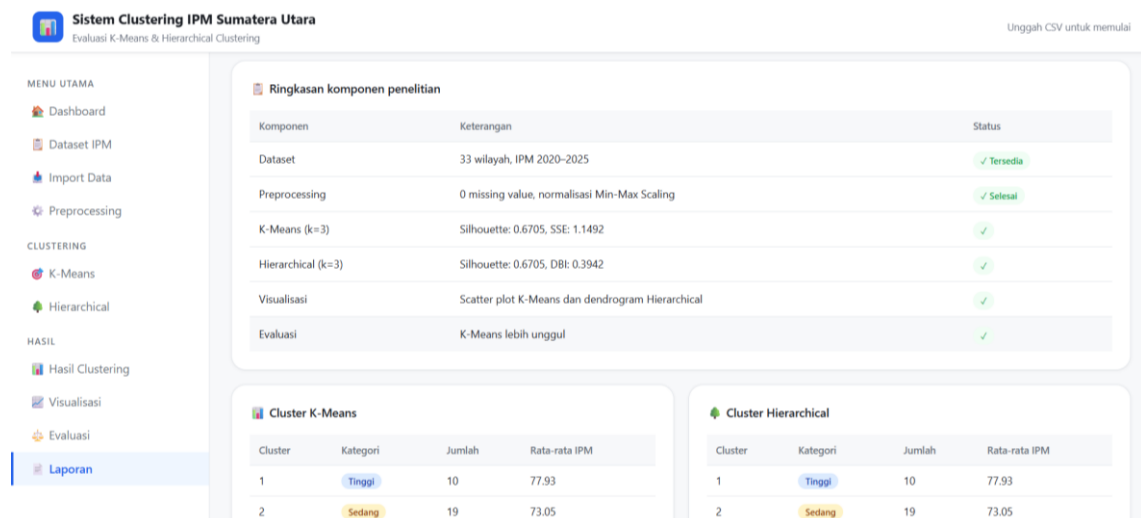
Halaman Evaluasi menyajikan perbandingan Silhouette Score, SSE, dan Davies-Bouldin Index kedua algoritma beserta kesimpulan evaluasi otomatis. Tampilan halaman ini telah disajikan pada Gambar 4.4 (subbab 4.7). Halaman ini penting karena memungkinkan pengguna membandingkan performa kedua algoritma secara langsung dan objektif berdasarkan metrik kuantitatif, sehingga pemilihan algoritma terbaik tidak dilakukan secara subjektif.



Gambar 4.14 Tampilan Halaman Evaluasi

4.11.10 Halaman Laporan

Halaman Laporan merangkum seluruh komponen penelitian (dataset, preprocessing, K-Means, Hierarchical, evaluasi) dalam satu tampilan, dilengkapi tombol cetak laporan dan ekspor CSV. Tampilan halaman ini telah disajikan pada Gambar 4.5 (subbab 4.10). Halaman ini dirancang sebagai ringkasan akhir yang memudahkan pengguna, misalnya pihak pengambil kebijakan, untuk melihat keseluruhan hasil penelitian dalam satu tampilan tanpa perlu membuka setiap halaman satu per satu, sekaligus mendukung dokumentasi hasil analisis melalui fitur cetak dan ekspor CSV.



Gambar 4.15 Tampilan Halaman Laporan

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai evaluasi performa algoritma K-Means dan Hierarchical Clustering dalam mengelompokkan data Indeks Pembangunan Manusia (IPM) 33 kabupaten/kota Provinsi Sumatera Utara periode 2020–2025, dapat ditarik kesimpulan sebagai berikut.

1. Pola pertumbuhan IPM pada masing-masing kabupaten/kota di Provinsi Sumatera Utara periode 2020–2025 menunjukkan tren yang terus meningkat setiap tahunnya, dengan rata-rata IPM provinsi naik dari 71,77 pada tahun 2020 menjadi 76,47 pada tahun 2025. Meskipun demikian, disparitas antarwilayah masih cukup besar, ditunjukkan oleh kesenjangan sekitar 17 poin antara Kota Medan sebagai wilayah dengan IPM tertinggi (83,74) dan Nias Barat sebagai wilayah dengan IPM terendah (66,65) pada tahun 2025.
2. Hasil pengelompokan kabupaten/kota menggunakan algoritma K-Means terhadap data pertumbuhan IPM menghasilkan 3 cluster optimal, yaitu Cluster Tinggi beranggotakan 10 wilayah (didominasi kota besar), Cluster Sedang beranggotakan 19 wilayah (kabupaten daratan Sumatera Utara), dan Cluster Rendah beranggotakan 4 wilayah (seluruh kabupaten di Kepulauan Nias), dengan nilai Silhouette Score sebesar 0,6705 dan SSE sebesar 1,1492 yang menunjukkan kualitas pengelompokan yang baik.
3. Hasil pengelompokan kabupaten/kota menggunakan algoritma Hierarchical Clustering (Ward Linkage) menghasilkan komposisi anggota cluster yang

identik dengan hasil K-Means, dengan nilai Silhouette Score sebesar 0,6705 dan Davies-Bouldin Index sebesar 0,3942, yang mengindikasikan struktur pengelompokan yang konsisten dan tidak bergantung pada pendekatan algoritma yang digunakan.

4. Berdasarkan hasil evaluasi, kedua algoritma menghasilkan kualitas clustering yang setara, sehingga tidak terdapat perbedaan signifikan dari segi akurasi pengelompokan. Namun demikian, algoritma K-Means direkomendasikan sebagai metode yang lebih optimal untuk diterapkan pada kasus ini karena memiliki kompleksitas komputasi yang lebih rendah ($O(n)$ per iterasi dibandingkan $O(n^2 \log n)$ pada Hierarchical Clustering), sehingga lebih efisien dan lebih skalabel apabila diterapkan pada data berukuran lebih besar di masa mendatang.

5.2 Saran

Berdasarkan kesimpulan penelitian, beberapa saran yang dapat disampaikan adalah sebagai berikut

1. Bagi pemerintah daerah, hasil pengelompokan ini dapat dijadikan salah satu dasar pertimbangan dalam menyusun kebijakan pembangunan yang lebih terarah sesuai karakteristik masing-masing cluster, terutama bagi wilayah Kepulauan Nias yang termasuk dalam Cluster Rendah agar mendapatkan perhatian dan intervensi pembangunan yang lebih intensif.
2. Bagi penelitian selanjutnya, disarankan untuk menambahkan indikator penyusun IPM secara terpisah (Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran per Kapita) sebagai

atribut clustering, sehingga dapat diketahui dimensi pembangunan manusia spesifik yang paling memengaruhi pengelompokan suatu wilayah.

3. Penelitian selanjutnya juga disarankan untuk memperluas cakupan wilayah penelitian ke provinsi lain atau skala nasional, serta menambah rentang periode data yang digunakan, agar hasil clustering dapat digeneralisasi lebih luas dan mencerminkan pola jangka panjang.
4. Bagi pengembangan sistem, disarankan untuk menambahkan penyimpanan data secara persisten (basis data) dan fitur unggah data berkelanjutan, sehingga sistem tidak hanya bergantung pada input CSV manual dan dapat digunakan secara berkelanjutan oleh instansi terkait.

DAFTAR PUSTAKA

- Aldahdooh, R. T., & Ashour, W. M. (2013). DIMK-means “Distance-based Initialization Method for K-means Clustering Algorithm.” *International Journal of Intelligent Systems and Applications*, 5(2), 41–51. <https://doi.org/10.5815/IJISA.2013.02.05>
- Agha, M. E., & Ashour, W. M. (2012). Efficient and Fast Initialization Algorithm for K- means Clustering. *International Journal of Intelligent Systems and Applications*, 4(1), 21–31. <https://doi.org/10.5815/IJISA.2012.01.03>
- Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The Advantages of Careful Seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07)*, 1027–1035.
- Bagirov, A. M., Aliguliyev, R. M., & Sultanova, N. (2023). Finding compact and well-separated clusters: Clustering using silhouette coefficients. *Pattern Recognition*, 135, 109144–109144. <https://doi.org/10.1016/j.patcog.2022.109144>
- Bartak, J., & Jastrzebska, A. (2022). Mining patterns of transitional growth using multivariate concept-based models. *Quality & Quantity*, 56(6), 4395–4419. <https://doi.org/10.1007/s11135-022-01318-8>
- Beecks, C., Berns, F., Hüwel, J. D., Linxen, A., Schlake, G. S., & Düsterhus, T. (2022). A Comparative Performance Analysis of Fast K-Means Clustering Algorithms. *Lecture Notes in Computer Science*, 119–125. https://doi.org/10.1007/978-3-031-21047-1_11
- Boyko, N., & Tkachyk, O. (2023). Hierarchical clustering algorithm for dendrogram construction and cluster counting. *Īnformatika Ta Matematičnì Metodi v Modelŭvannì*, 13(1–2), 5–15. <https://doi.org/10.15276/imms.v13.no1-2.5>
- Budiarti, W., Standsyah, R. E., Syahru, D., Ramadhan, Dr, U., & Soetomo, S. (2024). *Efforts to Alleviate Underdeveloped Areas by Clustering Regional Characteristics in Indonesia*. <https://doi.org/10.55927/ministal.v2i4.5531>

- Cai, R., Zhang, Z., Tung, A. K. H., Dai, C., & Hao, Z. (2014). A general framework of hierarchical clustering and its applications. *Information Sciences*, 272, 29–48. <https://doi.org/10.1016/J.INS.2014.02.062>
- Dolnicar, S., & Grün, B. (2022). *Cluster Analysis*. 509–511. <https://doi.org/10.4337/9781800377486.cluster.analysis>
- Fahmiyah, I., & Ningrum, R. A. (2023). *Human Development Clustering in Indonesia: Using K-Means Method and Based on Human Development Index Categories*. 2(1), 27–33. <https://doi.org/10.20473/jatm.v2i1.45070>
- Fadillah, M. R., Ardiansyah, M. R., & Fauzi, F. (2025). Penerapan K-Means Clustering dalam Mengelompokkan Kabupaten/Kota Berdasarkan Indikator IPM di Jawa Tengah. *Emerging Statistics and Data Science Journal*, 3(3), 674–685. <https://doi.org/10.20885/esds.vol3.iss.3.art20>
- Lawelai, H., & Baulete, E. (2025). *Human Development Inequality Inter-regional in Southeast Sulawesi Province 2019–2023*. <https://doi.org/10.36685/jlssm.v2i2.1553>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Liu, H. M., & Lu, J. G. (2015). Brief Survey of K-Means Clustering Algorithms. *Applied Mechanics and Materials*, 740, 624–628. <https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMM.740.624>
- Martín-Fernández, J. D., Luna-Romera, J. M., Pontes, B., & Riquelme-Santos, J. C. (2020). *Indexes to Find the Optimal Number of Clusters in a Hierarchical Clustering*. 3–13. https://doi.org/10.1007/978-3-030-20055-8_1
- M, S., V.J, R., & Abdoulelbaistoisaidabdou. (2018). Overview of Data Mining. *International Journal of Advanced Research in Computer and Communication Engineering*, 7(11), 109–113. <https://doi.org/10.17148/IJARCCE.2018.71123>
- Nurdin, N. H., Mahmud, M. N., Mirsa, S., & Hamson, Z. (2024). *Kinerja kebijakan pembangunan manusia di kabupaten bone*. 2(01), 01–13. <https://doi.org/10.62668/phenomenon.v2i01.960>

- Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). *Comprehensive Review of K-Means Clustering Algorithms*. 7(8), 64–69. <https://doi.org/10.31695/IJASRE.2021.34050>
- Permanyar, I., & Suppa, N. (2022). Racing ahead or lagging behind? Territorial cohesion in human development around the globe. *Regional Studies*, 56(12), 2086–2101. <https://doi.org/10.1080/00343404.2022.2038785>
- Rahmati, R., & Wijayanto, A. W. (2021). Analisis Cluster dengan Algoritma K-Means, Fuzzy C-Means dan Hierarchical Clustering (Studi Kasus: Indeks Pembangunan Manusia Tahun 2019). *JIKO (Jurnal Informatika dan Komputer)*, 5(2), 73–80. <https://doi.org/10.26798/jiko.v5i2.422>
- Ramadhan, R., & Mutaqin, A. K. (2025). Deskripsi Data pada Aspek Pembangunan Manusia di Provinsi Jawa Barat Tahun 2021-2023. *Bandung Conference Series Statistics*, 5(2). <https://doi.org/10.29313/bcss.v5i2.20002>
- Saraswat, P., & Raj, S. (2022). Data pre-processing techniques in data mining: a review. *International Journal of Innovative Research in Computer Science & Technology*, 122–125. <https://doi.org/10.55524/ijircst.2022.10.1.22>
- Sikana, A. M., & Wijayanto, A. W. (2021). Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering. *Jurnal Ilmu Komputer*, 14(2), 66–78. <https://doi.org/10.24843/JIK.2021.v14.i02.p01>
- Thant, A. A., & Aye, S. M. (2020). Euclidean, Manhattan and Minkowski Distance Methods For Clustering Algorithms. *International Journal of Scientific Research in Science, Engineering and Technology*, 7(3), 553–559. <https://doi.org/10.32628/IJSRSET2073118>
- Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>