

**PEMBOBOTAN FITUR PADA *HIERARCHICAL CLUSTERING*  
UNTUK SEGMENTASI PRODUKSI PADI DI SUMATERA  
UTARA**

**SKRIPSI**

**OLEH:**

**NORA IRAWANI SIREGAR**

**228160013**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
MEDAN  
2026**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 1/7/26

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Access From (repository.uma.ac.id)1/7/26

**PEMBOBOTAN FITUR PADA *HIERARCHICAL CLUSTERING*  
UNTUK SEGMENTASI PRODUKSI PADI DI SUMATERA  
UTARA**

**SKRIPSI**

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh  
Gelar Sarjana di Fakultas Teknik  
Universitas Medan Area**

**OLEH:**

**NORA IRAWANI SIREGAR**

**228160013**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
MEDAN  
2026**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

## HALAMAN PENGESAHAN

**Judul Skripsi : PEMBOBOTAN FITUR PADA *HIERARCHICAL CLUSTERING* UNTUK SEGMENTASI PRODUKSI PADI DI SUMATERA UTARA**

**Nama : NORA IRAWANI SIREGAR**

**NPM : 228160013**

**Fakultas : TEKNIK**

Disetujui Oleh  
Komisi Pembimbing

Nanda Novita, S.Kom., M.Kom  
Pembimbing I



....., S.T., M.T  
Fakultas Teknik



....., S.Kom., M.Kom  
Prodi Teknik Informatika

Tanggal Lulus : 12 Maret 2026

## **HALAMAN PERNYATAAN**

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 12 Maret 2026



Nora Irawani Siregar  
228160013



**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI  
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Nora Irawani Siregar  
NPM : 228160013  
Program Studi : Teknik Informatika  
Fakultas : Teknik  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul:

**PEMBOBOTAN FITUR PADA *HIERARCHICAL CLUSTERING* UNTUK  
SEGMENTASI PRODUKSI PADI DI SUMATERA UTARA**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan  
Pada tanggal : 12 Maret 2026  
Yang menyatakan

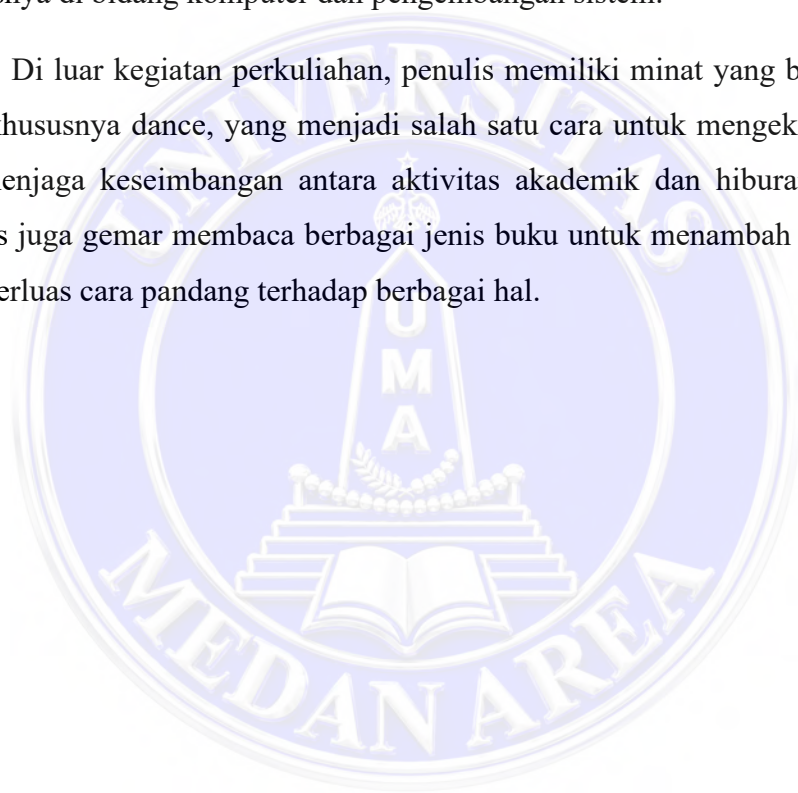


Nora Irawani Siregar  
228160013

## RIWAYAT HIDUP

Nora Irawani Siregar lahir di Padang Sidempuan pada tanggal 10 April 2004. Ia merupakan putri pertama dari tiga bersaudara yang tumbuh dan besar di Kota Sibolga. Pendidikan formal dimulai dari TK Bina Ikhwan, dilanjutkan ke SD Pandan 3, kemudian SMP Negeri 2 Pandan Nauli, dan menyelesaikan pendidikan menengah atas di SMAN 6 Padang Sidempuan pada tahun 2022. Setelah itu, penulis melanjutkan pendidikan ke Universitas Medan Area dengan mengambil program studi Teknik Informatika, yang didasari oleh ketertarikan terhadap dunia teknologi, khususnya di bidang komputer dan pengembangan sistem.

Di luar kegiatan perkuliahan, penulis memiliki minat yang besar terhadap seni, khususnya dance, yang menjadi salah satu cara untuk mengekspresikan diri dan menjaga keseimbangan antara aktivitas akademik dan hiburan. Selain itu, penulis juga gemar membaca berbagai jenis buku untuk menambah wawasan dan memperluas cara pandang terhadap berbagai hal.



## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, taufik, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Pembobotan Fitur pada *Hierarchical Clustering* untuk Segmentasi Produksi Padi di Sumatera Utara." Penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak yang telah memberikan motivasi serta arahan selama proses penelitian hingga penyusunan laporan ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Dengan penuh rasa syukur kepada Allah SWT. atas segala petunjuk, pertolongan, dan perlindungan-Nya selama penyusunan karya ini berlangsung.
2. Kedua orang tua tercinta, atas kasih sayang, doa, serta dukungan moral dan material yang tiada henti diberikan kepada penulis.
3. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
4. Bapak Dr. Eng. Suprianto, ST., MT., selaku Dekan Fakultas Teknik Universitas Medan Area.
5. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika.
6. Ibu Nanda Novita, S.Kom., M.Kom., selaku Dosen Pembimbing yang telah dengan sabar memberikan bimbingan, arahan, dan motivasi selama proses penyusunan skripsi ini.



7. Kepada adik-adik penulis yang terkasih Rayhand Ikhwal Siregar dan Tika Permata Sari Siregar yang sudah membuat penulis semangat untuk menyelesaikan skripsi ini.
8. Kepada kedua sahabat penulis dari semester satu yaitu Stevi Freshia Sihombing dan Cessy Cindi Regina Situmorang atas dukungan dan motivasi yang sangat berharga.
9. Kepada seluruh teman-teman yang tidak bisa penulis tulis satu-satu.
10. Kepada diri sendiri, penulis mengucapkan terimakasih dan syukur karena telah semangat sampai ditahap ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan baik dari segi penulisan maupun isi, yang disebabkan oleh keterbatasan pengetahuan dan pengalaman penulis. Oleh karena itu, penulis dengan rendah hati mengharapkan kritik dan saran yang bersifat membangun demi penyempurnaan karya ilmiah ini.

Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi bahan referensi dalam penelitian selanjutnya yang berkaitan dengan analisis menggunakan metode *Hierarchical Clustering*.

Medan  
Penulis



Nora Irawani Siregar  
228160013



## ABSTRAK

Pengelompokan wilayah produksi padi di Provinsi Sumatera Utara merupakan langkah penting dalam mendukung perencanaan dan pengambilan kebijakan pertanian yang tepat sasaran karena mampu memberikan gambaran yang lebih jelas mengenai karakteristik dan potensi produksi di setiap daerah. Sehingga penelitian ini bertujuan untuk melakukan segmentasi wilayah produksi padi di Provinsi Sumatera Utara dengan menerapkan pembobotan fitur menggunakan metode *Analytical Hierarchy Process* (AHP) yang diintegrasikan ke dalam algoritma *Hierarchical Clustering*. Data yang digunakan diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Metode penelitian meliputi pengumpulan data, pra-pemrosesan data, penentuan bobot fitur menggunakan AHP, normalisasi dan pembobotan data, proses klusterisasi, serta evaluasi menggunakan *Silhouette Index* (SI) dan *Davies–Bouldin Index* (DBI). Hasil penelitian menunjukkan bahwa pada fokus produksi padi diperoleh konfigurasi dua klaster dengan nilai SI 0,7182, DBI 0,2266, dan CHI 39,2039 sedangkan pada fokus produktivitas padi diperoleh empat klaster dengan nilai SI 0,5480, DBI 0,5209, dan CHI 72,3464 yang mampu menggambarkan variasi tingkat efisiensi antar wilayah. Dengan demikian, hasil segmentasi wilayah yang diperoleh dapat memberikan gambaran pengelompokan wilayah produksi padi di Provinsi Sumatera Utara.

**Kata Kunci:** Pembobotan Fitur, *Hierarchical Clustering*, *Analytical Hierarchy Process*, Produksi Padi, Produktivitas Padi, Segmentasi Wilayah.

## ABSTRACT

*The clustering of rice production regions in North Sumatra Province is an important step in supporting more targeted agricultural planning and policy-making, as it provides a clearer overview of the characteristics and production potential of each region. Therefore, this study aims to segment rice production regions in North Sumatra Province by applying feature weighting using the Analytical Hierarchy Process (AHP) integrated into the Hierarchical Clustering algorithm. The data used in this study were obtained from the Central Statistics Agency (BPS) of North Sumatra Province. The research methodology includes data collection, data preprocessing, feature weighting determination using AHP, data normalization and weighting, clustering processes, and evaluation using the Silhouette Index (SI) and Davies–Bouldin Index (DBI). The results show that, based on the rice production focus, a configuration of two clusters was obtained with an SI value of 0.7182 , a DBI value of 0.2266 and CHI value of 39,2039. Meanwhile, for the rice productivity focus, four clusters were obtained with an SI value of 0.5480 , a DBI value of 0.5209 and a CHI value of 72, 3464, which effectively describe variations in efficiency levels among regions. Thus, the resulting regional segmentation provides a clear overview of the grouping patterns of rice production areas in North Sumatra Province.*

**Keywords:** *Feature Weighting, Hierarchical Clustering, Analytical Hierarchy Process, Rice Production, Rice Productivity, Regional Segmentation.*

## DAFTAR ISI

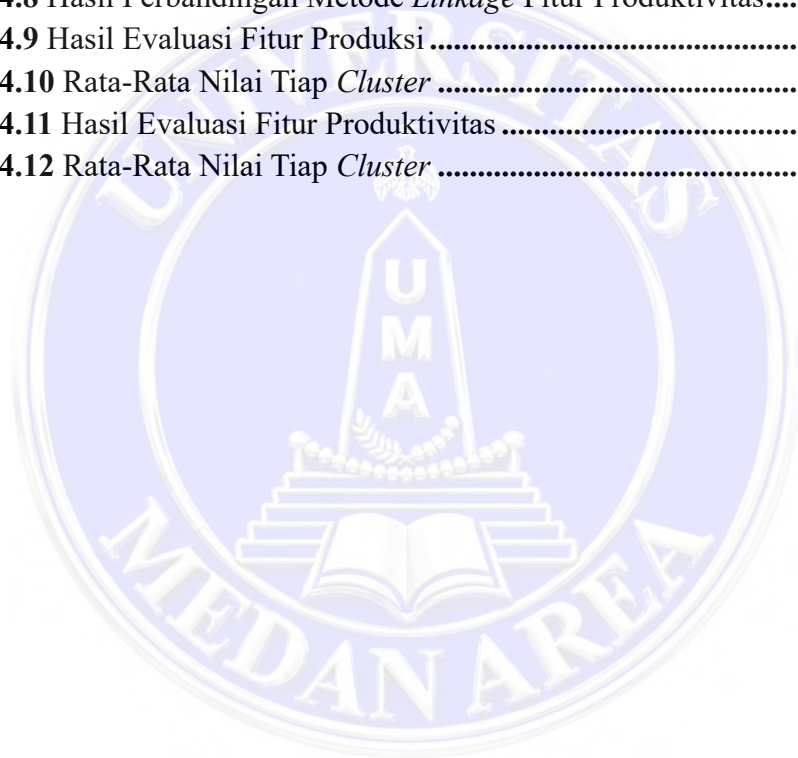
|                                                      |               |
|------------------------------------------------------|---------------|
| <b>HALAMAN PENGESAHAN</b> .....                      | <b>ii</b>     |
| <b>HALAMAN PERNYATAAN</b> .....                      | <b>iii</b>    |
| <b>RIWAYAT HIDUP</b> .....                           | <b>v</b>      |
| <b>KATA PENGANTAR</b> .....                          | <b>vi</b>     |
| <b>ABSTRAK</b> .....                                 | <b>viii</b>   |
| <b>ABSTRACT</b> .....                                | <b>ix</b>     |
| <b>DAFTAR ISI</b> .....                              | <b>x</b>      |
| <b>DAFTAR TABEL</b> .....                            | <b>xii</b>    |
| <b>DAFTAR GAMBAR</b> .....                           | <b>xiii</b>   |
| <br><b>BAB I PENDAHULUAN</b> .....                   | <br><b>1</b>  |
| 1.1 Latar Belakang Masalah.....                      | 1             |
| 1.2 Rumusan Masalah .....                            | 3             |
| 1.3 Batasan Masalah .....                            | 3             |
| 1.4 Tujuan Penelitian.....                           | 4             |
| 1.5 Manfaat Penelitian .....                         | 5             |
| <br><b>BAB II TINJAUAN PUSTAKA</b> .....             | <br><b>6</b>  |
| 2.1 Padi.....                                        | 6             |
| 2.2 Produksi dan Produktivitas Padi .....            | 7             |
| 2.3 <i>Analytic Hierarchy Process (AHP)</i> .....    | 8             |
| 2.4 <i>Machine Learning</i> .....                    | 9             |
| 2.5 <i>Hierarchical Clustering</i> .....             | 9             |
| 2.5.1 <i>Pseudocode Hierarcical Clustering</i> ..... | 12            |
| <i>Pseudocode Hierarcical Clustering</i> .....       | 12            |
| 2.5.2 <i>Flowchart Hierarchical Clustering</i> ..... | 13            |
| 2.6 <i>Silhouette Index (SI)</i> .....               | 13            |
| 2.7 <i>Davies-Bouldin Index (DBI)</i> .....          | 14            |
| 2.8 <i>Calinski Harabaz Index (CHI)</i> .....        | 14            |
| 2.9 Penelitian Terdahulu.....                        | 15            |
| <br><b>BAB III METODOLOGI PENELITIAN</b> .....       | <br><b>18</b> |
| 3.1 Sumber Data Penelitian.....                      | 18            |
| 3.2 Perangkat Yang Dibutuhkan.....                   | 18            |
| 3.3 Tahapan Penelitian .....                         | 19            |

|               |                                                         |           |
|---------------|---------------------------------------------------------|-----------|
| 3.3.1         | Pengumpulan Data .....                                  | 19        |
| 3.3.2         | Menentukan Bobot Fitur .....                            | 20        |
| 3.3.3         | Pra-Pemrosesan Data .....                               | 20        |
| 3.3.4         | <i>Exploratory Data Analysis</i> .....                  | 21        |
| 3.3.5         | Implementasi .....                                      | 21        |
| 3.3.6         | Evaluasi .....                                          | 24        |
| 3.4           | Perhitungan Manual .....                                | 25        |
| 3.4.1         | <i>Analytical Hierarchy Process (AHP)</i> .....         | 25        |
| 3.4.2         | Perhitungan Manual <i>Hierarchical Clustering</i> ..... | 27        |
| <b>BAB IV</b> | <b>HASIL DAN PEMBAHASAN .....</b>                       | <b>30</b> |
| 4.1           | Pra-Pemrosesan .....                                    | 30        |
| 4.2           | <i>Exploratory Data Analysis</i> .....                  | 31        |
| 4.3           | Pembobotan Fitur .....                                  | 34        |
| 4.4           | Normalisasi Data .....                                  | 35        |
| 4.5           | Implementasi .....                                      | 36        |
| 4.6           | Evaluasi .....                                          | 38        |
| 4.6.1         | Fitur Produksi .....                                    | 41        |
| 4.6.2         | Fitur Produktivitas .....                               | 44        |
| 4.7           | Pembahasan .....                                        | 48        |
| <b>BAB V</b>  | <b>KESIMPULAN .....</b>                                 | <b>50</b> |
| 5.1           | Kesimpulan .....                                        | 50        |
| 5.2           | Saran .....                                             | 50        |
|               | <b>DAFTAR PUSTAKA .....</b>                             | <b>52</b> |
|               | <b>LAMPIRAN .....</b>                                   | <b>56</b> |



## DAFTAR TABEL

|                                                                                    |           |
|------------------------------------------------------------------------------------|-----------|
| <b>Tabel 2.1</b> Penelitian Terdahulu .....                                        | <b>15</b> |
| <b>Tabel 3.1</b> Data Asli Penelitian .....                                        | <b>18</b> |
| <b>Tabel 4.1</b> Data Penelitian Setelah Agregasi.....                             | <b>30</b> |
| <b>Tabel 4.2</b> Statistik Deskriptif.....                                         | <b>31</b> |
| <b>Tabel 4.2</b> Hasil Statistik Deskriptif.....                                   | <b>31</b> |
| <b>Tabel 4.3</b> Hasil Pembobotan Fitur Fokus Produksi .....                       | <b>34</b> |
| <b>Tabel 4.4</b> Hasil Pembobotan Fitur Fokus Produktivitas .....                  | <b>35</b> |
| <b>Tabel 4.5</b> Normalisasi Data .....                                            | <b>35</b> |
| <b>Tabel 4.6</b> Normalisasi Data Dengan Bobot .....                               | <b>36</b> |
| <b>Tabel 4.7</b> Hasil Perbandingan Metode <i>Linkage</i> Fitur Produksi.....      | <b>38</b> |
| <b>Tabel 4.8</b> Hasil Perbandingan Metode <i>Linkage</i> Fitur Produktivitas..... | <b>39</b> |
| <b>Tabel 4.9</b> Hasil Evaluasi Fitur Produksi .....                               | <b>41</b> |
| <b>Tabel 4.10</b> Rata-Rata Nilai Tiap <i>Cluster</i> .....                        | <b>41</b> |
| <b>Tabel 4.11</b> Hasil Evaluasi Fitur Produktivitas .....                         | <b>44</b> |
| <b>Tabel 4.12</b> Rata-Rata Nilai Tiap <i>Cluster</i> .....                        | <b>45</b> |



## DAFTAR GAMBAR

|                                                                          |    |
|--------------------------------------------------------------------------|----|
| <b>Gambar 2.1</b> <i>Flowchart</i> Dasar AHC .....                       | 13 |
| <b>Gambar 3.1</b> <i>Diagram</i> Tahapan Penelitian .....                | 19 |
| <b>Gambar 3.2</b> <i>Flowchart Hierarchical Clustering</i> dan AHP ..... | 23 |
| <b>Gambar 4.1</b> Korelasi Produksi dan Produktivitas .....              | 32 |
| <b>Gambar 4.2</b> Hubungan Produksi dan Produktivitas .....              | 33 |
| <b>Gambar 4.3</b> Dendrogram Produksi .....                              | 36 |
| <b>Gambar 4.4</b> Dendrogram Produktivitas .....                         | 37 |
| <b>Gambar 4.5</b> Visualisasi Peta Klaster Produksi .....                | 43 |
| <b>Gambar 4.6</b> Visualisasi Peta Klaster Produktivitas .....           | 47 |



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Indonesia merupakan negara yang terkenal dengan pertaniannya, yang dapat dilihat dari mayoritas penduduknya menggantungkan hidup pada pertanian. Optimalisasi sektor pertanian menjadi kunci dalam pemanfaatan sumber daya alam dan geografi Indonesia, serta berperan penting dalam penyediaan pangan, bahan baku industri, dan ekspor (Nadilla et al., 2024). Tanaman padi adalah tanaman pangan strategis dan utama bagi ketahanan pangan di Indonesia (Shalaisya Azzahra Al-Meyda Hermawan et al., 2025). Upaya dalam memantau serta mengevaluasi kinerja produksi padi di berbagai wilayah menjadi suatu permasalahan tersendiri bagi pemerintah untuk menemukan kebijakan yang efisien dalam pengelolaannya. Provinsi Sumatera Utara memiliki potensi lahan yang luas dan beragam, yang jika dikelola dengan baik dapat meningkatkan hasil produksi padi secara signifikan. Namun, penilaian kinerja sering kali hanya bertumpu pada indikator tunggal seperti total volume produksi saja. Ketergantungan pada indikator tunggal ini dapat berpotensi mengaburkan gambaran kinerja yang lebih efisien, seperti produktivitas atau efisiensi lahan sering tidak mendapat posisi yang seimbang. Akibatnya, pemetaan potensi menjadi kurang merata (Nisa et al., 2025).

Untuk mengatasi permasalahan dalam pengelompokan yang kompleks ini, ilmu data (*data science*) seperti penggunaan *machine learning* dapat menjadi solusi yang efektif. Salah satu metode yang dipakai yaitu klaster (*clustering*), yang bertujuan untuk mengelompokkan objek data ke dalam beberapa klaster berdasarkan tingkat kemiripan karakteristik. Dalam berbagai penelitian untuk mengelompokkan

wilayah berdasarkan potensi pertanian dan indikator kesejahteraan Indonesia penerapan analisis klaster terbukti valid seperti penelitian dari (Septianto et al., 2025). Penggunaan metode klaster ini memungkinkan analisis multidimensi yang dapat mempertimbangkan beberapa variabel secara kompleks (Ihza Zuhendra & Hidayat, 2024).

Untuk mengatasi masalah tersebut, beberapa algoritma klasterisasi yang umum dapat digunakan. Namun dalam penelitian ini peneliti memilih algoritma *hierarchical clustering* (Klasterisasi Hierarkis). Algoritma ini menawarkan beberapa keunggulan untuk menganalisis masalah ini, yaitu berbeda dari algoritma klasterisasi yang lain, di mana metode lain mengharuskan penentuan jumlah klaster di awal, namun *hierarchical clustering* mampu membentuk struktur pengelompokan secara bertingkat sehingga dapat memvisualisasikan data melalui dendogram, yang dapat memberikan pemahaman yang lebih mendalam tentang hubungan kedekatan antar klaster. Meskipun demikian, penerapan standar dari algoritma ini masih memperlakukan seluruh variabel atau fitur dengan tingkat prioritas yang setara. Apabila kedua fitur diberi bobot yang sama tanpa mempertimbangkan peran strategisnya, hasil segmentasi dapat menjadi bias dan kurang mencerminkan kondisi sebenarnya di lapangan (Hasim Azari et al., 2024a).

Oleh karena itu, penelitian ini mengusulkan sebuah pengembangan metodologis dengan menerapkan pembobotan fitur. Setelah perhitungan bobot selesai dilakukan, bobot tersebut akan diterapkan pada fitur Produksi (volume) dan Produktivitas (efisiensi) sebelum data diproses oleh algoritma *Hierarchical clustering*. Lalu pada algoritma *hierarchical clustering* metode yang akan dipakai dalam tugas akhir ini ialah metode *ward*. Dengan demikian, hasil segmentasi tidak hanya berdasarkan



pada kedekatan numerik data, tetapi juga mempertimbangkan signifikansi variabel yang berpengaruh terhadap produksi padi.

Penerapan pembobotan fitur pada *hierarchical clustering* ini diharapkan dapat menghasilkan segmentasi kabupaten atau kota di Sumatera Utara yang lebih tajam dan juga relevan pada produksi dan produktivitasnya. Melalui segmentasi tersebut, pemerintah daerah dapat mengetahui wilayah mana yang memiliki performa tinggi dibagian produksi maupun produktivitas. Sehingga intervensi kebijakan pengelolaan seperti peningkatan sarana produksi, pelatihan petani atau pengelolaan lahan dapat dilakukan secara lebih tepat sasaran dan merata. Selain itu juga hasil analisis ini dapat menjadi dasar bagi perancangan pembangunan pertanian berkelanjutan kedepannya.

## 1.2 Rumusan Masalah

Berdasarkan permasalahan tersebut, maka rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana penerapan *Hierarchical clustering* dalam mengelompokkan wilayah berdasarkan hasil pembobotan fitur?
2. Bagaimana menerapkan bobot terhadap dua fitur sesuai dengan kepentingan yang dibutuhkan?

## 1.3 Batasan Masalah

Agar penelitian ini tetap fokus pada tujuan yang ingin dicapai, maka ruang lingkup penelitian dibatasi pada beberapa hal berikut:

1. Penelitian ini menggunakan algoritma *Hierarchical Clustering*, yang dikombinasikan dengan pembobotan fitur.

2. Penelitian ini berfokus pada dua variabel utama yang dianggap paling berpengaruh dalam menggambarkan kondisi produksi padi di Provinsi Sumatera Utara, yaitu produksi padi (ton) dan produktivitas padi (ku/ha). Kedua variabel tersebut dipilih karena secara langsung mencerminkan hasil dan efisiensi pertanian padi di setiap wilayah.
3. Data yang digunakan merupakan data sekunder yang diperoleh dari instansi Badan Pusat Statistik (BPS) dengan tautan <https://sumut.bps.go.id/id/statistics-table/2/MTU0IzI=/luas-panen-produksi-dan-rata-rata-produksi-padi-sawah-dan-padi-ladang-menurut-kabupaten-kota.html> .
4. Analisis difokuskan pada wilayah Provinsi Sumatera Utara, dengan data yang diambil berdasarkan periode tahun 2023–2024.
5. Penelitian ini tidak membahas perbandingan dengan metode *clustering* lain maupun analisis prediktif, tetapi hanya menitikberatkan pada penerapan pembobotan fitur dengan metode AHP dan segmentasi wilayah produksi padi.

#### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini secara umum ialah menerapkan pembobotan fitur dan mengintegrasikannya dengan metode *Hierarchical clustering* guna melakukan segmentasi wilayah berdasarkan produksi dan produktivitas padi di Provinsi Sumatera Utara. Secara lebih rinci, tujuan penelitian ini dapat diuraikan sebagai berikut:

1. Menentukan bobot kepentingan antara fitur produksi dan produktivitas.
2. Menerapkan metode *Hierarchical clustering* untuk mengelompokkan wilayah berdasarkan bobot fitur AHP yang diperoleh.

3. Menghasilkan peta segmentasi wilayah produksi padi di Sumatera Utara yang dapat dijadikan dasar dalam pengambilan keputusan dan perencanaan kebijakan pertanian.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Secara Teoritis, penelitian ini menambah wawasan dan referensi mengenai penerapan pembobotan fitur serta penerapan *Hierarchical clustering* untuk segmentasi data, khususnya di bidang pertanian.
2. Secara Praktis, hasil penelitian dapat membantu instansi terkait dalam mengidentifikasi dan mengelompokkan wilayah produksi padi di Sumatera Utara, sehingga kebijakan pengembangan pertanian dapat dilakukan lebih efektif dan tepat sasaran.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Padi

Padi (*Oryza sativa L.*) merupakan komoditas pangan utama di Indonesia yang memiliki peranan vital dalam menunjang ketahanan pangan nasional (Miranda Yustika Elmaria Marbun, 2024). Sebagian besar masyarakat Indonesia menjadikan beras sebagai makanan pokok, sehingga ketersediaan dan tingkat produksinya sangat memengaruhi stabilitas ekonomi dan sosial. Tanaman padi dapat tumbuh baik pada daerah beriklim tropis dengan curah hujan yang cukup dan kondisi tanah yang sesuai, terutama lahan sawah dengan tingkat kesuburan tinggi dan sistem irigasi yang baik (Alridiwersah, 2022).

Padi juga dapat dikategorikan berdasarkan varietas, umur tanaman, serta potensi hasil yang dihasilkan. Faktor-faktor lain seperti luas lahan, produktivitas, kualitas bibit, serta teknik budidaya sangat menentukan hasil produksi padi di suatu wilayah. Khususnya di Indonesia, produktivitas padi sering sekali digunakan sebagai indikator terhadap keberhasilan sektor pertanian, karena dipercaya mencerminkan efisiensi penggunaan lahan dan input pertanian dalam menghasilkan hasil berupa gabah atau beras (Sufriadi Dedi & Hamid. A, 2021).

Kemudian dalam konteks wilayah, tentunya setiap daerah mempunyai karakteristik produksi padi yang berbeda-beda, bisa dilihat dari tingkat hasil, efisiensi lahan yang ada. Maupun sistem irigasinya. Perbedaan inilah yang menjadi analisis terhadap produksi padi itu sangat penting, terutama dalam menentukan strategi peningkatan hasil pertaniannya (Gobel et al., 2024).



## 2.2 Produksi dan Produktivitas Padi

Produksi padi merupakan salah satu indikator penting dalam mengukur tingkat keberhasilan sektor pertanian, khususnya pada komoditas pangan utama di Indonesia. Secara umum, produksi padi adalah hasil total panen gabah yang diperoleh dari suatu luas lahan dalam periode waktu tertentu, biasanya dinyatakan dalam satuan ton per hektar. Tingkat produksi juga mencerminkan kemampuan suatu wilayah dalam memanfaatkan daya pertanian secara efektif dan juga efisien. Faktor-faktor yang dapat mempengaruhi bisa dibedakan menjadi faktor alamiah dan teknis. Peningkatan produksi padi tidak hanya bergantung pada perluasan lahan, tetapi juga pada efisiensi penggunaan faktor produksi dan penerapan teknologi yang tepat guna (Hilalullaily et al., 2021).

Produktivitas padi merupakan ukuran tingkat efisiensi hasil produksi tanaman padi terhadap satuan luas lahan yang digunakan. Secara umum, produktivitas dihitung sebagai jumlah hasil panen gabah dalam satuan ton per hektar. Tingkat produktivitas menjadi indikator penting untuk menilai keberhasilan sistem pertanian, karena mencerminkan sejauh mana sumber daya seperti lahan, tenaga kerja, air, dan input produksi dapat dimanfaatkan secara optimal. Peningkatan produksi padi tidak hanya menuntut peningkatan volume produksinya, tetapi juga mengoptimalkan dalam pemanfaatan sumber daya lahan dan penerapan teknologi pertanian yang adaptif terhadap perubahan lingkungan (Auliya et al., 2024).

### 2.3 *Analytic Hierarchy Process (AHP)*

AHP merupakan suatu metode pengambilan keputusan multi kriteria yang telah dikembangkan oleh Thomas L. Saaty pada tahun 1970-an. Pada metode ini memungkinkan pengambilan keputusannya dalam mengevaluasi dan memprioritaskan berbagai kriteria berdasarkan preferensi relatif (Rozali et al., 2023). Kekuatan AHP terletak pada perbandingan berpasangan (*pairwise comparison*) dan kemampuan untuk mengevaluasi konsistensi logika penilaian dengan menghitung nilai *Consistency Ratio (CR)* (Aruan & Novita, 2025). AHP memiliki prinsip sebagai berikut:

#### 1. *Decomposition* (Dekomposisi)

Permasalahan kompleks diuraikan menjadi elemen-elemen yang lebih sederhana dalam bentuk struktur hierarki. Hierarki tersebut umumnya terdiri dari tiga tingkat, yaitu tujuan, kriteria, dan alternatif. Dalam pembobotan fitur, struktur hierarki dapat disederhanakan menjadi tujuan dan kriteria (fitur).

#### 2. *Comparative Judgment* (Perbandingan Berpasangan)

Setiap kriteria dibandingkan secara berpasangan untuk menentukan tingkat kepentingannya relatif terhadap kriteria lain. Penilaian dilakukan menggunakan skala numerik yang telah ditetapkan oleh Saaty.

#### 3. *Synthesis of Priority* (Sintesis Prioritas)

Hasil perbandingan berpasangan diolah secara matematis untuk menghasilkan bobot prioritas masing-masing kriteria, yang mencerminkan tingkat kepentingannya dalam pengambilan keputusan.

## 2.4 *Machine Learning*

*Machine learning* adalah bagian dari *artificial intelligence* yang didefinisikan sebagai metode yang dimana metode tersebut belajar dari pengalaman dan dapat menyesuaikan dengan input baru untuk melakukan suatu tindakan otomatis dan membuat keputusan yang sesuai. *Machine learning* juga membantu peran manusia dalam pengambilan keputusan dengan otomatis, belajar dari data dan menentukan pola (Prasetyo & Dewayanto, 2024).

*Machine learning* secara umum dibagi menjadi tiga jenis utama, yaitu:

- a. *Supervised learning*, yaitu pembelajaran dengan data berlabel, dimana model dilatih untuk memprediksi output berdasarkan input yang diketahui.
- b. *Unsupervised learning*, yaitu pembelajaran tanpa label, dimana algoritma mencari pola atau struktur dalam data tanpa mengetahui output sebelumnya.
- c. *Reinforcement learning*, yaitu pembelajaran berbasis pengalaman melalui umpan balik atau reward dari lingkungan.

## 2.5 *Hierarchical Clustering*

Sebelum masuk ke *hierarchical clustering*, *clustering* adalah salah satu teknik dari data mining yang dimana bertujuan mengelompokkan objek yang dilihat berdasarkan kemiripan dari karakteristiknya lalu kemudian dipisahkan dari kelompok objek lainnya, sehingga dapat menciptakan kelompok objek yang homogen atau heterogen terhadap kelompok lainnya. *Hierarchical clustering* adalah salah satu teknik *clustering* yang membentuk hirarki seperti struktur pohon.

Sehingga proses pengelompokkannya dilakukan secara bertahap atau bertingkat (Pratama et al., 2021) .

Ada dua metode yang umum ada di *hierarchical clustering* yaitu:

a. *Agglomerative hierarchical*

Pada metode ini melakukan pendekatan dengan *bottom-up*, yang dimana awalnya menganggap setiap objek memiliki *cluster*-nya sendiri. Kemudian dilakukan penggabungan objek secara bertahap, dimulai dari objek-objek yang paling mirip hingga semua kelompok tergabung menjadi satu *cluster*.

b. *Divisive hierarchical*

Pada metode ini melakukan pendekatan dengan *top-up*, yang dimana pada awalnya semua objek akan berada dalam satu *cluster* yang sama. Lalu, disetiap tahapannya *cluster-cluster* tersebut dipecah menjadi *cluster-cluster* yang lebih kecil hingga setiap objeknya berada dalam *cluster*-nya tersendiri (Suraya & Wijayanto, 2022).

Dalam metode *agglomerative hierarchical* dibagi beberapa metode pendekatan yaitu:

a. *Single Linkage*

Metode pendekatan ini yaitu jarak antara dua klaster didefinisikan sebagai jarak minimum antara dua titik data dari masing-masing klaster. Dengan kata lain , kesamaan klaster ditentukan oleh titik yang paling dekat. Dalam metode ini cenderung menghasilkan klaster yang bentuknya memanjang atau seperti rantai dan sangat sensitif terhadap *noise* dan juga *outlier* (Fadhilah & Jananto, 2021).

b. *Complete Linkage*



Metode pendekatan ini yang dimana jarak antar kluster didefinisikan sebagai jarak maksimum antara dua titik data dari masing-masing kluster. Metode ini cenderung menghasilkan kluster-kluster yang lebih kompak (padat) dan berbentuk bola. Berbeda dengan *single linkage*, metode ini kurang sensitif terhadap *outlier* (Ulfatul Syahara et al., 2024).

c. *Average Linkage*

Metode ini bisa dikatakan gabungan dari *single* dan *complete linkage*. Dimana jarak antar klasternya dihitung sebagai jarak rata-rata dari semua kemungkinan pasangan titik data antar dua kluster tersebut (Butar Butar, 2023).

d. *Metode Ward*

*Ward hierarchical clustering* merupakan teknik pengelompokan yang bertujuan untuk meminimalkan jumlah kuadrat dari jarak antara titik-titik data dengan pusat kluster. Metode ini memiliki salah satu keunggulan yaitu kendalanya terhadap *noise* dan *outlier*. Ini disebabkan oleh pendekatannya yang mempertimbangkan semua titik data dalam kluster saat menghitung varians, sehingga pengaruh dari *outlier* bisa diminimalisir. Selain itu, metode ini juga sangat efektif dalam menghasilkan kluster yang seimbang dalam hal ukuran, karena berfokus pada meminimalkan varians antar kluster. Proses pengelompokannya yang iteratif dan hirarkis memungkinkan identifikasi struktur data yang lebih dalam, dan hasilnya mudah divisualisasikan dalam bentuk dendrogram. Selain itu, metode ini memungkinkan analisis yang lebih fleksibel dan adaptif terhadap karakteristik data (Aberson & Sidebang, 2024) .

Oleh karena itu penelitian ini menggunakan metode *ward* dalam proses *clustering* terhadap segmentasi wilayah produksi padi di Sumatera Utara. Agar lebih jelas, *Ward linkage* memiliki rumus sebagai berikut:

$$\Delta E = \frac{\eta_A \eta_B}{\eta_A + \eta_B} (\bar{X}_A - \bar{X}_B)^2 \quad (2.1)$$

Keterangan:

- $\eta_A, \eta_B$  = jumlah anggota tiap cluster
- $\bar{X}_A$  = rata – rata masing – masing cluster
- $(\bar{X}_A - \bar{X}_B)^2$  = selisih rata – rata kuadrat

### 2.5.1 Pseudocode Hierarcical Clustering

---

#### *Pseudocode Hierarcical Clustering*

---

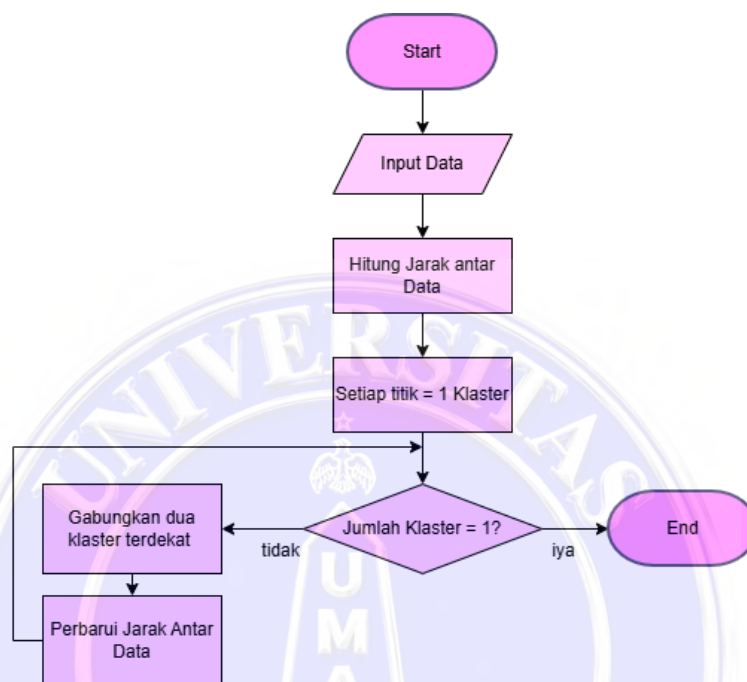
*Input: Data points D = {x<sub>1</sub>, x<sub>2</sub>, x<sub>3</sub>, ... x<sub>n</sub>}*

*Output: Dendrogram / Hierarchical cluster tree*

1. Mulai dengan setiap titik data sebagai *cluster* sendiri (*n cluster*)
  2. Hitung matriks jarak antar semua *cluster* (berdasarkan jarak antar titik)
  3. *WHILE* jumlah *cluster* > 1 *DO*
    - a. Temukan dua *cluster* terdekat A dan B berdasarkan metode *linkage* yang dipilih.
    - b. Gabungkan *cluster* A dan B menjadi *cluster* baru C
    - c. Update matriks jarak:
      - Hitung jarak antara C dan setiap *cluster* lain menggunakan rumus *linkage*
    - d. Hapus baris dan kolom matriks jarak untuk *cluster* A dan B
    - e. Tambahkan baris dan kolom untuk *cluster* C
  4. Kembalikan *dendrogram* atau hierarki *cluster*.
-

### 2.5.2 Flowchart Hierarchical Clustering

Adapun *flowchart* dasar dari algoritma *agglomerative hierarchical clustering* dapat dilihat pada gambar dibawah ini:



**Gambar 2.1** *Flowchart* Dasar AHC

### 2.6 Silhouette Index (SI)

*Silhouette Index* atau *silhouette coefficient* adalah penggabungan dari metode *cohesion* dan *separation* yang merupakan metode evaluasi untuk *cluster*. Jarak antara data nanti dapat dihitung menggunakan rumus *Euclidean Distance* dan *Manhattan Distance*. Nilai ini diperoleh untuk mendapatkan informasi tentang kualitas hasil *clustering* pada proses *clustering*. Jumlah *cluster* dinyatakan baik apabila bilai SI mendekati 1 atau tinggi (Suraya & Wijayanto, 2022).

Adapun rumus dari *Silhouette Index* yaitu:

$$SI = \frac{1}{n} \sum_{i=1}^n \left( \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \right) \quad (2.2)$$

## 2.7 Davies-Bouldin Index (DBI)

DBI pertama kali dikenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. DBI merupakan nilai yang dihitung berdasarkan jumlah *cluster* yang dimana dipakai sebagai dasar untuk mengevaluasi hasil proses dari pengelompokan. Tujuan evaluasi ini dilakukan yaitu untuk menilai kualitas dari pengelompokan. Jumlah *cluster* dikatakan baik apabila nilai DBI mendekati 0 atau rendah (Suraya & Wijayanto, 2022). Adapun rumus dari DBI yaitu :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j,...k}) \quad (2.2)$$

## 2.8 Calinski Harabaz Index (CHI)

*Calinski-Harabasz Index* yang dikenal dengan *Variance Ratio Criterion* adalah evaluasi hasil klaster berdasarkan perbandingan nilai jumlah kuadrat antar klaster (SSB) sebagai pemisahan dan nilai jumlah kuadrat dalam klaster (SSW) sebagai kekompakan dikalikan dengan faktor normalisasi yaitu selisih jumlah data dengan jumlah klaster dibagi banyaknya klaster dikurangi satu, dimana nilai yang lebih tinggi menunjukkan struktur klaster yang lebih baik (Sikana & Wijayanto, 2021). Adapun rumus CHI ialah:

$$CH = \frac{tr(SSB)}{tr(SSW)} \times \frac{n - k}{k - 1} \quad (2.3)$$



## 2.9 Penelitian Terdahulu

**Tabel 2.1** Penelitian Terdahulu

| No | Penulis (Tahun)                                                | Topik                                                                                                                                                                                 | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|----|----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. | Anne Mudya Yolanda, Husna Savira (2024)                        | Segmentasi provinsi-provinsi di Indonesia berdasarkan data deret waktu produksi padi menggunakan algoritma <i>Dynamic Time Warping</i> (DTW) dan teknik <i>K-Medoids clustering</i> . | Hasil penelitian menunjukkan bahwa terdapat tiga klaster utama provinsi berdasarkan karakteristik produksinya. Klaster 1 mencakup provinsi dengan tingkat produksi padi tertinggi, klaster 2 meliputi daerah dengan tingkat produksi sedang, sedangkan klaster 3 terdiri dari provinsi dengan produksi terendah. Dengan menggunakan algoritma DTW, peneliti mampu mengukur kesamaan pola produksi antarprovinsi, yang kemudian menunjukkan adanya perbedaan pola produksi akibat faktor lingkungan dan praktik pertanian yang beragam di tiap daerah.                                                                     |
| 2. | Rhayhana Putri Justitia, Nurul Hidayat, dan Edy Santoso (2021) | Implementasi Metode <i>Agglomerative Hierarchical clustering</i> Pada Segmentasi Pelanggan Barbershop (Studi Kasus : RichDjoe Barbershop Malang)                                      | Hasil dari penelitian ini menunjukkan bahwa metode <i>Agglomerative Hierarchical clustering</i> dengan parameter linkage seperti Single Linkage dan Average Linkage dapat secara efektif mengelompokkan data pelanggan RichDjoe Barbershops. Hasil terbaik didapat pada titik potong 391 dengan menghasilkan 8 <i>cluster</i> dan nilai silhouette coefficient rata-rata sebesar 0.968850698, yang menunjukkan bahwa sebagian besar data terkelompok dengan sangat baik. Sebanyak 376 dari 399 data yang tepat berada di dalam <i>cluster</i> yang sesuai, dan sebanyak 6 dari 8 <i>cluster</i> terbentuk secara optimal. |
| 3. | Sella Brina, Winona Rizky (2025)                               | Perbandingan Algoritma <i>K-Means</i> dan <i>Hierarchical clustering</i> dalam Segmentasi Kabupaten/ Kota di Jawa Timur Berdasarkan Data                                              | Penelitian ini membandingkan dua algoritma <i>clustering</i> , yaitu <i>K-Means</i> dan <i>Hierarchical clustering</i> , dalam melakukan segmentasi kabupaten/kota di Jawa Timur berdasarkan data perjalanan dan pergerakan                                                                                                                                                                                                                                                                                                                                                                                               |

|    |                                                          |                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----|----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                          | Perjalanan dan Pergerakan Wisatawan                                                                                  | wisatawan nusantara tahun 2023-2024. Hasil evaluasi menunjukkan bahwa <i>Hierarchical clustering</i> lebih unggul dalam hal stabilitas dan kejelasan struktur klaster, dilihat dari nilai Silhouette Score (0.5689) dan Dunn Index (0.2396). Penelitian menemukan tiga klaster utama, dengan Kota Surabaya menjadi outlier karena jumlah kunjungan tertinggi. Hasil ini diharapkan dapat mendukung strategi promosi pariwisata yang lebih terarah dan efisien di wilayah Jawa Timur                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| 4. | Hasim Azari, Dwi Hartanti, dan Aprilisa Arum Sari (2024) | Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode <i>Agglomerative Hierarchical clustering</i> | peneliti berhasil mengidentifikasi empat klaster utama berdasarkan tingkat Produktivitas padi, yaitu daerah dengan produktivitas tinggi sebanyak empat kabupaten, produktivitas sedang sebanyak tujuh kabupaten, produktivitas rendah sebanyak delapan belas kabupaten, dan produktivitas sangat rendah sebanyak sembilan kota. Nilai koefisien siluet sebesar 0,6347 menunjukkan bahwa hasil pengelompokan memiliki kualitas yang baik dan representatif terhadap kondisi nyata di lapangan. Temuan ini memberikan gambaran komprehensif mengenai kapasitas produksi padi di Jawa Timur, sehingga dapat membantu dalam mengenali wilayah yang unggul serta daerah yang perlu mendapat perhatian lebih dalam peningkatan produktivitas. Secara keseluruhan, penelitian ini memberikan kontribusi penting bagi pemerintah dan masyarakat sebagai acuan dalam pengambilan kebijakan dan strategi peningkatan ketahanan pangan, khususnya di sektor pertanian padi di Indonesia. |
| 5. | Nganta Ginting, (2024)                                   | Sistem Pendukung Keputusan Penerimaan                                                                                | Penelitian yang dilakukan membandingkan dua metode                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

|  |                      |                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |                      | Beasiswa dengan Metode SAW berbasis <i>website</i>                                                                                                                                                                | pengambilan keputusan, yaitu <i>Simple Additive Weighting</i> (SAW) dan <i>Weighted Product</i> (WP), dalam konteks pemilihan penerima beasiswa bagi mahasiswa di STIK-P Pembangunan Medan. Penelitian ini menekankan pentingnya penentuan bobot atribut secara sistematis untuk mendukung pengambilan keputusan yang lebih baik, sehingga dapat menghindari metode manual yang berpotensi di mana IPK dengan bobot 0,361 menjadi faktor yang paling dominan. Dengan penggunaan metode Information Gain memungkinkan penentuan bobot atribut secara objektif dan sistematis, sehingga dapat meningkatkan kualitas proses pengambilan keputusan dalam konteks alokasi beasiswa. |
|  | Karim et al., (2021) | Analisa Penerapan Metode Operational Competitiveness Rating Analysis (OCRA) dan Metode Multi Attribute Utility Theory (MAUT) Dalam Pemilihan Calon Karyawan Tetap Menerapkan Pembobotan Rank Order Centroid (ROC) | Hasil penelitian menunjukkan pembobotan kriteria menggunakan metode Rank Order Centroid (ROC) menghasilkan bobot untuk kriteria C1 hingga C7. Yang dimana Risa Sabrani (A5) memiliki nilai utilitas tertinggi 0.456, menjadikannya alternatif terbaik, dengan pembobotan untuk memberikan prioritas pada kriteria yang lebih penting, sehingga kriteria dengan bobot lebih tinggi, seperti Kinerja (C1), memiliki dampak yang lebih besar dalam menentukan nilai utilitas akhir dan peringkat alternatif.                                                                                                                                                                      |

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Sumber Data Penelitian

Pada penelitian ini, dataset yang dipakai yaitu data sekunder. Pemakaian data ini didasari oleh pertimbangan dalam efisiensi waktu dan juga ketersediaan data sudah terstruktur serta tervalidasi dari lembaga Badan Pusat Statistik (BPS) Sumatera Utara. Data yang peneliti gunakan yaitu data dari periode tahun 2023-2024 dengan jumlah data 64 baris. Data tersebut mencakup beberapa atribut yaitu luas panen (hektar), produktivitas (kuintal per hektar), total produksi (ton) dan luas baku sawah(ha).

**Tabel 3.1** Data Asli Penelitian

| Kabupaten/Kota   | Luas Baku Sawah(Ha) | Luas Panen Tanaman Padi (Ha) | Produktivitas Tanaman Padi (Ku/ha) | Rekap Produksi Padi (Ton) |
|------------------|---------------------|------------------------------|------------------------------------|---------------------------|
| Nias             | 5.884               | 8481,7                       | 44,99                              | 38155,62                  |
| Mandailing Natal | 11.963              | 19833,89                     | 39,47                              | 78282,95                  |
| Tapanuli Selatan | 11.942              | 18748,74                     | 48,64                              | 91188,83                  |
| Tapanuli Tengah  | 8.517               | 12079,15                     | 40,31                              | 48693,65                  |
| Tapanuli Utara   | 20.380              | 25293,86                     | 50,14                              | 126829,9                  |

#### 3.2 Perangkat Yang Dibutuhkan

Dalam penyusunan tugas akhir ini, diperlukan perangkat yang dapat membantu peneliti, baik dari segi *hardware* maupun *software*. Rincian perangkat yang digunakan sebagai berikut:

Perangkat keras (*Hardware*):

1. *Laptop ASUS*
2. *Processor Intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz (1.19 GHz)*
3. *Memory (RAM) 8 GB*

Perangkat lunak (*Software*)

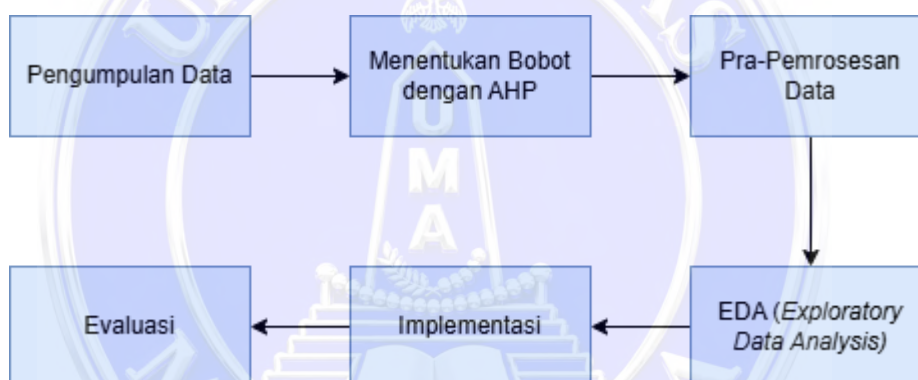
1) *Microsoft Word*

2) *Microsoft Excel*

3) *Google Colab*

### 3.3 Tahapan Penelitian

Pelaksanaan penelitian ini mengikuti serangkaian tahapan yang telah disusun secara sistematis untuk memastikan bahwa tujuan penelitian dapat tercapai secara efektif dan efisien. Tahapan-tahapan ini mencakup keseluruhan proses. Secara garis besar, prosedur penelitian yang dilakukan diuraikan sebagai berikut:



**Gambar 3.1** *Diagram Tahapan Penelitian*

#### 3.3.1 Pengumpulan Data

Proses pengumpulan data yaitu proses dimana data tersebut dikumpulkan, data pada penelitian ini didapat dari data sekunder yang dimana diakses melalui web Badan Pusat Statistik Sumatera Utara yang dimana rentang datanya diambil dari periode tahun 2023-2024.



### 3.3.2 Menentukan Bobot Fitur

Pada tahap menentukan bobot fitur ini ialah tahap dimana peneliti menentukan tingkat prioritas bobot yang akan diteliti, pada tahap inilah bobot produksi maupun produktivitas diberi bobot sesuai prioritas yang akan diterapkan pada proses klasterisasi nantinya.

### 3.3.3 Pra-Pemrosesan Data

Pada tahap pra pemrosesan data ini merupakan langkah yang penting sebelum dilakukan proses klasterisasi, dikarenakan untuk memastikan bahwa data yang akan digunakan berada dalam skala yang bentuknya itu sesuai untuk dianalisis. Pada penelitian ini, pra-pemrosesan dilakukan terhadap dua variabel utama, yaitu produktivitas dan produksi padi.

Langkah pertama yang akan dilakukan yaitu normalisasi data menggunakan metode *Min-Max Scaler*, yang akan merubah nilai setiap fiturnya ke rentang 0-1. Normalisasi ini diperlukan agar kedua variabel ini mempunyai skala yang *balance* atau sebanding. Setelah data dinormalisasi, dilakukan penerapan pembobotan yang telah dilakukan. Pembobotan ini bertujuan untuk memberikan tingkat kepentingan atau prioritas yang berbeda pada masing-masing variabel sesuai hasil perhitungan bobot yang telah dilakukan.

Pada tahap akhir pra-pemrosesan data ini akan menghasilkan data baru yang bernama data berbobot, data bobot ini berisi nilai produksi dan produktivitas padi yang telah dinormalisasi dan juga dibobotkan. Data inilah yang nantinya akan digunakan sebagai input utama dalam proses klasterisasi menggunakan algoritma *hierarchical clustering*, sehingga segmentasi yang dihasilkan akan lebih

merepresentasikan prioritas antarvariabel berdasarkan bobot yang ditentukan sebelumnya.

### 3.3.4 *Exploratory Data Analysis*

Tahap *Exploratory Data Analysis* (EDA) ini bertujuan untuk memahami karakteristik dan pola umum dari data yang akan digunakan dalam proses klusterisasi. Pada penelitian ini, EDA dilakukan setelah tahap pra-pemrosesan data, yaitu setelah data melalui proses normalisasi dan pembobotan fitur. Hal ini dilakukan bertujuan agar eksplorasi yang dilakukan merepresentasikan data dalam bentuk yang telah siap untuk proses klusterisasi.

Analisis yang akan dilakukan mencakup visualisasi distribusi nilai pada setiap fitur, pemeriksaan hubungan antar variabel, serta identifikasi kecenderungan atau kelompok awal yang mungkin muncul dari data yang telah dibobotkan. Melalui tahapan ini, diharapkan dapat melihat bagaimana pengaruh pembobotan terhadap persebaran data serta memastikan bahwa hasil transformasi dan pembobotan tidak menimbulkan distorsi atau ketidakseimbangan antar fitur.

Selain itu, EDA juga berfungsi sebagai langkah validasi terhadap kualitas data sebelum dilakukan proses *Hierarchical Clustering*, dengan memastikan bahwa tidak terdapat nilai ekstrem (*outlier*) yang dapat memengaruhi hasil pengelompokan.

### 3.3.5 Implementasi

Tahapan ini merupakan tahapan utama dalam penelitian ini, di mana dilakukan proses *hierarchical clustering* menggunakan metode *ward* dengan data yang telah melalui tahapan pembobotan fitur sebelumnya. Pembobotan fitur inilah

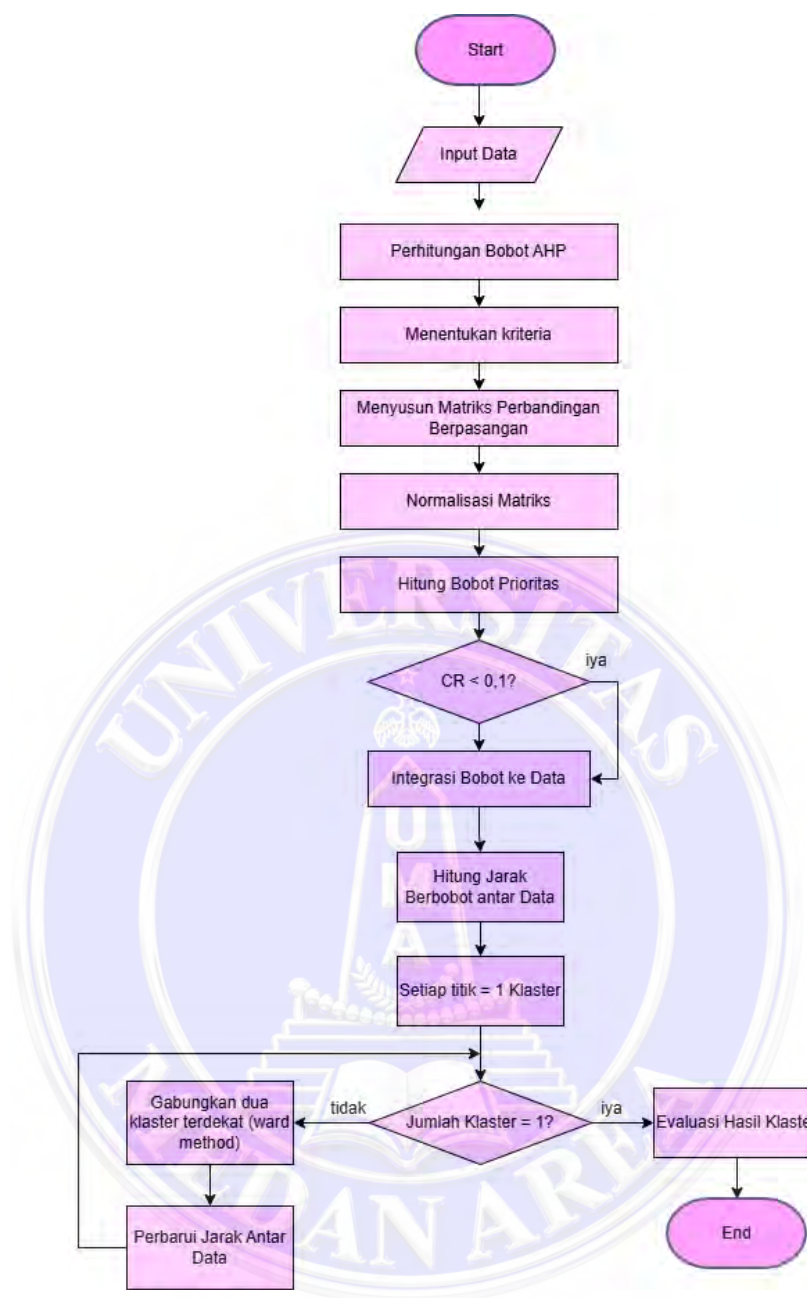
yang memberikan proporsi nilai kepentingan yang berbeda terhadap masing-masing variabel, yaitu produksi dan produktivitas, sehingga fitur dengan tingkat kepentingan lebih tinggi memiliki pengaruh yang lebih besar dalam proses klasterisasi.

Pada tahap ini, data yang telah dibobotkan digunakan untuk menghitung jarak *Euclidean* antar wilayah. Meskipun rumus dasar perhitungan jarak tidak diubah, hasil pembobotan menjadikan perhitungan tersebut setara dengan konsep *weighted Euclidean distance*, karena setiap fitur telah dikalikan dengan bobotnya masing-masing. Jarak yang dihasilkan menjadi dasar pembentukan *linkage matrix* menggunakan metode *Ward*, yang bertujuan meminimalkan variansi antar kelompok dalam setiap proses penggabungan.

Proses ini kemudian divisualisasikan dengan dendrogram, yang menampilkan struktur hierarki pengelompokan antar kabupaten/kota berdasarkan kesamaan karakteristik produksi dan produktivitas padi. Dengan demikian, tahap implementasi ini tidak hanya menghasilkan hasil klasterisasi, tetapi juga memperlihatkan bagaimana pembobotan fitur mampu memengaruhi hasil segmentasi wilayah, sehingga klaster yang terbentuk lebih merepresentasikan prioritas analisis sesuai dengan tingkat kepentingan yang ditetapkan.

#### a. *Flowchart Hierarchical Clustering*

*Flowchart hierarchical clustering* yang sudah dimodifikasi dapat dilihat melalui gambar dibawah ini. Perbedaan dengan *flowchart* dasar yaitu adanya penambahan integrasi bobot ke data yang dilakukan setelah memasukkan data.



**Gambar 3.2** Flowchart Hierarchical Clustering dan AHP

#### b. Jarak *Euclidean*

Sama halnya dengan *flowchart*, jarak *euclidean* juga sudah dimodifikasi untuk penelitian ini, yaitu adanya penambahan bobot yang akan dikalikan dengan setiap dataset. Rumus nya dapat dilihat dengan jelas dibawah ini:

$$x_{rs} = \sqrt{\sum_{k=1}^p w_k (y_{rk} - y_{sk})^2} \quad (3.1)$$

Keterangan :

- $x_{rs}$  = Jarak antara objek ke  $r$  dan ke  $s$
- $y_{rk}$  = Nilai objek ke  $r$  pada variabel ke  $k$
- $y_{sk}$  = Nilai objek ke  $s$  pada variabel ke  $k$
- $w_k$  = Bobot hasil perhitungan untuk ke fitur  $k$
- $p$  = Jumlah variabel

### 3.3.6 Evaluasi

Tahap evaluasi dilakukan untuk menilai kualitas hasil klusterisasi yang telah diperoleh dari proses *Hierarchical Clustering* dengan pembobotan fitur. Dua metrik utama yang digunakan dalam penelitian ini adalah *Davies–Bouldin Index* (DBI) dan *Silhouette Index* (SI).

*Davies–Bouldin Index* (DBI) digunakan untuk mengukur seberapa baik pemisahan antar kluster serta tingkat kekompakan dalam masing-masing kluster. Nilai DBI yang lebih rendah menunjukkan bahwa kluster yang terbentuk memiliki jarak yang cukup jauh antar kelompok (*well-separated*) dan data di dalam setiap kluster lebih homogen. Dengan demikian, semakin kecil nilai DBI yang diperoleh, semakin baik kualitas hasil pengelompokan.

Sementara itu, *Silhouette Index* (SI) digunakan untuk mengevaluasi tingkat kesesuaian setiap objek terhadap kluster yang ditempatinya dibandingkan dengan kluster lain. Nilai SI berkisar antara -1 hingga 1, di mana nilai mendekati 1 menandakan bahwa objek berada pada kluster yang tepat dan memiliki kesamaan



tinggi dengan anggota klaster lainnya. Nilai mendekati 0 menunjukkan batas atau tumpang tindih antar klaster, sedangkan nilai negatif menandakan kemungkinan kesalahan pengelompokan.

Dengan mengombinasikan kedua metrik ini, hasil evaluasi dapat memberikan gambaran yang efisien terhadap struktur klaster yang terbentuk, sekaligus memastikan bahwa penerapan pembobotan fitur menghasilkan segmentasi wilayah yang optimal dan representatif terhadap karakteristik produksi serta produktivitas padi di Sumatera Utara.

### 3.4 Perhitungan Manual

Untuk memperjelas proses pengelompokan yang dilakukan oleh AHP dan algoritma *hierarchical clustering*, pada bagian ini disajikan contoh perhitungan manual menggunakan data terpilih. Perhitungan ini bertujuan untuk menunjukkan tahapan pembobotan fitur berdasarkan metode *Analytical Hierarchy Process* (AHP), normalisasi data, serta perhitungan jarak antar wilayah sebagai dasar pembentukan klaster.

#### 3.4.1 *Analytical Hierarchy Process* (AHP)

##### a. Pembobotan Fitur Produksi Padi

1. Langkah-langkah terdiri dari: Penentuan Kriteria

| Kode | Kriteria                   |
|------|----------------------------|
| C1   | Produktivitas tanaman padi |
| C2   | Rekap produksi padi        |
| C3   | Luas panen padi            |
| C4   | Luas baku tanah            |

Dikarenakan tahap ini berfokus pada produksi padi, maka kriteria C2 diprioritaskan.

## 2. Penyusunan Matriks Perbandingan Berpasangan

Matriks perbandingan disusun berdasarkan skala Saaty (1–9).

$$A = \begin{bmatrix} 1 & \frac{1}{3} & 3 & 5 \\ 3 & 1 & 5 & 7 \\ \frac{1}{3} & \frac{1}{5} & 1 & 3 \\ \frac{1}{5} & \frac{1}{7} & \frac{1}{3} & 1 \end{bmatrix}$$

Makna perbandingan:

- Produksi (C2) lebih penting dari Produktivitas (C1)
- Produksi sangat lebih penting dari Luas Panen dan Luas Baku Sawah
- Produktivitas lebih penting dari Luas Panen dan Luas Baku Sawah

## 3. Normalisasi Matriks Perbandingan

### 3.1 Hitung jumlah setiap kolom

| Kolom | Jumlah |
|-------|--------|
| C1    | 4,533  |
| C2    | 1,676  |
| C3    | 9,333  |
| C4    | 16,000 |

### 3.2 Normalisasi matriks

Setiap elemen dibagi dengan total kolomnya.

Contoh:

$$a_{11}^{norm} = \frac{1}{4,533} = 0,2266$$

Hasil normalisasi:

$$A_{norm} = \begin{bmatrix} 0,2206 & 0,1990 & 0,3214 & 0,3125 \\ 0,6618 & 0,5965 & 0,5357 & 0,4375 \\ 0,0735 & 0,1193 & 0,1071 & 0,1875 \\ 0,0441 & 0,0852 & 0,0357 & 0,0625 \end{bmatrix}$$

## 4. Perhitungan Bobot

Bobot diperoleh dengan rata-rata setiap baris:

| Kriteria             | Bobot  |
|----------------------|--------|
| Produktivitas (C1)   | 0,2633 |
| Produksi (C2)        | 0,5579 |
| Luas Panen (C3)      | 0,1219 |
| Luas Baku sawah (C4) | 0,0569 |

## 5. Uji Konsistensi (*Consistency Ratio*)

5.1 Hitung nilai  $\lambda$  dan menemukan  $\lambda_{max} = 4,116$

5.2 *Consistency Index* (CI)

$$CI = \frac{\lambda_{max} - n}{n - 1} = \frac{4,116 - 4}{3} = 0,0387$$

5.3 *Consistency Rasio* (CR)

Nilai *Random Index* (RI) untuk  $n = 4$  adalah **0.90**

$$CR = \frac{CI}{RI} = \frac{0,0387}{0,90} = 0,0433$$

## 6. Interpretasi Konsisten

Karena  $CR = 0,0433 < 0,10$

Maka Matriks perbandingan dinyatakan konsisten dan bobot dapat digunakan dalam analisis klaster.

### 3.4.2 Perhitungan Manual *Hierarchical Clustering*

1. Data yang digunakan:

| Wilayah             | Produktivitas (ku/ha) | Produksi (ton) | Luas Panen (Ha) | Luas Baku Sawah (Ha) |
|---------------------|-----------------------|----------------|-----------------|----------------------|
| Deli Serdang (A)    | 62,44                 | 313546,40      | 50213,59        | 34280                |
| Serdang Begadai (B) | 65,17                 | 333128,70      | 51117,61        | 28337                |
| Simalungun (C)      | 52,16                 | 143824,15      | 27573,66        | 25444                |

2. Normalisasi Data

$$\text{Rumus: } x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Contoh :

Normalisasi Produktivitas:

$$\text{Min} = 52,16 \text{ max} = 65,17$$

$$\text{Wilayah A} = \frac{(62,44 - 52,16)}{(65,17 - 52,16)} = 0,790$$

### 3. Matriks Data Ternormalisasi

| Wilayah | Produktivitas | Produksi | Luas Panen | Luas Baku |
|---------|---------------|----------|------------|-----------|
| A       | 0,790         | 0,897    | 0,963      | 1,000     |
| B       | 1,000         | 1,000    | 1,000      | 0,327     |
| C       | 0,000         | 0,000    | 0,000      | 0,000     |

### 4. Pembobotan dengan AHP

| Kriteria             | Bobot  |
|----------------------|--------|
| Produktivitas (C1)   | 0,2633 |
| Produksi (C2)        | 0,5579 |
| Luas Panen (C3)      | 0,1219 |
| Luas Baku sawah (C4) | 0,0569 |

### 5. Data Berbobot

$$x_{\text{bobot}} = x' \times w$$

| Wilayah | Produktivitas | Produksi | Luas Panen | Luas Baku |
|---------|---------------|----------|------------|-----------|
| A       | 0,208         | 0,500    | 0,117      | 0,057     |
| B       | 0,263         | 0,558    | 0,122      | 0,019     |
| C       | 0,000         | 0,000    | 0,000      | 0,000     |

### 6. Perhitungan Jarak Euclidien

Jarak A-B

$$\sqrt{(0,208 - 0,263)^2 + (0,500 - 0,558)^2 + (0,117 - 0,122)^2 + (0,057 - 0,019)^2}$$

$$= \sqrt{0,0030 + 0,0034 + 0,0000 + 0,0014} = \sqrt{0,0078} = 0,088$$

- Jarak A-C

$$\sqrt{(0,208)^2 + (0,500)^2 + (0,117)^2 + (0,057)^2} = 0,557$$

- Jarak B-C

$$\sqrt{(0,263)^2 + (0,558)^2 + (0,122)^2 + (0,019)^2} = 0,630$$

### 7. Matrix Jarak

$$M = \begin{bmatrix} 0 & 0,080 & 0,557 \\ 0,088 & 0 & 0,630 \\ 0,557 & 0,630 & 0 \end{bmatrix} \text{ atau}$$

|   | A     | B     | C     |
|---|-------|-------|-------|
| A | 0     | 0,080 | 0,557 |
| B | 0,088 | 0     | 0,630 |
| C | 0,557 | 0,630 | 0     |

#### 8. Proses *Hierarchical Clustering (Ward)*

- Klaster awal: {A}, {B}, {C}
- Jarak terkecil: **A–B (0.088)** → Digabung menjadi klaster {A,B}
- Klaster {A,B} kemudian dibandingkan dengan C





## BAB V

### KESIMPULAN

#### 5.1 Kesimpulan

Penelitian ini menghasilkan suatu segmentasi wilayah produksi padi di Provinsi Sumatera Utara yang disusun melalui penerapan pembobotan fitur menggunakan metode *Analytical Hierarchy Process* (AHP) dan algoritma *Hierarchical Clustering*, sehingga mampu menggambarkan perbedaan karakteristik wilayah secara lebih objektif dan terstruktur. Dengan menghasilkan pembentukan kluster terbaik dengan metode *ward* pada fokus produksi padi, dengan evaluasi menggunakan *Silhouette Index* (SI), *Davies–Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI) menunjukkan bahwa dua kluster merupakan konfigurasi paling optimal, ditandai dengan nilai SI 0,7182, DBI 0,2266, dan CHI 39,2039. Hasil ini menunjukkan bahwa wilayah di Sumatera Utara terbagi ke dalam dua kelompok utama, yaitu wilayah dengan produksi padi tinggi dan wilayah dengan produksi menengah hingga rendah.

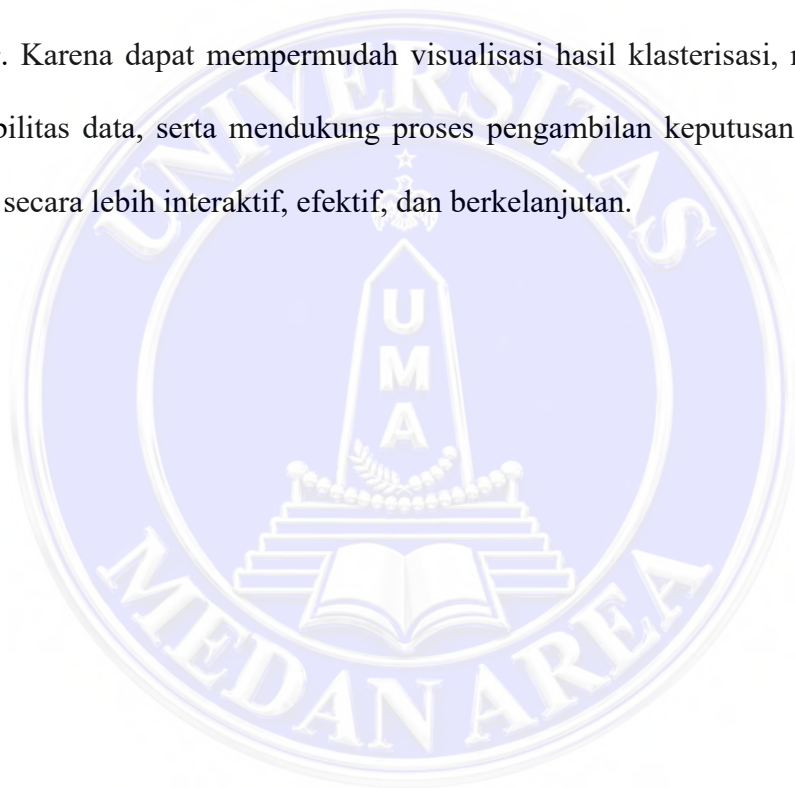
Sementara itu, pada fokus produktivitas padi, konfigurasi empat kluster dengan nilai SI 0,5480, DBI 0,5209, dan CHI 72,3464. Empat kluster tersebut mencerminkan variasi tingkat efisiensi produksi padi dari sangat tinggi hingga rendah, yang menunjukkan bahwa pola produktivitas padi lebih beragam dibandingkan produksi padi.

#### 5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, disarankan agar penelitian selanjutnya menggunakan jumlah kriteria yang lebih banyak dalam proses

pembobotan fitur, sehingga perbedaan hasil segmentasi dapat dianalisis dan dibandingkan secara lebih jelas antara penggunaan kriteria yang terbatas dan kriteria yang lebih beragam. Penambahan kriteria diharapkan mampu memberikan gambaran yang lebih menyeluruh terhadap karakteristik wilayah serta meningkatkan ketajaman hasil klasterisasi.

Selain itu, penelitian ini dapat dikembangkan lebih lanjut dengan mengimplementasikan hasil analisis ke dalam bentuk aplikasi berbasis *web* atau *mobile*. Karena dapat mempermudah visualisasi hasil klasterisasi, meningkatkan aksesibilitas data, serta mendukung proses pengambilan keputusan oleh instansi terkait secara lebih interaktif, efektif, dan berkelanjutan.



## DAFTAR PUSTAKA

- Aberson, A. M., & Sidebang, S. (2024). Pemetaan Kemiskinan Digital Kabupaten/Kota di Sumatera Utara Menggunakan Ward Hierarchical Clustering Digital Poverty Mapping of Regencies/Cities in North Sumatra Using Ward Hierarchical Clustering. In (Vol. 3, Issue 2).
- Alridiwersah, dkk. (2022). Evaluasi Kesesuaian Lahan Padi Sawah Di Desa Tanjung Kubah Kecamatan Air Putih. *Agrica Ekstensia*, 16(1). <https://doi.org/10.55127/ae.v16i1.110>
- Aruan, Y. E. L., & Novita, N. (2025). Sistem Pendukung Keputusan Berbasis AHP untuk Pemilihan Hunian di Kawasan Perkotaan yang Padat. *Sistem Pendukung Keputusan Dengan Aplikasi*, 4(2), 57–71. <https://doi.org/10.55537/spk.v4i2.1289>
- Auliya, D., Rosandi, A. H., & Subroto, W. T. (2024). Analisis Perubahan Iklim terhadap Produktivitas Padi di Jawa Timur. *Diponegoro Journal of Economics*, 13(3), 55–65. <https://doi.org/10.14710/djoe.47595>
- Brina, S., & Rizky, W. (2025). Perbandingan Algoritma K-Means Dan Hierarchical Clustering Dalam Segmentasi Kabupaten/Kota Di Jawa Timur Berdasarkan Data Perjalanan Dan Pergerakan Wisatawan. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Issue 5).
- Butar Butar, R. P. (2023). Analisis Hierarchical Dan Non-Hierarchical Clustering Untuk Pengelompokkan Potensi Ekonomi Kelautan Indonesia 2021. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 11(3), 543. <https://doi.org/10.26418/justin.v11i3.67283>
- Fadhilah, A. N., & Jananto, A. (2021). Klasterisasi Literatur Mahasiswa Menggunakan Metode AHCDi Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah. *Dinamika Informatika*, 13(1), 9–17.
- Gobel, Y. A., Hasan, D., Indrianti, A., Djibran, M. M., & Djaini, A. (2024). *Analysis Of Rice Production In Limboto District, Gorontalo* (Vol. 7, Issue 1).

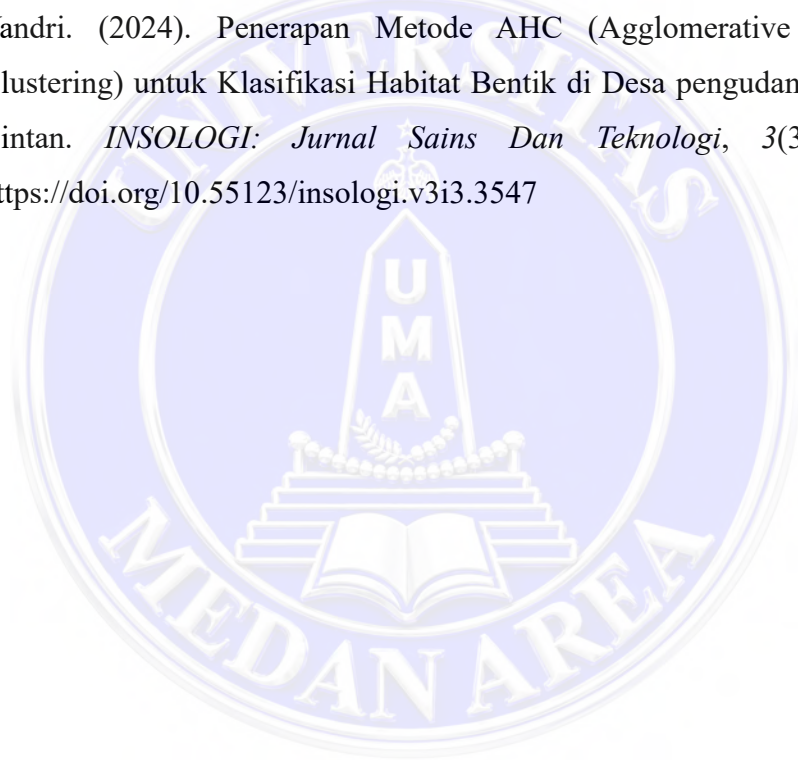
- Hasim Azari, Dwi Hartanti, & Aprilisa Arum Sari. (2024a). Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode Agglomerative Hierarchical Clustering. *Infotek: Jurnal Informatika Dan Teknologi*, 7(2), 379–389. <https://doi.org/10.29408/jit.v7i2.26016>
- Hasim Azari, Dwi Hartanti, & Aprilisa Arum Sari. (2024b). Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode Agglomerative Hierarchical Clustering. *Infotek: Jurnal Informatika Dan Teknologi*, 7(2), 379–389. <https://doi.org/10.29408/jit.v7i2.26016>
- Hilalullaili, R., Kusnadi, N., & Rachmina, D. (2021). Analisis Efisiensi Usahatani Padi di Jawa dan Luar Jawa, Kajian Prospek Peningkatan Produksi Padi Nasional. *Jurnal Agribisnis Indonesia*, 9(2), 143–153. <https://doi.org/10.29244/jai.2021.9.2.143-153>
- Homepage, J., Pratama Simanjuntak, K., & Khaira, U. (2021). *MALCOM: Indonesian Journal of Machine Learning and Computer Science Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering*. 1, 7–16.
- Ihza Zuhendra, M., & Hidayat, R. (2024). *Penerapan Data Mining Untuk Klasterisasi Tingkat Kemiskinan Berdasarkan Data Terpadu Kesejahteraan Sosial (DTKS)*. 7(1), 32–41.
- Karim, A., Esabella, S., Kusmanto, K., Mesran, M., & Hasanah, U. (2021). Analisa Penerapan Metode Operational Competitiveness Rating Analysis (OCRA) dan Metode Multi Attribute Utility Theory (MAUT) Dalam Pemilihan Calon Karyawan Tetap Menerapkan Pembobotan Rank Order Centroid (ROC). *Jurnal Media Informatika Budidarma*, 5(4), 1674. <https://doi.org/10.30865/mib.v5i4.3265>
- Miranda Yustika Elmaria Marbun. (2024). Pelestarian plasma nutfah padi untuk mendukung ketahanan pangan beras Indonesia. *Journal Of Crutucal Ecology*. <https://doi.org/10.61511/jcreco.v1i2.1170>



- Mudya, A., Dan, Y., & Savira, H. (2024). *Segmentasi Provinsi di Indonesia berdasarkan Data Runtun Waktu Produksi Padi dengan Algoritma DTW dan K-Medoids Clustering Segmentasi Provinsi di Indonesia*.
- Nadilla, S., Syahrifa, W., Razi, F., Maulana, R., & Ula, M. (2024). *Analisis Perbandingan Kinerja Algoritma K-Means Dan k-Medoids Untuk Klasterisasi Produksi Padi Di Pulau Sumatera*.
- Nganta Ginting Teknik Komputer Politeknik Unggulan Cipta Mandiri Jl Bambu, J. (2024). Sistem Pendukung Keputusan Penerimaan Beasiswa Dengan Metode Simple Additive Weight Berbasis Website. In *Durian, Kec. Medan Tim* (Issue 1). <http://ejournal.stmik-time.ac.id>
- Nisa, H., Daud, M., & Retno, S. (2025). Evaluasi kinerja K-Medoids Clustering Daerah Produktivitas Panen Padi Di Kabupaten Bireuen. *Journal Of Information Systems And Informatics Engineering*, 9(2). <https://doi.org/10.35145/joisie.v9i2.4965>
- Prasetyo, S., & Dewayanto, T. (2024). Penerapan Machine Learning, Deep Learning, Dan Data Mining Dalam Deteksi Kecurangan Laporan Keuangan. *Diponegoro Journal Of Accounting*, 13(3), 1–12. <http://ejournal-s1.undip.ac.id/index.php/accounting>
- Rozali, C., Zein, A., & Farizy, S. (2023). *Penerapan Analytic Hierarchy Process (AHP) untuk Pemilihan Penerimaan Karyawan Baru*. <https://doi.org/10.55903/jitu.v1i2.153>
- Septianto, M. A., Faqih, A., & Rinaldi, A. R. (2025). Klasterisasi data Produksi Pertanian Di Kabupaten Cirebon Dengan Algoritma K-Means. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(2). <https://doi.org/10.23960/jitet.v13i2.6174>
- Shalaisya Azzahra Al-Meyda Hermawan, Yayan Nuryanto, & Agus Taryana. (2025). Efektivitas Program Ketahanan Pangan Berupa Bantuan Pemberian Bibit Tanaman Padi di Desa Bojongkoneng. *JANE (Jurnal Administrasi Negara)*.



- Sufriadi Dedi, & Hamid. A. (2021). Effisiensi Penggunaan Faktor-Faktor Produksi Usaha Tani Padi Sawah di Kabupaten Aceh Besar (Studi Kasus di Kecamatan Indapuri). *Jurnal Pendidikan Tambusai*, 5.
- Suraya, G. R., & Wijayanto, A. W. (2022). Comparison of Hierarchical Clustering, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting. *Indonesian Journal of Statistics and Its Applications*, 6(2), 180–201. <https://doi.org/10.29244/ijsa.v6i2p180-201>
- Ulfatul Syahara, Esty Kurniawati, Mario Putra Suhana, Rika Anggraini, & Falmi Yandri. (2024). Penerapan Metode AHC (Agglomerative Hierarchical Clustering) untuk Klasifikasi Habitat Bentik di Desa pengudang, Kabupaten Bintan. *INSOLOGI: Jurnal Sains Dan Teknologi*, 3(3), 306–314. <https://doi.org/10.55123/insologi.v3i3.3547>



## LAMPIRAN

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import davies_bouldin_score, silhouette_score,
calinski_harabasz_score
from scipy.cluster.hierarchy import dendrogram, linkage
from google.colab import files
from scipy.cluster.hierarchy import fcluster
import geopandas as gpd
import matplotlib.pyplot as plt

print(" Library berhasil diimpor")

df = pd.read_excel("Data Skripsi.xlsx")

print("\n Kolom dataset:")
print(df.columns.tolist())

features = [
    'Produktivitas Tanaman Padi (Ku/ha)',
    'Rekap Produksi Padi (Ton)',
    'Luas Panen Tanaman Padi (Ha)',
    'Luas Baku Sawah(Ha)'
]

# pastikan numerik
for col in features:
    df[col] = pd.to_numeric(df[col], errors='coerce')

jumlah_awal = len(df)
df = df.dropna(subset=features)
jumlah_akhir = len(df)

print("\n PRA-PEMROSESAN DATA")
print(f"Jumlah data awal : {jumlah_awal}")
print(f"Jumlah data akhir : {jumlah_akhir}")
print(f>Data terhapus : {jumlah_awal - jumlah_akhir}")

df = (
    df
    .groupby('Kabupaten/Kota')[features]
    .mean()
    .reset_index()
    # boleh mean (paling umum)
)

print("\n Data setelah agregasi per Kabupaten/Kota:")
print(f"Jumlah wilayah unik: {len(df)}")
display(df.head())

print("\n EDA - Statistik Deskriptif")
display(df[features].describe().T)

# Korelasi
plt.figure(figsize=(6,4))
sns.heatmap(df[features].corr(), annot=True, cmap="coolwarm")
plt.title("Korelasi Antar Variabel")
plt.tight_layout()
plt.show()

# Distribusi tiap fitur
for col in features:
    plt.figure(figsize=(6,3))
    sns.histplot(df[col], kde=True)
    plt.title(f"Distribusi {col}")
```

```
plt.xlabel(col)
plt.ylabel("Frekuensi")
plt.tight_layout()
plt.show()

# Scatter Produksi vs Produktivitas
plt.figure(figsize=(6,4))
sns.scatterplot(
    data=df,
    x='Produktivitas Tanaman Padi (Ku/ha)',
    y='Rekap Produksi Padi (Ton)'
)
plt.title("Hubungan Produktivitas dan Produksi Padi")
plt.tight_layout()
plt.show()

AHP = np.array([
    [1, 1/3, 3, 5], # Produktivitas
    [3, 1, 5, 7], # Produksi (PRIORITAS)
    [1/3, 1/5, 1, 3], # Luas Panen
    [1/5, 1/7, 1/3, 1] # Luas Baku Sawah
])

# Normalisasi matriks
norm_AHP = AHP / AHP.sum(axis=0)
weights = norm_AHP.mean(axis=1)

# Hitung Consistency Ratio (CR)
eigvals = np.linalg.eigvals(AHP)
lambda_max = np.max(eigvals.real)
n = AHP.shape[0]
CI = (lambda_max - n) / (n - 1)
RI = 0.90 # n = 4
CR = CI / RI

print("\nHASIL PEMBOBOTAN AHP - FOKUS PRODUKSI")
for f, w in zip(features, weights):
    print(f"{f:45s} : {w:.4f}")

print(f"\nJumlah bobot : {weights.sum():.4f}")
print(f"Consistency Ratio (CR) : {CR:.4f}")

scaler = MinMaxScaler()
X_norm = scaler.fit_transform(df[features])

print("\nContoh hasil normalisasi (tanpa bobot):")
display(pd.DataFrame(X_norm, columns=features).head())

X_weighted = X_norm * weights

print("\nContoh hasil normalisasi + bobot:")
display(pd.DataFrame(X_weighted, columns=features).head())

Z = linkage(X_weighted, method='ward')

plt.figure(figsize=(25,12))
dendrogram(Z)
plt.title("Dendrogram - Fokus Produksi (Ward + Bobot AHP)")
plt.xlabel("Objek")
plt.ylabel("Jarak")
plt.show()

n_clusters = 2
model = AgglomerativeClustering(n_clusters=n_clusters, linkage='ward')
labels = model.fit_predict(X_weighted)
df['Cluster'] = labels

print(f"\n🌀 Jumlah klaster yang digunakan: {n_clusters}")

n_clusters = 2 # ganti sesuai kebutuhan (2, 4, 6, dst.)
```

```

labels = fcluster(Z, t=n_clusters, criterion='maxclust')

df['Cluster'] = labels

print(f"\n✂ Jumlah klaster (fcluster): {len(np.unique(labels))}")

df_cluster = df[['Kabupaten/Kota', 'Cluster']].copy()

df_cluster['Kabupaten_Kota'] = (
    df_cluster['Kabupaten/Kota']
    .str.upper()
    .str.strip()
)

nama_mapping = {
    # Kepulauan Nias
    "NIAS": "NIAS",
    "NIAS BARAT": "NIAS BARAT",
    "NIAS SELATAN": "NIAS SELATAN",
    "NIAS UTARA": "NIAS UTARA",
    "GUNUNGSITOLI": "GUNUNGSITOLI",

    # Pantai Timur
    "LANGKAT": "LANGKAT",
    "DELI SERDANG": "DELI SERDANG",
    "SERDANG BEDAGAI": "SERDANG BEDAGAI",
    "BATU BARA": "BATU BARA",
    "ASAHAN": "ASAHAN",
    "LABUHAN BATU": "LABUHAN BATU",
    "LABUHAN BATU UTARA": "LABUHAN BATU UTARA",
    "LABUHAN BATU SELATAN": "LABUHAN BATU SELATAN",

    # Kota
    "MEDAN": "MEDAN",
    "BINJAI": "BINJAI",
    "SIBOLGA": "SIBOLGA",
    "TEBING TINGGI": "TEBING TINGGI",
    "PEMATANG SIANTAR": "PEMATANGSIANTAR",
    "PADANGSIDIMPUAN": "PADANGSIDIMPUAN",
    "TANJUNG BALAI": "TANJUNG BALAI",

    # Pantai Barat
    "TAPANULI TENGAH": "TAPANULI TENGAH",
    "TAPANULI UTARA": "TAPANULI UTARA",
    "TAPANULI SELATAN": "TAPANULI SELATAN",
    "MANDAILING NATAL": "MANDAILING NATAL",

    # Pegunungan
    "HUMBANG HASUNDUTAN": "HUMBANG HASUNDUTAN",
    "PAKPAK BHARAT": "PAKPAK BHARAT",
    "KARO": "KARO",
    "DAIRI": "DAIRI",
    "SIMALUNGUN": "SIMALUNGUN",

    # Penyesuaian ejaan GADM
    "LABUHAN BATU": "LABUHANBATU",
    "LABUHAN BATU UTARA": "LABUHANBATU UTARA",
    "LABUHAN BATU SELATAN": "LABUHANBATU SELATAN",
    "PEMATANG SIANTAR": "PEMATANGSIANTAR",
}

df_cluster['Kabupaten_Kota'] = (
    df_cluster['Kabupaten/Kota']
    .str.upper()
    .str.replace('KOTA ', '', regex=False)
    .str.replace('KABUPATEN ', '', regex=False)
    .str.strip()
)

```

```
# Mapping ke nama GADM
df_cluster['NAME_2'] = df_cluster['Kabupaten_Kota'].replace(nama_mapping)

print("df_cluster berhasil dibuat")
display(df_cluster.head())

dbi = davies_bouldin_score(X_weighted, labels)
si = silhouette_score(X_weighted, labels)
chi = calinski_harabasz_score(X_weighted, labels)

print("\nEVALUASI KLASSTER")
print(f"Davies-Bouldin Index : {dbi:.4f}")
print(f"Silhouette Score : {si:.4f}")
print(f"Calinski-Harabasz Index : {chi:.4f}")

print("\nRata-rata tiap klaster:")
display(df.groupby('Cluster')[features].mean())

print("\nDAFTAR WILAYAH PADA SETIAP KLASSTER")

for k in sorted(df['Cluster'].unique()):
    print("\n" + "="*45)
    print(f" Klaster {k}")
    print("="*45)

    wilayah = df[df['Cluster'] == k][
        ['Kabupaten/Kota',
         'Rekap Produksi Padi (Ton)',
         'Produktivitas Tanaman Padi (Ku/ha)',
         'Luas Panen Tanaman Padi (Ha)',
         'Luas Baku Sawah(Ha)']
    ].sort_values(
        by='Rekap Produksi Padi (Ton)',
        ascending=False
    )

    display(wilayah.reset_index(drop=True))

gdf_map = gpd.read_file("gadm36_IDN_2.shp")

print("Total wilayah Indonesia:", len(gdf_map))
print("Nama provinsi unik:", gdf_map["NAME_1"].unique())

gdf_map = gdf_map[
    gdf_map["NAME_1"].str.upper() == "SUMATERA UTARA"
]

print("Jumlah wilayah Sumatera Utara:", len(gdf_map))

gdf_map = gdf_map[gdf_map.geometry.notnull()]
gdf_map["geometry"] = gdf_map["geometry"].buffer(0)

print("Geometry valid:", gdf_map.is_valid.all())

gdf_map["NAME_2"] = (
    gdf_map["NAME_2"]
    .str.upper()
    .str.replace("KOTA ", "", regex=False)
    .str.replace("KABUPATEN ", "", regex=False)
    .str.strip()
)

df_cluster["NAME_2"] = (
    df_cluster["NAME_2"]
    .str.upper()
    .str.replace("KOTA ", "", regex=False)
    .str.replace("KABUPATEN ", "", regex=False)
    .str.strip()
)

print("Tidak cocok:",
```



```

        set(df_cluster["NAME_2"]) - set(gdf_map["NAME_2"]))

# Mapping khusus perbedaan GADM 3.6
nama_fix = {
    "LABUHAN BATU UTARA": "LABUHANBATU UTARA",
    "TEBING TINGGI": "TEBINGTINGGI",
    "TANJUNG BALAI": "TANJUNGBALAI"
}

df_cluster["NAME_2"] = df_cluster["NAME_2"].replace(nama_fix)

# ulangi cek
print(
    "Tidak cocok setelah fix:",
    set(df_cluster["NAME_2"]) - set(gdf_map["NAME_2"])
)

final_fix = {
    # Labuhan Batu (gabung kata)
    "LABUHAN BATU": "LABUHANBATU",
    "LABUAN BATU UTARA": "LABUHANBATU UTARA",
    "LABUHAN BATU SELATAN": "LABUHANBATU SELATAN",

    # Pakpak
    "PAKPAK BHARAT": "PAKPAK BARAT"
}

df_cluster["NAME_2"] = df_cluster["NAME_2"].replace(final_fix)

print(
    "Tidak cocok FINAL:",
    set(df_cluster["NAME_2"]) - set(gdf_map["NAME_2"])
)

gdf_merge = gdf_map.merge(
    df_cluster[["NAME_2", "Cluster"]],
    on="NAME_2",
    how="left"
)

print("Wilayah setelah merge:", len(gdf_merge))

# Hitung centroid (untuk label)
gdf_merge["centroid"] = gdf_merge.geometry.centroid

from matplotlib.colors import ListedColormap

fig, ax = plt.subplots(figsize=(10, 12))

# Colormap manual (AMAN UNTUK SKRIPSI)
cmap_cluster = ListedColormap([
    "#66c2a5", # Klaster 1 (hijau)
    "#fc8d62" # Klaster 2 (oranye)
])

gdf_merge.plot(
    column="Cluster",
    categorical=True,
    cmap=cmap_cluster,
    linewidth=0.5,
    edgecolor="black",
    legend=True,
    ax=ax
)

# Tambahkan label nama kab/kota
for idx, row in gdf_merge.iterrows():
    ax.text(
        row["centroid"].x,

```

PAPER NAME

NORA IRAWANI SIREGAR\_File Skripsi\_O  
HhdWUotNQOUTL7DKTDuJL9eDZzMq8r  
MWEKp5GQn.docx

AUTHOR

NORA IRAWANI SIREGAR\_

WORD COUNT

8583 Words

CHARACTER COUNT

57282 Characters

PAGE COUNT

58 Pages

FILE SIZE

935.8KB

SUBMISSION DATE

Feb 9, 2026 2:50 PM GMT+7

REPORT DATE

Feb 9, 2026 2:52 PM GMT+7

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 5% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 6% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Cited material
- Abstract
- Small Matches (Less than 15 words)



# UNIVERSITAS MEDAN AREA

## FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate / Jalan Gedung PBSI, Medan 20223  
Kampus II : Jalan Sei Serayu Nomor 70 A / Jalan Setia Budi Nomor 79 B, Medan 20112 Telepon : (061) 8225602, 8201994  
Fax : (061) 8226331 HP : 0811 607 259 website: www.uma.ac.id Email : univ\_medanarea@uma.ac.id

Nomor : 2569/FT/01.10/X/2025  
Lampiran : -  
Hal : Pembimbing Tugas Akhir

14 Oktober 2025

Yth. Pembimbing Tugas Akhir  
**Nanda Novita S.Kom, M.Kom ( Sebagai Pembimbing )**  
di Tempat

Dengan hormat, sehubungan telah dipenuhinya persyaratan untuk memperoleh Tugas Akhir dari mahasiswa atas :

Nama : NORA IRAWANI SIREGAR  
NIM : 228160013  
Jurusan : TEKNIK INFORMATIKA

Maka dengan hormat kami mengharapkan kesediaan saudara :

**Nanda Novita S.Kom, M.Kom ( Sebagai Pembimbing )**

Adapun Tugas Akhir Skripsi berjudul :

**Pembobotan Fitur Pada Hierarchical Clustering untuk Segmentasi Produksi Padi di Sumatera Utara**

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

Dekan,



Dr Eng. Supriatno.ST, MT.





# UNIVERSITAS MEDAN AREA

## FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, Medan, 20223  
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 42402994, Medan, 20122  
Website: [www.teknik.uma.ac.id](http://www.teknik.uma.ac.id) E-mail: [univ\\_medanarea@uma.ac.id](mailto:univ_medanarea@uma.ac.id)

Nomor : 351/FT.6/01.10/XI/2025  
Lamp : -  
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

27 November 2025

Yth. Wakil Rektor Bidang Mutu Sumber Daya Manusia dan Perekonomian  
Jln. Kolam No.1  
Di  
Medan

Dengan hormat, kami mohon kesediaan bapak kiranya berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

| NO | N A M A              | N P M     | PRODI              |
|----|----------------------|-----------|--------------------|
| 1  | Nora Irawani Siregar | 228160013 | Teknik Informatika |

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul :

**Pembobotan Fitur Pada *Hierarchical Clustering* untuk Segmentasi Produksi Padi di Sumatera Utara.**

Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya Surat ini.

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.

Dekan

  
Priatno, ST, MT

Tembusan :

1. Ka. BPMPP

2. Mahasiswa

3. File

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 1/7/26

Access From ([repository.uma.ac.id](http://repository.uma.ac.id))1/7/26





# UNIVERSITAS MEDAN AREA

Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, 7366878, 7364348 📠 (061) 7368012 Medan 20223  
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602 📠 (061) 8226331 Medan 20122  
Website: [www.uma.ac.id](http://www.uma.ac.id) E-Mail: [univ\\_medanarea@uma.ac.id](mailto:univ_medanarea@uma.ac.id)

## SURAT KETERANGAN SELESAI PENELITIAN

Nomor : 57/UMA/B/01.7/I/2026

Yang bertanda tangan di bawah ini :

Nama : Dr. Ir. Rahmad Syah, M.Kom, IPM, ASEAN Eng, APEC Eng  
Jabatan : Wakil Rektor Bidang Mutu Sumber Daya dan Perekonomian  
NIDN : 0105058804

Dengan ini menerangkan bahwa mahasiswa yang Namanya tercantum di bawah ini :

Nama : Nora Irawani Siregar  
NPM : 228160013  
Program Studi : Teknik Informatika  
Fakultas : Teknik  
Status : (Mahasiswa / Dosen / Peneliti)

Telah melaksanakan dan menyelesaikan riset (penelitian) di lingkungan Universitas Medan Area dengan rincian sebagai berikut:

Judul Penelitian : Permohonan Fitur Pada *Hierarchical Clustering* untuk Segmentasi Produksi Padi di Sumatera Utara  
Tujuan Penelitian : Menentukan bobot kepentingan antara fitur produksi dan produktivitas, menerapkan metode *Hierarchical Clustering* untuk mengelompokkan wilayah, dan menghasilkan peta segmentasi wilayah produksi padi di Sumatera Utara  
Lokasi Penelitian : Laboratorium Komputer Program Studi Teknik Informatika Universitas Medan Area  
Waktu Pelaksanaan : November - Desember 2025  
Hasil Penelitian : Berdasarkan analisis fitur produksi dan produktivitas padi menggunakan Ward method, jumlah 2 klaster terpilih sebagai yang paling optimal karena memiliki nilai SI tertinggi (0,7182) dan DBI terendah (0,2266). Meskipun fitur produktivitas menunjukkan hasil stabil pada 4 klaster.

Berdasarkan laporan hasil riset dan verifikasi data yang kami terima, yang bersangkutan telah menyelesaikan seluruh rangkaian kegiatan penelitiannya dengan baik.

Demikian surat ini diterbitkan untuk dapat digunakan seperlunya.

Medan, 08 Januari 2026  
An. Rektor  
Wakil Rektor Bidang Mutu Sumber  
Daya dan Perekonomian,



Dr. Ir. Rahmad Syah, M.Kom,  
IPM, ASEAN Eng, APEC Eng

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area