

***CLUSTERING JENIS SAYURAN DI DAERAH DESA SEMPAJAYA  
DENGAN ALGORITMA GAUSIAN MIXTURE MODEL***

**SKRIPSI**

**DISUSUN OLEH**

**HAFIZAN SYAFIK**  
**2109010075**



**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2025**

**CLUSTERING JENIS SAYURAN DI DAERAH DESA SEMPAJAYA  
DENGAN ALGORITMA *GAUSSIAN MIXTURE MODEL***

**SKRIPSI**

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

**HAFIZAN SYAFIK  
NPM. ISI 21019010075**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI  
INFORMASI UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA  
MEDAN  
2025**

## LEMBAR PENGESAHAN

Judul Skripsi : CLUSTERING JENIS SAYURAN DI DAERAH DESA  
SEMPAJAYA MENGGUNAKAN ALGORITMA  
GAUSSIAN MIXTURE MODEL

Nama Mahasiswa : HAFIZAN SYAFIK

NPM : 2109010075

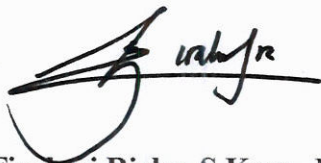
Program Studi : SISTEM INFORMASI

Menyetujui  
Komisi Pembimbing



(Mulkan Azhari, S.kom, M.kom)  
NIDN. 0108129402

Ketua Program Studi



Dr. Firahmi Rizky, S.Kom., M.Kom  
NIDN. 0116079201

Dekan



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

## **PERNYATAAN ORISINALITAS**

### **CLUSTERING JENIS SAYURAN DI DAERAH DESA SEMPAJAYA DENGAN ALGORITMA GAUSSIAN MIXTURE MODEL**

#### **SKRIPSI**

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

Medan, 29 Agustus 2025

Yang membuat pernyataan



Hafizan Syafik

NPM. 2109010075

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

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

|               |                    |
|---------------|--------------------|
| Nama          | : Hafizan Syafik   |
| NPM           | : 2109010075       |
| Program Studi | : Sistem Informasi |
| Karya Ilmiah  | : Skripsi          |

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

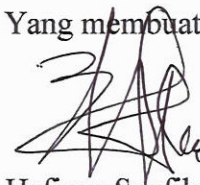
**CLUSTERING JENIS SAYURAN DI DAERAH DESA SEMPAJAYA  
DENGAN ALGORITMA GAUSSIAN MIXTURE MODEL**

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, 29 Agustus 2025

Yang membuat pernyataan



Hafizan Syafik

NPM. 210901007

## **`RIWAYAT HIDUP**

### **DATA PRIBADI**

|                          |                               |
|--------------------------|-------------------------------|
| Nama Lengkap             | : Hafizan Syafik              |
| Tempat dan Tanggal Lahir | : Berastagi 05-06-2003        |
| Alamat Rumah             | : Lembah Kemakmuran Berastagi |
| Telepon/Faks/HP          | : 082180926571                |
| E-mail                   | : Hafizsyafik014@gmail.com    |
| Instansi Tempat Kerja    | : -                           |
| Alamat Kantor            | : -                           |

### **DATA PENDIDIKAN**

|     |                             |             |
|-----|-----------------------------|-------------|
| SD  | : SD SWASTA AL WASLHIYAH    | TAMAT: 2015 |
| SMP | : SMPN 1 BERASTAGI          | TAMAT: 2018 |
| SMA | : MADRASAH ALIYAH KABANJAHE | TAMAT: 2021 |

## KATA PENGANTAR



Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “*Clustering Jenis Sayuran di Daerah Sempajaya dengan Algoritma Gaussian Mixture Model*”. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan program Sarjana Strata satu ( S1 ) program studi sistem informasi fakultas ilmu komputer dan teknologi informasi ( FIKTI )

Dalam penyusunan skripsi ini, penulis menyadari bahwa banyak tantangan yang dihadapi, baik dari segi pengumpulan data, analisis, maupun penyusunan laporan. Namun berkat doa, dukungan, bimbingan, dan bantuan dari berbagai pihak, skripsi ini dapat terselesaikan dengan baik. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Agussani, M.AP., 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 Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer Dan Teknologi Informasi
4. Bapak Halim Maulana, S.T., M.Kom., MTA Selaku Wakil Dekan 1 Fakultas Ilmu Komputer Dan Teknologi Informasi

5. Bapak Mulkan Azhari, S.Kom, M.Kom selaku dosen pembimbing yang telat meluangkan waktu untuk membimbing penulis selama penyusunan skripsi
6. Pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang terdalem kepada kedua orang tua tercinta. Rahmayana Nasution dan Syaiful Dengan penuh kesabaran, kasih sayang, doa, dan dukungan tanpa henti, Ayah dan Ibu telah menjadi sumber semangat terbesar dalam setiap langkah penulis. Semua pengorbanan, kerja keras, dan cinta Skripsi ini penulis persembahkan sebagai bentuk kecil dari rasa hormat dan terima kasih yang tak terhingga kepada Ayah dan Ibu. Semoga karya sederhana ini dapat menjadi kebanggaan serta doa penulis agar selalu dapat membahagiakan dan membalas sedikit dari semua kebaikan yang telah diberikan. kasih yang telah diberikan tidak akan pernah bisa terbalas dengan apapun.
7. Teman seperjuangan dan seangkatan, Aldi Subari, Aldi Pradana, Dimas AryaPrayoga, DimasWijayanto, MikoAryandi, RioFerdinand, Rizaldi, Ronggo Hernanda, dan Setyo Harry Nugroho yang selalu memberikan motivasi dan dukungan serta semangat selama penyusunan skripsi ini dimulai hingga selesai.



## **CLUSTERING JENIS SAYURAN DI DAERAH DESA SEMPAJAYA DENGAN ALGORITMA GAUSSIAN MIXTURE MODEL**

### **ABSTRAK**

Sektor pertanian di Desa Sempajaya, Kabupaten Karo, memiliki potensi besar dalam menghasilkan berbagai jenis sayuran yang menjadi sumber utama pemenuhan kebutuhan gizi masyarakat sekaligus penyokong perekonomian lokal. Namun, pengelolaan lahan dan pemetaan jenis sayuran unggulan masih menghadapi kendala karena belum adanya sistem pengelompokan data yang akurat.

Penelitian ini bertujuan untuk mengelompokkan jenis sayuran di Desa Sempajaya menggunakan algoritma Gaussian Mixture Model (GMM) sebagai metode clustering yang mampu menangani distribusi data yang kompleks. Metode penelitian dilakukan melalui beberapa tahap, yaitu pengumpulan data (observasi, wawancara, dan dokumentasi), pra-pemrosesan data, analisis faktor menggunakan diagram Fishbone untuk menentukan atribut relevan, serta implementasi algoritma GMM dengan pendekatan Expectation-Maximization (EM). Data yang digunakan mencakup enam komoditas utama, yaitu cabai, tomat, sawi, wortel, terung, dan buncis, dengan variabel meliputi ukuran, berat, warna, serta luas lahan.

Hasil penelitian menunjukkan bahwa algoritma GMM berhasil mengelompokkan data sayuran ke dalam tiga kategori kluster produksi, yaitu rendah, sedang, dan tinggi, dengan visualisasi hasil clustering yang lebih representatif dibandingkan metode konvensional. Sistem ini mampu memberikan informasi potensi sayuran unggulan di setiap wilayah Desa Sempajaya, yang dapat dimanfaatkan oleh petani, masyarakat, maupun pemerintah daerah dalam pengambilan keputusan strategis terkait diversifikasi, distribusi, dan pengembangan pertanian berkelanjutan.

**Kata kunci:** Clustering, Gaussian Mixture Model, Sayuran, Sempajaya, Data Mining

## CLUSTERING OF VEGETABLE TYPES IN SEMPAJAYA VILLAGE USING THE GAUSSIAN MIXTURE MODEL ALGORITHM

### ABSTRACT

*The agricultural sector in Sempajaya Village, Karo Regency, holds significant potential in producing various types of vegetables that serve as a primary source of nutrition as well as a contributor to the local economy. However, land management and the mapping of superior vegetable types still face challenges due to the absence of an accurate data clustering system. This research aims to cluster vegetable types in Sempajaya Village using the Gaussian Mixture Model (GMM) algorithm, a clustering method capable of handling complex data distributions.*

*The research methodology consists of several stages, including data collection (observation, interviews, and documentation), data preprocessing, factor analysis using the Fishbone diagram to identify relevant attributes, and the implementation of GMM with the Expectation-Maximization (EM) approach. The dataset involves six main vegetable commodities, namely chili, tomato, mustard greens, carrot, eggplant, and beans, with variables such as size, weight, color, and cultivated land area.*

*The results show that the GMM algorithm successfully clustered vegetable data into three production categories: low, medium, and high, providing a more representative visualization compared to conventional methods. This system is able to present valuable insights on superior vegetable potentials across Sempajaya Village, which can be utilized by farmers, communities, and local governments to support strategic decision-making related to diversification, distribution, and sustainable agricultural development.*

**Keywords:** *Clustering, Gaussian Mixture Model, Vegetables, Sempajaya, Data Mining*

## DAFTAR ISI

|                                                                        |             |
|------------------------------------------------------------------------|-------------|
| <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>xii</b>  |
| <b>BAB I PENDAHULUAN .....</b>                                         | <b>1</b>    |
| 1.1 Latar Belakang Masalah.....                                        | 1           |
| 1.2 Rumusan Masalah .....                                              | 3           |
| 1.3 Batasan Masalah .....                                              | 4           |
| 1.4 Tujuan Penelitian .....                                            | 5           |
| 1.5 Manfaat Penelitian .....                                           | 5           |
| <b>BAB II LANDASAN TEORI .....</b>                                     | <b>7</b>    |
| 2.1 Clustering.....                                                    | 7           |
| 2.2 Sejarah Desa Sempajaya .....                                       | 8           |
| 2.3 Jenis Sayuran .....                                                | 10          |
| 2.4 Komoditi Jenis Sayuran .....                                       | 12          |
| 2.5 Lahan Pertanian.....                                               | 13          |
| 2.6 Segmentasi Data.....                                               | 15          |
| 2.7 Konsep Segmentasi Data .....                                       | 15          |
| 2.8 Algoritma Gaussian Mixture Model.....                              | 17          |
| 2.8.1 Pengertian Algoritma <i>Gaussian Mixture Model</i> .....         | 17          |
| 2.8.2 Langkah Langkah dan Rumus Algoritma Gaussian Mixture Model ..... | 18          |
| 2.8.3 Rumus Gaussian Mixture Model.....                                | 19          |
| 2.9 Pengertian Python .....                                            | 21          |
| 2.10 Software ( Perangkat Lunak ) .....                                | 21          |
| 2.10.1 Pengertian Website .....                                        | 21          |
| 2.10.2 Pengertian Database.....                                        | 22          |
| 2.11 Flowchart .....                                                   | 23          |

|                                                          |           |
|----------------------------------------------------------|-----------|
| 2.12 Pengujian Blackbox Texting.....                     | 25        |
| 2.13 Literatur Riview .....                              | 26        |
| <b>BAB III METODOLOGI PENELITIAN .....</b>               | <b>30</b> |
| 3.1 Jenis Penelitian.....                                | 30        |
| 3.2 Tempat Dan Waktu Penelitian .....                    | 31        |
| 3.2.1 Tempat Penelitian .....                            | 31        |
| 3.2.2 Waktu Penelitian .....                             | 31        |
| 3.3 Teknik Pengumpulan Data.....                         | 32        |
| 3.4 Teknik Analisis Data.....                            | 32        |
| 3.5 Kerangka Berpikir.....                               | 33        |
| 3.6 Variabel Penelitian .....                            | 34        |
| 3.7 Desain Penelitian.....                               | 35        |
| 3.8 Diagram Fishbone sebagai Analisis Awal Fitur .....   | 35        |
| 3.9 Diagram Fishbone.....                                | 36        |
| 3.10 Sumber Dan Pengumpulan Data.....                    | 36        |
| 3.11 Pra – pemrosesan Data.....                          | 37        |
| 3.12 Implementasi Algoritma Gaussian Mixture Model ..... | 37        |
| <b>BAB IV HASIL DAN UJI COBA .....</b>                   | <b>39</b> |
| 4.1 Hasil .....                                          | 39        |
| 4.2 Pembahasan.....                                      | 42        |
| 4.3 Hasil Uji Coba.....                                  | 54        |
| 4.4 Kekurangan Aplikasi.....                             | 55        |
| <b>BAB V PENUTUP.....</b>                                | <b>56</b> |
| 5.1 Kesimpulan .....                                     | 56        |
| 5.2 Saran .....                                          | 56        |
| <b>DAFTAR PUSTAKA.....</b>                               | <b>57</b> |

## DAFTAR TABEL

|                                     |    |
|-------------------------------------|----|
| Tabel 2.1 Struktur Desa .....       | 9  |
| Table 2.2 Gambar Flowchart .....    | 24 |
| Tabel 2.3 Struktur Desa .....       | 29 |
| Tabel 3.1 Waktu Penelitian .....    | 31 |
| Tabel 3.2 Variable Penelitian ..... | 34 |

## DAFTAR GAMBAR

|                                         |    |
|-----------------------------------------|----|
| Gambar 2.1 Struktur Desa .....          | 9  |
| Gambar 2.2 Gambar Flowchart .....       | 24 |
| Gambar 2.3 Struktur Desa .....          | 29 |
| Gambar 3.1 Alur Metode Penelitian ..... | 30 |
| Gambar 3.2 Kerangka Berpikir .....      | 33 |
| Gambar 3.3 Diagram Fishbone.....        | 36 |

## **BAB I PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Sektor pertanian memainkan peran penting dalam menyediakan makanan bagi masyarakat Indonesia, termasuk di Sempajaya. Pertanian Sempajaya memiliki potensi besar untuk menghasilkan berbagai jenis sayuran. Banyak jenis sayuran yang dihasilkan menjadi keuntungan sekaligus tantangan yang terakhir terutama berkaitan dengan manajemen data dan distribusi hasil panen. Namun, pengambilan keputusan strategis seperti manajemen lahan, distribusi hasil tani, dan pengembangan pasar dapat terhambat oleh kurangnya pemetaan yang jelas tentang jenis sayuran yang dominan di setiap wilayah Sempajaya. Menurut (Nugraha, 2022) menyatakan bahwa kurangnya pemetaan yang jelas mengenai jenis sayuran yang dominan di setiap wilayah dapat menghambat pengambilan keputusan strategis.

Sempajaya Berastagi merupakan salah satu wilayah di Provinsi Sumatera Utara yang memiliki potensi pertanian dan perkebunan yang sangat luas. Tanaman pangan seperti sayuran dan umbi. sangat penting bagi pemenuhan kebutuhan gizi masyarakat. Hasil produksi dari pertanian dan perkebunan ini tidak hanya memenuhi kebutuhan makanan sehari-hari, tetapi juga menjadi sumber pendapatan bagi masyarakat. Seiring dengan meningkatnya kesadaran akan pentingnya gizi, masyarakat semakin memanfaatkan hasil pertanian untuk memenuhi kebutuhan mereka, baik untuk konsumsi pribadi maupun untuk dijual di pasar. Menurut (H.Wijayanto,2021) analisis kelayakan finansial usahatani sayuran menunjukkan potensi yang signifikan dalam meningkatkan pendapatan petani, yang menunjukkan bahwa sektor ini dapat berkontribusi pada perekonomian lokal.

Sayuran adalah bahan makanan yang kaya akan serat, protein, dan vitamin, yang sangat penting bagi manusia. Setiap jenis sayuran memiliki manfaat dan rasa sendiri, sehingga berkontribusi pada asupan nutrisi yang seimbang. Karena manfaat kesehatannya, masyarakat semakin menyukai produksi sayuran yang ditawarkan serta nilai ekonomis yang tinggi. Permintaan produk sayuran di pasar domestik terus meningkat, mendorong masyarakat untuk membudidayakan sayuran secara maksimal pada lahan yang tersedia.. Menurut (Pradjasasmitha et al., 2024). Berbagai faktor biologis, termasuk jenis tanaman, teknik pertanian, dan kondisi lingkungan, dapat memengaruhi kualitas produk sayuran. Petani lokal di pasar tradisional menggunakan teknik pertanian yang lebih organik dan lingkungan yang lebih terjaga dibandingkan dengan pemasok pasar saat ini. Dalam perspektif biologi, berbagai faktor biologis dan nonbiologis dapat memengaruhi kualitas produk sayuran, dan perbedaan antara pasar modern dan tradisional dalam perolehan, pengolahan, dan penyimpanan produk sayuran dapat memengaruhi kualitas produk sayuran yang dijual di kedua pasar tersebut. Akibatnya, penting untuk memilih produk sayuran yang berkualitas tinggi.

Namun, tantangan dalam penanaman sayuran di Sempajaya Berastagi tidak dapat diabaikan. Banyak petani yang mengalami kendala dalam pemilihan lahan yang sesuai dengan jenis sayuran yang ditanam. Ketidaksesuaian lahan dapat mengakibatkan kerugian, baik dari segi tenaga kerja maupun finansial. (Nugraha et al., 2022) menekankan bahwa pemilihan lahan yang tidak sesuai dengan jenis sayurannya dapat menimbulkan kerugian bagi petani tersebut. Oleh karena itu, evaluasi lahan menjadi langkah penting untuk meningkatkan kualitas produksi



sayuran.

Dengan menggunakan metode seperti *Gaussian Mixture Model* (GMM), petani dapat memperoleh rekomendasi mengenai jenis sayuran yang cocok untuk ditanam berdasarkan karakteristik lahan. Penelitian yang dilakukan oleh Wijayanto et al. (2021) menunjukkan bahwa metode GMM dapat meningkatkan produktivitas tanaman sekunder yang ditanam dan memperbaiki komoditas ekonomi di daerah tersebut, sehingga diharapkan dapat membantu petani membuat keputusan yang lebih baik. Penelitian ini bertujuan untuk mengidentifikasi komoditas potensial di setiap wilayah Sempajaya; tanaman dengan potensi di daerah tersebut akan terus dipertahankan dan produksinya ditingkatkan, sementara tanaman sayuran dengan produksi rendah akan menjadi prioritas untuk meningkatkan hasilnya..

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang di atas, rumusan masalah yang perlu diteliti adalah sebagai berikut. perlu diidentifikasi jenis sayuran unggul yang dapat ditemukan di daerah Sempajaya serta bagaimana distribusinya di wilayah tersebut. Hal ini penting untuk memahami potensi pertanian sayuran di daerah tersebut. penelitian ini juga akan mengkaji kendala-kendala yang dihadapi oleh petani dalam pengelolaan lahan dan produksi sayuran di Sempajaya, yang dapat mencakup faktor-faktor seperti cuaca, hama, dan akses terhadap teknologi pertanian akan dilakukan analisis untuk menentukan jenis sayuran apa saja yang dapat dikelompokkan menggunakan metode *clustering*, yang dapat membantu dalam pengelompokan sayuran berdasarkan karakteristik tertentu. Terakhir, penelitian ini akan mengeksplorasi bagaimana metode *Gaussian Mixture Model* (GMM) dapat

diterapkan untuk merekomendasikan jenis sayuran yang paling cocok untuk ditanam di Sempajaya, dengan mempertimbangkan faktor-faktor lingkungan dan kebutuhan pasar. Dengan menjawab rumusan masalah ini, diharapkan dapat memberikan wawasan yang lebih baik mengenai pengembangan pertanian sayuran di daerah Sempajaya.

### **1.3 Batasan Masalah**

Variabel Penelitian:

Penelitian ini akan membatasi fokus pada variabel-variabel berikut:

1. Penelitian akan dilakukan di beberapa wilayah di daerah Sempajaya
2. Jenis Sayuran Jenis sayuran yang akan dianalisis meliputi cabai, tomat, sawi, wortel, terung dan buncis
3. Produksi sayuran akan diukur dalam satuan kuantitas (misalnya, kilogram) yang dihasilkan per hektar.
4. Luas Tanaman Luas lahan yang digunakan untuk menanam masing-masing jenis sayuran akan dicatat dalam satuan hektar.
5. Komoditi yang Diteliti: Penelitian ini akan terbatas pada enam komoditi sayuran, yaitu: Cabai, Tomat, Sawi, Wortel, Terung dan Buncis.
6. Klasifikasi Hasil Panen Hasil panen dari masing-masing jenis sayuran akan dikelompokkan menjadi tiga kategori kluster berdasarkan tingkat produksi

Rendah : Produksi di bawah rata-rata yang ditentukan.

Sedang : Produksi mendekati rata-rata.

Tinggi : Produksi di atas rata-rata yang ditentukan.

#### **1.4 Tujuan Penelitian**

1. Untuk merancang sistem yang dapat mengelompokkan data jenis sayuran unggulan di daerah Sempajaya.
2. Untuk memperoleh hasil klasterisasi dari perhitungan menggunakan algoritme Gaussian Mixture Model. Menyediakan Dasar untuk Riset Lanjutan.
3. Rekomendasi untuk Diversifikasi Pertanian Memberikan rekomendasi berbasis data untuk diversifikasi jenis sayuran yang ditanam, berdasarkan hasil clustering yang menunjukkan potensi pasar dan keberlanjutan.

#### **1.5 Manfaat Penelitian**

Adapun manfaat dari penelitian tugas akhir ini yaitu ;

1. Penyajian Data yang Informatif: Penelitian ini dapat digunakan oleh Dinas Pertanian dan Ketahanan Pangan serta Badan Pusat Statistik Kabupaten Karo (Karo Regency) dalam penyajian data yang lebih informatif. Hal ini akan membantu mereka dalam memahami potensi pertanian, khususnya di bidang sayuran di daerah Sempajaya
2. Informasi Potensi Sayuran Unggulan: Sistem yang dihasilkan dari penelitian ini dapat dimanfaatkan oleh masyarakat untuk mengetahui potensi sayuran unggulan di daerah Sempajaya. Dengan informasi ini, masyarakat dapat lebih memahami jenis sayuran yang memiliki nilai ekonomi tinggi di daerah mereka.
3. Peningkatan Kebijakan Pertanian: Informasi yang diperoleh dari penelitian ini dapat digunakan oleh pemerintah daerah untuk merumuskan kebijakan

yang lebih baik. Dengan memahami potensi komoditas, pemerintah dapat fokus pada peningkatan daerah dengan komoditas yang memiliki potensi rendah, sambil tetap mempertahankan dan mengembangkan komoditas dengan potensi tinggi.

4. Dukungan untuk Pengembangan Pertanian Berkelanjutan: Penelitian ini juga dapat memberikan dasar bagi pengembangan program-program pertanian berkelanjutan, yang bertujuan untuk meningkatkan produktivitas dan keberlanjutan sektor pertanian di daerah Sempajaya.
5. Peningkatan Kualitas Hasil Pertanian: Dengan informasi yang lebih baik mengenai potensi sayuran, petani di daerah Sempajaya dapat menerapkan praktik budidaya yang lebih efektif, yang pada gilirannya dapat meningkatkan kualitas hasil pertanian dan kesejahteraan petani.

## BAB II

### LANDASAN TEORI

#### 2.1 Clustering

Klastering merupakan metode untuk membagi sekumpulan data menjadi beberapa kelompok berdasarkan tingkat kesamaan yang diharapkan. Dalam konteks data mining, clustering melibatkan pengorganisasian data atau objek ke dalam kelompok-kelompok (*cluster*) sehingga setiap kelompok terdiri dari data yang serupa satu sama lain dan berbeda dari objek di kelompok lain. Pengelompokan ini diperlukan karena volume data mentah yang sangat besar sering kali sulit dianalisis atau dipelajari secara langsung. Tujuan utama pengelompokan dalam skenario ini adalah untuk memperdalam pemahaman terhadap data serta mengevaluasi kualitasnya. (Saputra & Nataliani, 2021)

Clustering merupakan salah satu teknik yang berperan penting dalam analisis data karena memungkinkan pengelompokan objek berdasarkan kesamaan karakteristik tertentu. Di antara berbagai algoritma dalam bidang data mining, **K-Means** dikenal sebagai metode clustering yang didasarkan pada pengelompokan berdasarkan jarak antar data. Penelitian ini bertujuan untuk menganalisis penerapan teknik clustering menggunakan algoritma K-Means dalam konteks data mining melalui kajian literatur yang komprehensif. Analisis dilakukan dengan meninjau secara mendalam berbagai penelitian yang relevan dan mengevaluasi literatur terpilih berdasarkan kriteria tertentu, sehingga hasil kajian ini dapat digunakan untuk menjawab pertanyaan penelitian yang telah dirumuskan.

Tinjauan Pustaka Sistematis (*Systematic Literature Review/SLR*) adalah metode penelitian yang bertujuan untuk mengidentifikasi dan mengevaluasi hasil penelitian menggunakan teknik terbaik berdasarkan prosedur tertentu dari hasil perbandingan. Berdasarkan pemilihan literatur publikasi jurnal, Pola Jurnal-jurnal internasional seperti *Recognition*, *Knowledge-Based Systems*, *Applied Soft Computing*, dan *IEEE Access* dapat dijadikan sebagai rujukan utama dalam kajian yang berkaitan dengan algoritma **K-Means**. Berdasarkan hasil beberapa penelitian yang dimuat dalam jurnal-jurnal tersebut, diperoleh perbandingan metode perhitungan jarak yang digunakan dalam proses *clustering*. Dari hasil kajian tersebut diketahui bahwa **Euclidean Distance** memiliki keunggulan dalam hal akurasi dan efisiensi perhitungan jarak antar data. Oleh karena itu, metode ini dapat dijadikan sebagai pilihan utama dalam penerapan teori perhitungan jarak pada algoritma K-Means. (Sekar Setyaningtyas et al., 2022)

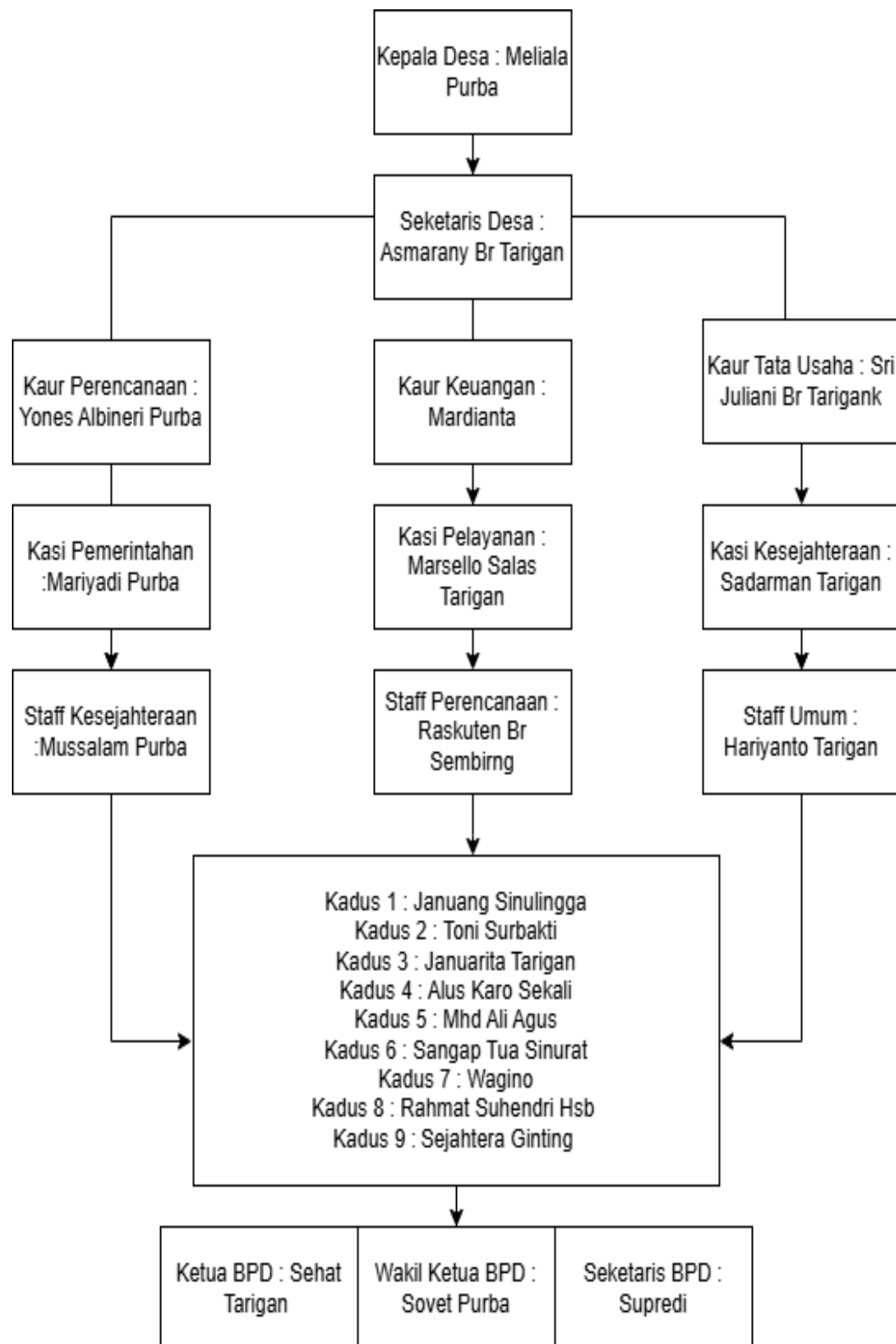
## 2.2 Sejarah Desa Sempajaya

Desa Sempajaya terletak di Kecamatan Berastagi, Kabupaten Karo, Sumatera Utara. Sejarah desa ini mencakup perkembangan pemerintahan dan masyarakat yang telah berlangsung selama bertahun-tahun. Desa Sempajaya, yang lebih dikenal dengan sebutan Desa Peceren, dipimpin oleh Kepala Desa Bapak Meliala Purba sejak tahun 2017 hingga saat ini. Berdasarkan hasil observasi peneliti, lokasi kantor Desa Sempajaya tergolong strategis dan mudah diakses oleh masyarakat. Bangunan kantor desa tersebut telah dirancang sesuai dengan kebutuhan aparatur pemerintahan desa. Kantor ini terdiri atas beberapa ruangan, antara lain ruang tunggu, ruang kepala desa, ruang operator desa, serta ruang baca yang berfungsi

sebagai perpustakaan kecil.

Selain itu, data yang diperoleh peneliti menunjukkan bahwa jumlah penduduk Desa Sempajaya mengalami peningkatan dari waktu ke waktu. Berdasarkan pengamatan lapangan, pemerintah desa terus berupaya melakukan pembenahan dengan melengkapi berbagai sarana dan prasarana guna memenuhi kebutuhan masyarakat. Upaya tersebut dilakukan untuk mewujudkan pembangunan desa yang lebih baik dan berkualitas.

Berhasilnya pembangunan di Desa Sempajaya sebagai aplikasi Pancasila tergantung pada partisipasi warga, serta sikap mental, tekad, ketaatan dan disiplin para Pemerintah Desa serta seluruh masyarakat Desa. Struktur perangkat desa biasanya terdiri dari kepala desa, sekretaris desa, dan berbagai bidang yang menangani urusan administrasi, pembangunan, dan pelayanan Masyarakat. struktur pemerintahan Desa Sempajaya dapat dilihat berikut ini.



**Gambar 2.1** Struktur Desa



## 2.3 Jenis Sayuran

Sayuran merupakan bahan pangan yang berasal dari tumbuhan dan memiliki kandungan air yang relatif tinggi. Umumnya, sayuran dikonsumsi dalam keadaan segar atau setelah melalui proses pengolahan sederhana. Terdapat beragam jenis sayuran yang berbeda dalam hal warna, rasa, aroma, dan teksturnya. Namun, secara umum, dari aspek kandungan gizinya, sayuran berfungsi sebagai sumber utama vitamin, mineral, dan serat yang penting bagi kesehatan tubuh. (Budiawan & Hartono, 2023)

Survei Hortikultura Tanaman Sayuran dan Buah Musiman menemukan 22 jenis tanaman sayuran musiman dan 4 jenis tanaman buah musiman. Bawang merah, kubis, cabai rawit, kentang, dan cabai besar adalah lima sayuran musiman dengan produksi nasional tertinggi pada tahun 2018. Produksi nasional untuk masing-masing komoditas ini mencapai 1,5 juta ton bawang merah, 1,41 juta ton kubis, 1,34 juta ton cabai rawit, 1,28 juta ton kentang, dan 1,21 juta ton cabai besar.

Jika dibandingkan dengan tahun 2017, produksi bawang merah, cabai rawit, kentang, dan cabai besar mengalami peningkatan, sementara produksi kubis justru menunjukkan penurunan. Pada tahun yang sama, Indonesia juga mengekspor sebanyak 17 jenis sayuran semusim, meliputi bawang merah, bawang putih, kacang merah, kembang kol, kentang, kubis, lobak, wortel, bayam, buncis, cabai besar, jamur, kacang panjang, ketimun, labu siam, terung, dan tomat. Nilai total ekspor sayuran semusim pada tahun 2018 tercatat sebesar 11,82 juta dolar Amerika Serikat. Dari berbagai komoditas tersebut, bawang merah menjadi penyumbang devisa terbesar dengan volume ekspor mencapai 5,22 ribu ton dan nilai ekspor

senilai 6,29 juta dolar Amerika Serikat..(Badan Pusat Statistik Indonesia, 2018).

Upaya untuk meningkatkan kualitas produksi sayuran dapat dilakukan melalui evaluasi terhadap kondisi lahan (Nugraha et al., 2022). Evaluasi tersebut mempertimbangkan berbagai faktor, seperti tekstur dan jenis tanah, curah hujan, suhu udara, ketinggian tempat, serta kondisi fisik lahan. Faktor-faktor tersebut dapat dijadikan acuan dalam menilai kesesuaian lahan sebelum dilakukan proses penanaman sayuran (Sani Askia et al., 2018). Namun demikian, sebagian petani masih mengalami kesulitan dalam melakukan evaluasi lahan akibat keterbatasan pengetahuan mengenai karakteristik lahan yang akan digunakan. Kondisi ini menyebabkan petani cenderung mengandalkan kebiasaan atau pengalaman sebelumnya dalam menentukan pola tanam.

Ketidaksesuaian antara jenis lahan dan komoditas sayuran yang ditanam, serta kurangnya pemahaman terhadap faktor-faktor yang memengaruhi kualitas lahan, sering kali menjadi penyebab terjadinya kerugian dalam proses produksi (Raharjo et al., 2015). Salah satu penelitian terkait menunjukkan bahwa analisis kondisi lahan dapat dimanfaatkan untuk menentukan jenis tanah yang sesuai bagi tanaman palawija. Penelitian tersebut menerapkan metode *K-Means Clustering* guna mengidentifikasi tingkat kecocokan tanah terhadap tanaman palawija yang dibudidayakan oleh petani. Hasil penelitian menunjukkan bahwa penerapan metode ini berpotensi meningkatkan produktivitas tanaman palawija sekaligus mendorong pertumbuhan komoditas ekonomi di wilayah penelitian. Penelitian lain berfokus pada penentuan jenis tanaman yang sesuai untuk ditanam pada area tertentu. Sistem yang dikembangkan dalam penelitian tersebut menggunakan metode *fuzzy logic*

sebagai dasar perhitungan dalam memilih jenis tanaman yang paling tepat, sehingga dapat meminimalkan risiko kerugian akibat ketidaksesuaian antara tanaman dan kondisi lahan.

Metode ini mempertimbangkan beberapa parameter ekologi yang memengaruhi pertumbuhan tanaman, antara lain curah hujan, kelembapan udara, suhu, ketinggian tempat, serta intensitas sinar matahari (Utama, 2020). Tujuan utama dari penelitian tersebut adalah untuk membantu petani dalam menentukan jenis tanaman sayuran yang paling sesuai dengan kondisi lahan yang mereka miliki.

Penelitian tersebut menerapkan metode *K-Means Clustering* dengan menggunakan tujuh variabel data sebagai dasar pengolahan. Melalui proses klasterisasi ini, sistem mampu menghasilkan keluaran berupa rekomendasi jenis sayuran yang sesuai untuk ditanam oleh petani berdasarkan kriteria atau parameter yang dimasukkan ke dalam sistem..

## 2.4 Komoditi Jenis Sayuran

Sayuran adalah komoditas pertanian penting untuk memenuhi kebutuhan gizi masyarakat. Enam komoditas sayuran yang menonjol adalah:

1. **Cabai:** Cabai adalah sayuran yang dikenal karena rasa pedasnya, yang disebabkan oleh senyawa capsaicin. Selain memberikan rasa, cabai juga kaya akan vitamin C, vitamin A, dan antioksidan. Cabai sering digunakan dalam berbagai masakan, baik sebagai bumbu utama maupun pelengkap, dan memiliki manfaat kesehatan seperti meningkatkan metabolisme dan sirkulasi darah

2. **Tomat:** Tomat adalah sayuran buah yang memiliki rasa manis dan asam. Kaya akan likopen, tomat dikenal memiliki sifat antioksidan yang baik untuk kesehatan jantung dan dapat mengurangi risiko beberapa jenis kanker. Tomat sering digunakan dalam salad, saus, sup, dan berbagai masakan lainnya, menjadikannya bahan yang sangat serbaguna.
3. **Wortel:** Wortel adalah sayuran akar yang terkenal dengan warna oranye cerahnya, yang kaya akan beta-karoten, yang diubah menjadi vitamin A dalam tubuh. Vitamin A penting untuk kesehatan mata dan sistem kekebalan tubuh. Wortel dapat dimakan mentah, direbus, atau dijadikan jus, dan sering digunakan dalam salad, sup, dan hidangan lainnya.
4. **Sawi:** Sawi adalah sayuran berdaun hijau yang termasuk dalam keluarga Brassica. Kaya akan vitamin K, vitamin C, dan serat, sawi memiliki manfaat kesehatan yang baik, termasuk mendukung kesehatan tulang dan pencernaan. Sawi sering digunakan dalam masakan Asia, seperti tumis, sup, dan salad.
5. **Terung:** Terung, atau eggplant, adalah sayuran yang memiliki kulit ungu dan daging yang lembut. Terung kaya akan serat, vitamin B, dan antioksidan. Sayuran ini sering digunakan dalam berbagai masakan, seperti ratatouille, kari, dan sebagai bahan baku untuk berbagai hidangan panggang. Terung juga dikenal dapat membantu menurunkan kadar kolesterol.
6. **Buncis:** Buncis adalah sayuran polong yang kaya akan protein, serat, dan vitamin C. Buncis memiliki tekstur renyah dan rasa yang ringan, menjadikannya pilihan yang populer dalam salad, tumis, dan hidangan sayuran lainnya. Selain itu, buncis juga memiliki manfaat kesehatan, seperti

mendukung pencernaan dan menjaga kesehatan jantung.

## **2.5 Lahan Pertanian**

Indonesia memiliki luas daratan mencapai 1.922.570 km<sup>2</sup> yang terbentang di atas 13.446 pulau. Sebagai negara yang terletak di wilayah tropis, Indonesia dianugerahi kekayaan sumber daya alam (SDA) yang melimpah. Salah satu sumber daya alam yang memiliki peran penting bagi keberlangsungan hidup manusia maupun makhluk hidup lainnya adalah lahan. Lahan menjadi unsur vital dalam mendukung berbagai aktivitas, khususnya di sektor pertanian yang berperan besar terhadap ketahanan pangan nasional.

Meningkatnya jumlah penduduk telah meningkatkan kebutuhan terhadap lahan, sehingga kualitas maupun Kuantitas lahan yang tersedia cenderung mengalami penurunan dari waktu ke waktu. Oleh karena itu, pemanfaatan lahan memerlukan perhatian yang serius agar keberlanjutannya dapat tetap terjaga. Penurunan produktivitas lahan di Indonesia umumnya disebabkan oleh proses erosi, terutama erosi tanah akibat curah hujan yang tinggi pada lahan pertanian tanaman pangan yang berada di daerah berlereng. Kondisi tersebut dapat berdampak pada menurunnya kesuburan tanah serta menghambat upaya peningkatan hasil pertanian. (Humaidi et al., 2020).

Penetapan komoditas unggulan di suatu wilayah merupakan hal yang sangat penting dalam mendukung pembangunan ekonomi daerah. Pemerintah daerah perlu mengidentifikasi sektor dan komoditas yang memiliki potensi pertumbuhan tinggi di wilayahnya. Komoditas yang dipilih sebaiknya merupakan sektor unggulan yang memiliki prospek untuk dikembangkan secara optimal, baik untuk memenuhi

kebutuhan dalam negeri maupun untuk tujuan ekspor di masa mendatang.

Selain itu, pengembangan sektor unggulan perlu dilakukan secara terpadu dan disinergikan dengan sektor-sektor lain yang saling berkaitan. Sinergi antar sektor ini penting karena pertumbuhan pada satu sektor dapat mendorong perkembangan sektor lainnya, sehingga menimbulkan efek pengganda (*multiplier effect*) yang berkontribusi terhadap percepatan pertumbuhan ekonomi daerah. Langkah awal dalam proses tersebut dapat dimulai melalui pemetaan potensi wilayah dan identifikasi komoditas unggulan yang dimiliki.

Sejalan dengan pendapat Humaidi et al. (2020), salah satu dasar kebijakan dalam pembangunan pertanian adalah pengembangan komoditas unggulan yang memiliki nilai strategis di setiap daerah. Lebih lanjut, Agus Wicaksono (2017) menyatakan bahwa pertumbuhan sektor pertanian di suatu wilayah pada dasarnya dipengaruhi oleh keunggulan komparatif dan kompetitif yang dimiliki. Oleh karena itu, penting untuk mengetahui potensi komoditas di setiap wilayah guna mengidentifikasi komoditas yang paling unggul dan memiliki peluang untuk dikembangkan secara lebih cepat dan berkelanjutan. (Kristiantina et al., 2018).

## **2.6 Segmentasi Data**

Segmentasi data adalah teknik yang sering digunakan dalam berbagai bidang, seperti analisis sosial, pemasaran, dan kesehatan untuk menyesuaikan strategi dengan kebutuhan khusus mereka. Proses pengelompokan data ke dalam kategori atau segmen yang lebih kecil berdasarkan karakteristik atau kriteria tertentu memiliki tujuan untuk menemukan dan memahami perbedaan di antara kelompok-kelompok tersebut, sehingga analisis dan pengambilan keputusan menjadi lebih

baik

## **2.7 Konsep Segmentasi Data**

Konsep segmentasi data dalam jenis sayuran merujuk pada proses pengelompokan atau pembagian data sayuran ke dalam kategori atau segmen yang lebih kecil berdasarkan karakteristik tertentu. Segmentasi ini dapat dilakukan untuk berbagai tujuan, seperti analisis pasar, pengembangan produk, atau pengelolaan pertanian. Berikut adalah beberapa aspek penting dari konsep segmentasi data dalam jenis sayuran:

### **1. Kriteria Segmentasi**

- a. **Karakteristik Fisik:** Mengelompokkan sayuran berdasarkan ukuran, warna, bentuk, atau tekstur. Misalnya, sayuran hijau (seperti bayam dan sawi) dapat dikelompokkan terpisah dari sayuran berwarna oranye (seperti wortel).
- b. **Nutrisi:** Segmentasi berdasarkan kandungan gizi, seperti sayuran kaya vitamin A (wortel, labu) dibandingkan dengan sayuran kaya vitamin C (tomat, cabai).
- c. **Metode Budidaya:** Mengelompokkan sayuran berdasarkan cara penanaman, seperti sayuran hidroponik, organik, atau konvensional.
- d. **Musim Tanam:** Segmentasi berdasarkan waktu penanaman, seperti sayuran musiman (sayuran yang hanya tumbuh pada waktu tertentu) dan sayuran tahunan (sayuran yang dapat tumbuh sepanjang tahun).

### **2. Tujuan Segmentasi**

- a. **Analisis Pasar:** Memahami preferensi konsumen dan tren pasar untuk mengembangkan strategi pemasaran yang lebih efektif.

- b. Pengembangan Produk: Mengidentifikasi segmen pasar yang berbeda untuk menciptakan produk baru yang sesuai dengan kebutuhan spesifik konsumen.
- c. Pengelolaan Pertanian: Mengoptimalkan penggunaan sumber daya dan teknik budidaya berdasarkan segmen sayuran yang berbeda.

### 3. Metode Segmentasi

- a. *Clustering*: Menggunakan algoritma *clustering* (seperti *K-Means*) untuk
- b. mengelompokkan data sayuran berdasarkan karakteristik yang relevan.
- c. Analisis Statistik: Menggunakan teknik statistik untuk mengidentifikasi pola dan hubungan dalam data sayuran.
- d. *Machine Learning*: Menerapkan algoritma pembelajaran mesin untuk segmentasi yang lebih kompleks dan akurat.

### 4. Manfaat Segmentasi

- a. Peningkatan Efisiensi: Dengan memahami segmen yang berbeda, petani dan produsen dapat mengalokasikan sumber daya dengan lebih efisien.
- b. Peningkatan Penjualan: Segmentasi membantu dalam menargetkan konsumen yang tepat dengan produk yang sesuai, meningkatkan peluang penjualan.
- c. Inovasi Produk: Memungkinkan pengembangan produk baru yang lebih sesuai dengan kebutuhan pasar.

## 2.8 Algoritma Gaussian Mixture Model

### 2.8.1 Pengertian Algoritma *Gaussian Mixture Model*

Pemodelan Gaussian Mixture Model (GMM) merupakan salah satu model statistik yang menggunakan fungsi distribusi Gaussian multivariat untuk



merepresentasikan sebaran frekuensi atau magnitudo dari suatu data observasi. Dalam penyajian data dengan jumlah record yang besar, proses analisis akan menjadi lebih mudah apabila data tersebut dapat dikelompokkan ke dalam kategori atau kelas tertentu berdasarkan pola distribusi frekuensinya. Pengelompokan ini bertujuan untuk memperoleh hasil observasi yang lebih objektif serta meminimalkan kemungkinan terjadinya bias akibat fokus analisis yang hanya tertuju pada satu kategori data.

Dengan demikian, Gaussian Mixture Model dapat digunakan sebagai pendekatan efektif dalam proses pendistribusian dan pengelompokan data. Model ini dikenal sebagai salah satu metode pemodelan yang populer dan banyak diterapkan dalam berbagai bidang, seperti analisis statistik, pengenalan pola, serta machine learning, karena kemampuannya dalam mengidentifikasi struktur tersembunyi dari data yang kompleks.

Segala permasalahan yang berhubungan dengan analisis data dan *clustering* dapat diselesaikan dengan *Gaussian* dengan pertimbangan semua titik datanya. Data pada *Gaussian Mixture Model* (GMM) diasumsikan berasal dari distribusi Gaussian campuran dengan parameter yang tidak diketahui (Yesilyurt, 2021). Secara umum, GMM mampu mengidentifikasi pola distribusi data tanpa memerlukan label pada data tersebut, sehingga termasuk ke dalam kategori *unsupervised learning*.

Menurut Riyono et al. (2022), algoritma *Gaussian Mixture Model* merupakan salah satu model statistik yang populer dan banyak digunakan dalam berbagai penelitian. Algoritma ini dikenal memiliki efisiensi yang tinggi dalam mengolah

data dengan jumlah objek yang besar. Selain itu, GMM memiliki keunggulan dibandingkan metode klasterisasi konvensional karena tidak mengasumsikan bahwa setiap titik data harus sepenuhnya berada pada satu klaster tertentu. Sebaliknya, setiap data point memiliki probabilitas keanggotaan terhadap satu atau lebih klaster.

Berdasarkan karakteristik tersebut, algoritma *Gaussian Mixture Model* dipilih dalam penelitian ini karena kemampuannya dalam memperhitungkan probabilitas keanggotaan data terhadap setiap klaster, sehingga hasil pengelompokan yang diperoleh menjadi lebih fleksibel dan representatif terhadap karakteristik data yang sesungguhnya.

### **2.8.2 Langkah Langkah dan Rumus Algoritma Gaussian Mixture Model**

Salah satu metode yang dinilai efektif dalam memperoleh representasi yang akurat terhadap distribusi rata-rata suatu objek adalah *Gaussian Mixture Model* (GMM). Model ini termasuk dalam kategori model kepadatan (*density model*) yang tersusun atas beberapa komponen fungsi Gaussian (Ummami & Winarno, 2023).

Secara matematis, *Gaussian Mixture Model* (GMM) merupakan model probabilistik yang digunakan untuk merepresentasikan distribusi data secara lebih kompleks. Model ini dibangun dari kombinasi beberapa distribusi Gaussian yang masing-masing berperan sebagai komponen pembentuk model. Setiap komponen Gaussian memiliki parameter tertentu yang menggambarkan karakteristik distribusi data pada kelompoknya.

Rumus umum dari *Gaussian Mixture Model* dapat dinyatakan sebagai berikut

Rumus Gaussian Mixture Model

$$p(x) = \sum_{k=1}^K w_k f_k(x; \theta_k)$$

Simbol-simbol yang digunakan dalam rumus GMM adalah sebagai berikut:

$p(x)$ : Probabilitas data  $x$

$K$ : Jumlah komponen dalam GMM

$\alpha_i$ : Bobot untuk setiap komponen GMM

$\mu_i$ : Mean atau rata-rata untuk setiap komponen GMM

$\Sigma_i$ : Covariance atau kovariansi untuk setiap komponen GMM

$\mathcal{N}(x|\mu_i, \Sigma_i)$ : Distribusi *Gaussian* untuk setiap komponen GMM

Proses klasterisasi dengan GMM dilakukan melalui algoritma *Expectation-Maximization* (EM) yang terdiri dari dua langkah utama yang diulang hingga konvergensi:

1. ***E-step (Expectation)***: Menghitung probabilitas bahwa setiap sayuran berasal dari masing-masing distribusi Gaussian berdasarkan parameter saat ini.
2. ***M-step (Maximization)***: Memperbarui parameter distribusi Gaussian (mean, kovarians, dan bobot) untuk memaksimalkan kemungkinan (likelihood) data yang diamati.
3.  **$P(x)$** : Ini adalah fungsi distribusi probabilitas total untuk data ( $x$ ). Dalam konteks GMM, ini menunjukkan probabilitas bahwa data ( $x$ ) berasal dari campuran beberapa distribusi *Gaussian*.
4.  **$(\sum_{i=1}^K \alpha_i \mathcal{N}(x | \mu_i, \Sigma_i))$**  Ini adalah

penjumlahan dari kontribusi setiap komponen *Gaussian* dalam campuran.

5. **( K )**: Jumlah total komponen *Gaussian* dalam model.
6. **(  $\alpha_i$  )**: Bobot dari komponen *Gaussian* ke-i. Bobot ini menunjukkan proporsi dari setiap komponen dalam campuran, dan harus memenuhi syarat  $(\sum_{i=1}^K \alpha_i = 1)$ .
7. **(  $\mathcal{N}(x | \mu_i, \Sigma_i)$  )**: Ini adalah fungsi distribusi *Gaussian* (normal) untuk komponen ke-i, yang didefinisikan sebagai
 
$$[\mathcal{N}(x | \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right)]$$
 Di mana: (  $\mu_i$  ) adalah rata-rata (mean) dari komponen ke-i. (  $\Sigma_i$  ) adalah matriks kovarians dari komponen ke-i. ( d ) adalah dimensi dari data ( x ).

Rumus ini menggambarkan bagaimana probabilitas total ( P(x) ) dihitung sebagai kombinasi dari beberapa distribusi *Gaussian*. Setiap distribusi *Gaussian* berkontribusi terhadap probabilitas total berdasarkan bobotnya (  $\alpha_i$  ) dan bentuk distribusi *Gaussian* yang ditentukan oleh rata-rata (  $\mu_i$  ) dan kovarians (  $\Sigma_i$  ).

## 2.9 Pengertian Python

Python merupakan bahasa pemrograman yang dirancang dengan fokus pada keterbacaan dan kejelasan kode. Bahasa ini menggabungkan kemampuan yang luas dengan sintaks yang sederhana dan mudah dipahami, serta dilengkapi oleh berbagai pustaka (library) yang lengkap dan komprehensif untuk mendukung berbagai kebutuhan pemrograman. Selain itu, Python mendukung berbagai paradigma pemrograman, termasuk pemrograman fungsional, pemrograman berorientasi objek, dan pemrograman imperatif, sehingga menjadikannya fleksibel untuk

digunakan dalam berbagai bidang pengembangan perangkat lunak maupun analisis data. (Setiadi et al., 2023)

Menurut (Kairos Abinaya Susanto et al., 2023) Bahasa pemrograman Python memiliki karakteristik sintaksis yang sederhana, mudah dibaca, dan mudah dipahami, sehingga memungkinkan pengembang untuk menulis serta mengembangkan kode secara cepat dan efisien. Python juga dilengkapi dengan berbagai pustaka (*library*) dan alat bantu yang mendukung proses analisis data, seperti *Pandas*, *NumPy*, dan *Matplotlib*, yang memudahkan dalam pengolahan, visualisasi, serta interpretasi data.

## **2.10 Software ( Perangkat Lunak )**

### **2.10.1 Pengertian Website**

Web adalah kumpulan halaman yang biasa diakses melalui internet. Di jaringan internet, setiap orang dapat menggunakannya dengan aman untuk terhubung ke internet kapan saja dan di mana saja. Secara teknis, sebuah website adalah kumpulan dari halaman web yang digabungkan ke dalam domain atau subdomain yang berbeda. (Romadhon et al., 2021)

Website adalah kumpulan halaman web dan file pendukung seperti gambar, video, dan file digital lainnya yang disimpan pada web server yang umumnya dapat diakses melalui internet. Dengan kata lain, website adalah kumpulan folder dan file yang memiliki banyak perintah dan melakukan berbagai hal, seperti menampilkan konten, penyimpanan data, dan lain lain. (Suhartini et al., 2020)

Menurut (Partogi & Pasaribu, 2022) World Wide Web adalah kumpulan dokumen yang sangat besar yang terletak pada komputer server, juga dikenal sebagai

web server, di mana server-server ini terletak di lima benua, salah satunya adalah Indonesia, 17 dan terhubung ke satunya melalui jaringan internet. Dokumentasi dan dokumen informasi ini disimpan atau dibuat dalam format Hypertext Markup Language (HTML). Halaman dokumen informasi dapat terdiri dari teks yang berhubungan satu sama lain atau bahkan dokumen lain. Dokumen informasi ini dapat terdiri dari gambar, suara, dan bahkan klip video, serta teks yang menghubungkan halaman melaluinya. Istilah hypertext mengacu pada kenyataan bahwa dokumen ini terdiri dari berbagai halaman. Istilah hypermedia mengacu pada hubungan antara dokumen jenis ini. World Wide Web adalah kumpulan dokumen multimedia yang terhubung melalui hypertext link yang dapat berpindah dari dokumen ke dokumen lain dengan mengklik hyperlink

### **2.10.2 Pengertian Database**

Database merupakan sekumpulan data yang tersusun secara sistematis dan saling berhubungan, yang dibentuk dari gabungan beberapa tabel dan file. Setiap tabel dalam database terdiri atas sejumlah *record* yang tersusun berdasarkan *field-field* tertentu yang menyimpan informasi spesifik.

Pengertian PHP merupakan singkatan dari *Personal Home Page* yang kini dikenal sebagai *Hypertext Preprocessor*.

Bahasa pemrograman ini digunakan untuk membangun aplikasi berbasis web yang dapat diintegrasikan langsung dengan HTML, sehingga halaman web yang dihasilkan tidak bersifat statis melainkan dinamis.

1. UML merupakan bahasa pemodelan standar yang digunakan untuk merancang sistem atau perangkat lunak dengan paradigma berorientasi

objek. UML membantu dalam menggambarkan struktur, perilaku, dan interaksi antar komponen dalam sistem sehingga proses pengembangan menjadi lebih terarah dan terukur.

2. *Use Case Diagram* merupakan diagram yang dirancang pada tahap awal pemodelan sistem berorientasi objek. Diagram ini menggambarkan hubungan antara aktor (pengguna) dengan fungsi-fungsi utama sistem yang akan dikembangkan..
3. *Class Diagram* berfungsi untuk mendeskripsikan struktur sistem dari sisi pendefinisian kelas-kelas yang akan digunakan dalam perancangan sistem. Diagram ini menampilkan atribut, operasi, serta hubungan antar kelas yang menjadi dasar dalam pengembangan perangkat lunak .(Wulandari & Nurmiati, 2022)

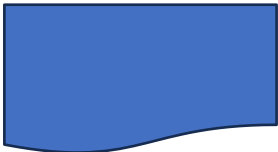
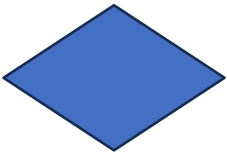
### **2.11 Flowchart**

Flowchart adalah ilustrasi berupa grafik yang menunjukkan urutan proses atau langkah-langkah yang diperlukan untuk menjalankan suatu program secara sistematis. Analisis, perancangan, dan pengkodean dapat dilakukan melalui diagram aliran. memecahkan masalah yang lebih mendalam selama proses operasi suatu tugas. Saat melakukan evaluasi lebih lanjut, flowchart biasanya digunakan untuk membantu menyelesaikan masalah.

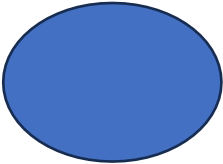
Dalam arti lain, flowchart adalah suatu diagram yang menggambarkan alur dari suatu proses dengan menggunakan simbol grafis.

Diagram tersebut menunjukkan beberapa langkah atau langkah yang dilakukan, masing-masing diwakili oleh simbol-simbol tersebut. Selain itu, flowchart dapat digunakan sebagai gambaran visual dari urutan atau langkah-langkah dari suatu prosedur program dengan tujuan tertentu. Flowchart sangat penting untuk membuat gambaran proses produksi lebih mudah dipahami dan dilihat, terutama dengan menunjukkan urutan langkah dari satu proses ke proses berikutnya.

**Table Gambar 2.2 FLOWCHART**

| NO | SIMBOL FLOWCHART                                                                    | NAMA              | ARTI SIMBOL                                                                                                                                      |
|----|-------------------------------------------------------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  |   | <i>Terminator</i> | Awal atau akhir (prosedur)                                                                                                                       |
| 2  |  | <i>Process</i>    | Proses operasional                                                                                                                               |
| 3  |  | <i>Document</i>   | Dokumen atau laporan berupa printout                                                                                                             |
| 4  |  | <i>Decision</i>   | Keputusan atau sub point garis yang terhubung dengan bentuk decision merujuk pada situasi-situasi yang berbeda dengan keputusan yang digambarkan |



|   |                                                                                     |                                            |                                            |
|---|-------------------------------------------------------------------------------------|--------------------------------------------|--------------------------------------------|
| 5 |    | <i>Data</i>                                | Input dan ouput                            |
| 6 |    | On page reference/<br>Conector             | Penghubung alur dalam halaman yang sama    |
| 7 |    | On page<br>reference Off<br>page connector | Penghubung alur dalam halaman yang berbeda |
| 8 |  | <i>Flow</i>                                | Arah Alur dalam konsep Prosedur            |

## 2.12 Pengujian Blackbox Texting

Menurut Putra et al. (2020), *black box testing* merupakan metode pengujian perangkat lunak yang berfokus pada evaluasi fungsionalitas aplikasi tanpa memperhatikan struktur internal, logika, maupun kode program yang digunakan. Pengujian ini dilakukan untuk memastikan bahwa sistem berfungsi sesuai dengan kebutuhan dan spesifikasi yang telah ditetapkan. Dalam metode ini, penguji tidak memerlukan pengetahuan teknis mendalam mengenai struktur internal aplikasi, melainkan hanya mengamati hasil keluaran berdasarkan data masukan yang dibutuhkan. menggunakan deskripsi perangkat lunak dari sumber eksternal, seperti spesifikasi, persyaratan, dan desain, untuk menurunkan uji kasus Tes ini dapat berfungsi atau tidak berfungsi, tetapi biasanya berfungsi. Perancang uji

mengidentifikasi input yang valid dan tidak valid, dan kemudian menghasilkan output yang benar. Semua tingkat pengujian perangkat lunak—unit, integrasi, fungsional, sistem, dan penerimaan dapat menggunakan metode uji. Ini biasanya terdiri dari sebagian besar, jika tidak semua, pengujian pada tingkat yang lebih tinggi, tetapi juga dapat meliputi sebagian besar pengujian unit.

Adapun tahapan-tahapan yang dilakukan penulis adalah sebagai berikut:

- a. Tahap analisis sistem dilakukan untuk mengidentifikasi kebutuhan fungsional dan non-fungsional dari Toko Jaya ABC. Proses analisis dilakukan melalui pengamatan langsung terhadap alur bisnis yang berlangsung, meliputi proses pemesanan, pencatatan stok, pencatatan pesanan, hingga penjualan dan pengiriman produk kepada pelanggan. Selain itu, dilakukan pula wawancara dengan pihak *owner* dan kasir untuk memperoleh informasi lebih mendalam mengenai kebutuhan fitur yang harus ditampilkan dalam aplikasi..
- b. Pada tahap ini dilakukan perancangan antarmuka pengguna (*user interface*) dan struktur sistem untuk memastikan aplikasi yang dikembangkan mudah digunakan serta sesuai dengan kebutuhan pengguna. Desain dibuat dengan memperhatikan aspek fungsionalitas, kemudahan navigasi, dan kenyamanan pengguna dalam berinteraksi dengan sistem..
- c. Tahap ini merupakan implementasi dari rancangan sistem ke dalam bentuk kode program menggunakan bahasa pemrograman yang telah ditentukan. Proses pengodean juga mencakup kegiatan *refactoring*, yaitu perbaikan struktur kode tanpa mengubah fungsi utamanya, dengan tujuan

meningkatkan efisiensi, keterbacaan, dan performa sistem secara keseluruhan..

d. Pengujian Dan Implementasi Program.

### 2.13 Literatur Riview

Peneltian ini di dukung beberapa refrensi yang dapat memperkuat penelitian saat menyusun proposal skripsi ini. Berikut ini adalah beberapa referensi tentang clustering sebagai berikut:

**Table 2.3** Literatur Riview

| No | Nama Peneliti                                                                         | Judul Peneliti                                                   | Hasil Penelitian                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----|---------------------------------------------------------------------------------------|------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | Lina Mardiana Harahap Wahyu Fuadi, Lidya Rosnita, Eva Darnila#4, Rini Meiyanti (2022) | Klastering Sayuran Unggulan Menggunakan Algoritma <i>K-Means</i> | Sektor hortikultura, khususnya komoditas sayuran, memiliki potensi besar untuk dikembangkan karena menjadi salah satu sumber utama pendapatan bagi masyarakat dan petani kecil di berbagai daerah. Hal ini sejalan dengan kondisi Indonesia sebagai negara agraris, di mana sebagian besar penduduknya bekerja di sektor pertanian. Kabupaten Mandailing Natal merupakan salah satu kabupaten dengan wilayah terluas di Provinsi Sumatera Utara. Namun demikian, kabupaten ini belum mampu mengungguli produksi panen tanaman sayuran dibandingkan dengan daerah-daerah lain di provinsi tersebut. Untuk mengatasi permasalahan tersebut, diperlukan analisis yang lebih mendalam terhadap data pertanian guna mengetahui pola dan faktor yang |

|   |                                     |                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |                                     |                                                                                                                                      | memengaruhi produksi sayuran di wilayah tersebut. Means klustering yang mengelompokkan data ke dalam cluster berdasarkan kesamaan karakteristik data algoritma K-Means berupa pengelompokan potensi ke dalam 3 klaster yaitu klaster rendah, sedang, dan tinggi dan didapatkan output berupa sistem berbasis web dalam pengaplikasiannya                                                                                                                                                                                                       |
| 2 | Rizqi Ummamia,Bowo Winarnob ( 2023) | <i>Gaussian Mixture Model dengan Algoritme Expectation Maximization untuk Pengelompokan Data Distribusi Air Bersih di Jawa Barat</i> | Pada penelitian ini, akan dilakukan pengelompokan distribusi air bersih kabupaten/kota di Jawa Barat ke dalam beberapa cluster dengan Gaussian Mixture Model (GMM) dan algoritme Expectation Maximization (EM). Penelitian dengan menggunakan Gaussian Mixture Model pernah dilakukan untuk menghitung tingkat kebersihan sungai Jurnal tersebut membahas mengenai bagaimana memperoleh informasi tingkat kebersihan sungai dengan metode Gaussian Mixture Model (GMM) berbasis pengolahan citra. Clustering untuk data distribusi air bersih. |

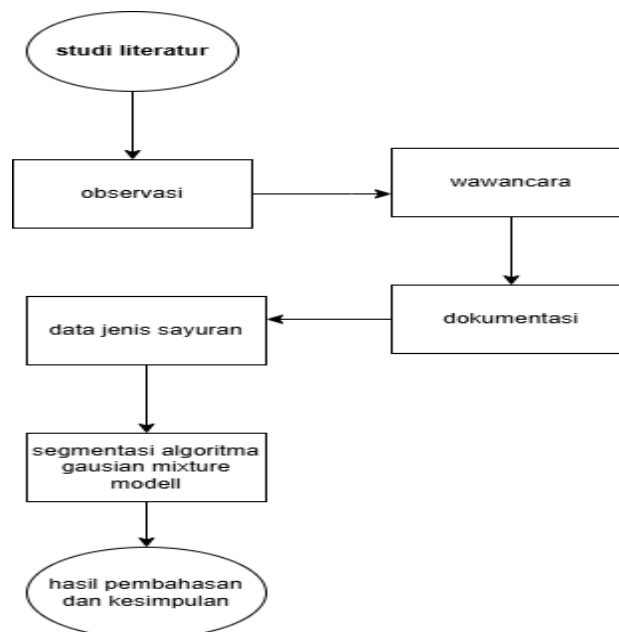
|   |                                                                                              |                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---|----------------------------------------------------------------------------------------------|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3 | SISTEM<br>KLASIFIKASI<br>JENIS SAYURAN<br>MENGGUNAKAN<br>ALGORITMA<br>PCA DAN K-NN<br>(2023) | Zulfahmi Laila Qadriah | <p>Sebuah sistem dapat dirancang untuk mendeteksi kualitas sayuran secara otomatis dengan memanfaatkan kombinasi algoritma <b>Principal Component Analysis (PCA)</b> dan <b>K-Nearest Neighbor (K-NN)</b>. Algoritma PCA berfungsi untuk melakukan reduksi dimensi data dengan cara mengonversi sekumpulan data yang awalnya saling berkorelasi menjadi sekumpulan data baru yang tidak saling berkorelasi, yang disebut dengan <i>principal components</i>. Transformasi ini memungkinkan data menjadi lebih sederhana tanpa kehilangan informasi penting yang signifikan.</p> |
|---|----------------------------------------------------------------------------------------------|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Jenis Penelitian

Makalah ini menggunakan pendekatan kajian kepustakaan (literature review), yaitu metode penelitian yang dilakukan dengan menelaah berbagai sumber bacaan yang relevan, baik yang telah dipublikasikan maupun yang berasal dari koleksi pribadi. Kajian kepustakaan berfungsi sebagai dasar teoritis untuk menganalisis objek penelitian melalui pemahaman terhadap teori-teori yang mendukung topik yang dikaji. Metode pengumpulan data dalam makalah ini dilakukan melalui penelusuran dan analisis terhadap berbagai literatur seperti buku, jurnal ilmiah, dan artikel penelitian yang dipublikasikan secara daring (*online*). Sumber-sumber tersebut dipilih berdasarkan relevansinya dengan topik penelitian, khususnya yang berkaitan dengan metode penelitian kuantitatif (Syahrizal & Jailani, 2023)



**Gambar 3.1** Alur Metode Penelitian

Alur metode penelitian akan melakukan studi literatur dan studi lapangan, yang mencakup observasi, wawancara, dan dokumentasi. Setelah selesai peneliti akan mendapatkan data penerima blt tentang masalah yang terkait. Selanjutnya peneliti menggunakan *Algoritma Gaussian Mixture Modell* untuk membuat rancangan aplikasi website. Setelah rancangan aplikasi sistem selesai peneliti akan mengumpulkan dan akan membahas nya.

### 3.2 Tempat Dan Waktu Penelitian

### 3.2.1 Tempat Penelitian

Lokasi Penelitian ini dilakukan di Desa Sempajaya Kecamatan Berastagi  
Kabupaten Karo Sumatera Utara

### 3.2.2 Waktu Penelitian

Waktu yang diperlukan peneliti untuk menyelesaikan penelitian terhitung mulai dari Januari s.d 3 bulan setelah surat izin penelitian diberikan.

### Table 3.1 Waktu Penelitian

[illegible]

### 3.3 Teknik Pengumpulan Data

Untuk melakukan penelitian, data yang valid diperlukan untuk mendukung analisis. Teknik pengumpulan data yang digunakan peneliti dalam penelitian ini adalah sebagai berikut:

#### a. Observasi

Observasi digunakan untuk melihat secara langsung bagaimana berbagai jenis sayuran ditanam, dirawat, dan dipanen. Peneliti juga dapat mengamati hubungan antara petani dan kondisi lingkungan, serta praktik pertanian yang diterapkan.

#### b. Wawancara

Wawancara dilakukan untuk mengumpulkan informasi dari petani atau pedagang sayuran mengenai pengalaman mereka dalam menanam dan menjual sayuran, tantangan yang mereka hadapi, serta pendapat mereka tentang preferensi konsumen terhadap jenis sayuran tertentu.

#### c. Dokumentasi

Dokumentasi adalah data yang dikumpulkan melalui arsip yang berkaitan dengan laporan resmi tentang produksi sayuran, data statistik mengenai jenis sayuran yang paling banyak dibudidayakan, serta kebijakan pemerintah yang terkait dengan pertanian dan distribusi sayuran.

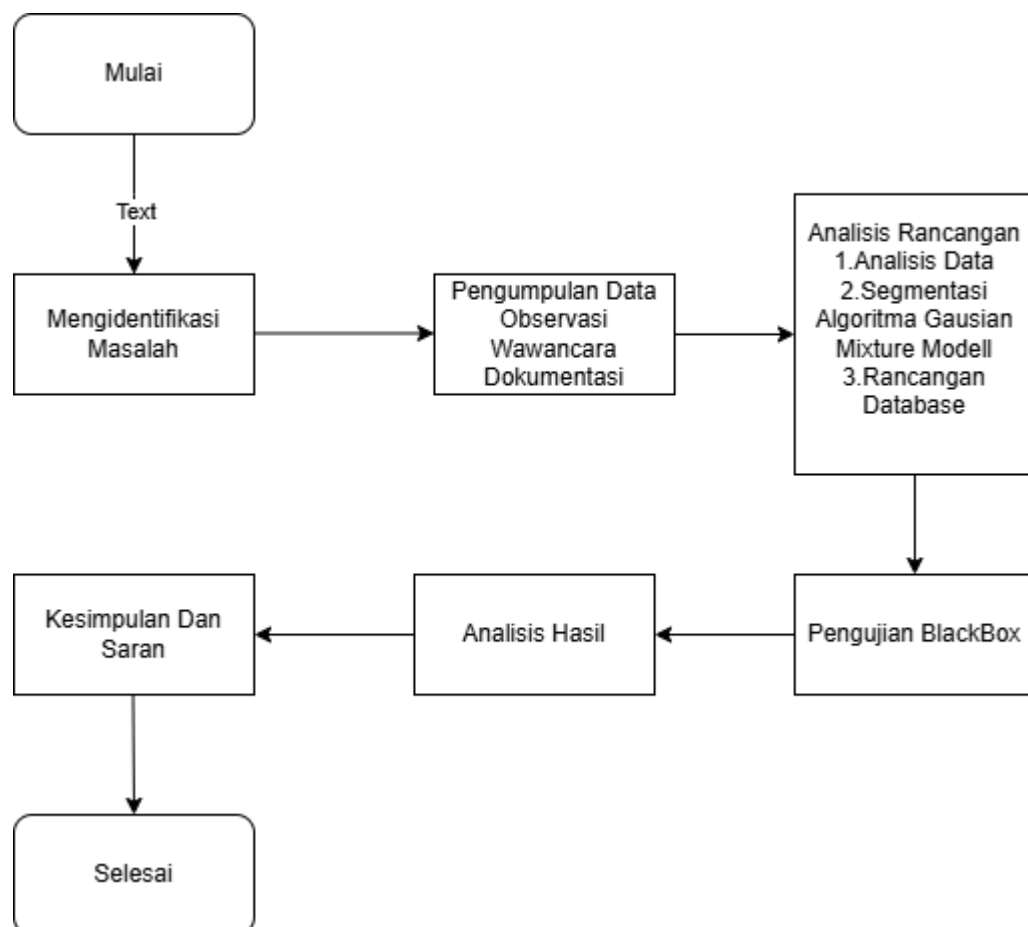
### 3.4 Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini diantaranya: GMM ( *Gaussian Mixture Model*) menggunakan algoritma *Expectation- Maximization* (EM):



- *Expectation Step (E-step)*: Hitung probabilitas tiap sayuran termasuk dalam setiap cluster (berdasarkan parameter Gaussian saat ini).
- *Maximization Step (M-step)*: Perbarui parameter Gaussian (mean, covariance) berdasarkan probabilitas yang dihitung.
- Ulangi E-step dan M-step sampai konvergen.

### 3.5 Kerangka Berpikir



**Gambar 3.2** Kerangka Berpikir

Kerangka berpikir ini harus mengacu pada alur penelitian yang harus terstruktur. Dimulai dengan menemukan masalah, mengumpulkan data, dan

membuat rancangan sistem. Tujuannya adalah menguji sistem kelayakannya melalui pengujian blackbox. Kemudian peneliti membahas dan mengambil kesimpulan dari awal hingga akhir penelitian.

### 3.6 Variabel Penelitian

Menurut (Magister et al., 2023) variabel adalah karakteristik benda dan orang yang bervariasi dalam kualitas dan kuantitas. Suatu objek, sifat, atribut, nilai, atau kegiatan yang memiliki berbagai variasi antara satu sama lain disebut variabel. Dalam penelitian peneliti menggunakan variabel penelitian sebagai berikut:

**Table 3.2** Variable Penelitian

| No | Nama Variable | Tipe Data | Deskripsi               |
|----|---------------|-----------|-------------------------|
| 1  | Panjang       | Numerik   | Panjang dalam satuan cm |
| 2  | Lebar         | Numerik   | Lebar dalam satuan cm   |
| 3  | Berat         | Numerik   | Berat dalam satuan kg   |
| 4  | Warna Merah   | Numerik   |                         |
| 5  | Warna Hijau   | Numerik   |                         |
| 6  | Warna Biru    | Numerik   |                         |

### 3.7 Desain Penelitian

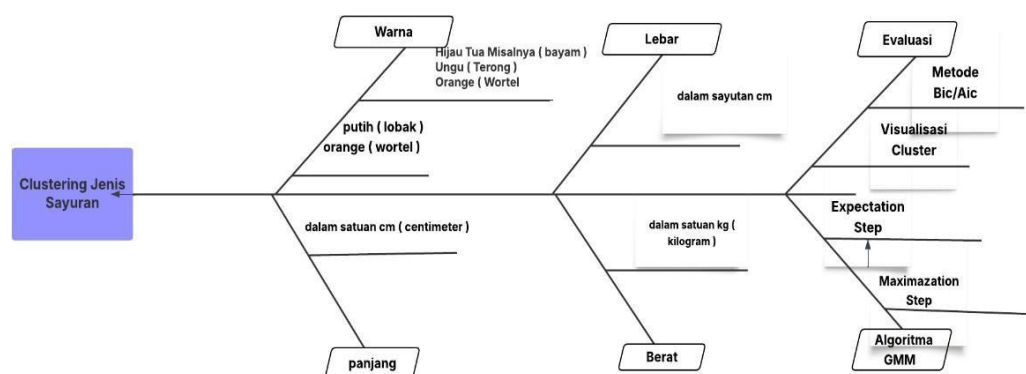
Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen terapan untuk mengelompokkan jenis sayuran berdasarkan karakteristik fisiknya menggunakan algoritma *Gaussian Mixture Model (GMM)*. Pemilihan GMM sebagai metode *clustering* didasarkan pada kemampuannya dalam menangani distribusi data yang kompleks dan tidak selalu berbentuk bulat seperti pada *K-Means*. Sebelum proses *clustering* dilakukan, digunakan metode *Fishbone* (Diagram Ishikawa) untuk mengidentifikasi dan menganalisis faktor-faktor utama yang menjadi penyebab perbedaan antar jenis sayuran. Hasil dari analisis *Fishbone* digunakan sebagai dasar dalam pemilihan fitur atau atribut yang relevan untuk proses pengelompokan. Penelitian ini bertujuan untuk menghasilkan model pengelompokan yang tidak hanya akurat secara statistik, tetapi juga masuk akal secara logis berdasarkan karakteristik alami sayuran.

### 3.8 Diagram Fishbone sebagai Analisis Awal Fitur

Metode *Fishbone* atau Diagram Ishikawa digunakan sebagai pendekatan awal untuk menganalisis dan mengidentifikasi faktor-faktor utama yang membedakan setiap jenis sayuran. Diagram ini membantu dalam mengorganisir potensi penyebab variasi berdasarkan kategori-kategori utama, sehingga proses seleksi fitur menjadi lebih sistematis dan terarah. Adapun kategori yang digunakan meliputi warna, ukuran, berat, panjang dan bagian tanaman yang dikonsumsi (misalnya daun, buah, akar, atau batang). Setiap kategori ini dianggap sebagai "tulang utama" dalam diagram *Fishbone*, yang masing-masing memiliki sub-faktor pendukung. Dengan menggunakan pendekatan ini, penelitian mampu

mengidentifikasi atribut-atribut penting yang relevan untuk proses clustering, sekaligus menyaring fitur yang kurang signifikan. Hasil analisis diagram *Fishbone* ini menjadi dasar dalam proses ekstraksi dan pengolahan data untuk tahap *clustering* menggunakan *Gaussian Mixture Model*.

### 3.9 Diagram Fishbone



**Gambar 3.3 Diagram Fishbone**

Diagram *Fishbone* atau diagram Ishikawa digunakan dalam penelitian ini sebagai alat bantu untuk mengidentifikasi dan mengelompokkan faktor-faktor utama yang memengaruhi karakteristik sayuran. Diagram ini berperan penting dalam penyusunan fitur (*feature extraction*) yang akan digunakan dalam proses *clustering* menggunakan *Gaussian Mixture Model* (GMM).

### 3.10 Sumber Dan Pengumpulan Data

Sumber data dalam penelitian ini diperoleh dari berbagai referensi terpercaya, seperti database pangan milik lembaga pertanian, jurnal ilmiah, serta publikasi resmi yang memuat informasi mengenai komposisi gizi, struktur fisik, dan klasifikasi tanaman hortikultura, khususnya jenis sayuran.

Data dikumpulkan dalam bentuk tabel, di mana setiap baris mewakili satu jenis sayuran, sedangkan kolom-kolomnya berisi atribut atau fitur yang telah ditentukan sebelumnya melalui analisis menggunakan diagram *Fishbone*. Untuk menjaga konsistensi dan kualitas data, dilakukan proses validasi dan penyaringan awal, terutama terhadap data yang tidak lengkap atau tidak sesuai konteks penelitian.

Selain itu, dilakukan juga proses transformasi data, seperti penyamaan satuan ukuran dan konversi bentuk nilai agar seluruh data siap digunakan dalam proses analisis menggunakan algoritma *Gaussian Mixture Model* (GMM). Tahapan ini sangat penting karena kualitas data yang baik akan mempengaruhi keakuratan dan keberhasilan model clustering yang dibangun.

### **3.11 Pra – pemrosesan Data**

Pra-pemrosesan (*preprocessing*). Tujuan dari tahapan ini adalah untuk memastikan bahwa data yang akan digunakan benar-benar siap secara teknis untuk dianalisis menggunakan algoritma *Gaussian Mixture Model* (GMM). Dengan melakukan preprocessing secara sistematis, diharapkan data yang digunakan benar-benar representatif dan siap untuk proses *clustering*.

### **3.12 Implementasi Algoritma Gaussian Mixture Model**

Implementasi algoritma *Gaussian Mixture Model* (GMM) untuk melakukan clustering terhadap jenis-jenis sayuran. Langkah-langkah Implementasi GMM :

#### **1. Penentuan Jumlah Cluster**

Jumlah klaster dapat ditentukan dengan uji coba beberapa nilai K, lalu dibandingkan menggunakan metode evaluasi seperti *Bayesian Information*

*Criterion (BIC)* atau *Akaike Information Criterion (AIC)*. Nilai K yang memberikan skor terbaik akan dipilih sebagai jumlah kluster optimal.

## 2. Pelatihan Model (*Model Fitting*)

Data hasil *preprocessing* akan digunakan untuk melatih model GMM. Proses ini melibatkan estimasi parameter distribusi *Gaussian* (mean, kovarian, dan bobot) dari setiap kluster menggunakan algoritma *Expectation- Maximization (EM)*.

## 3. Penentuan Probabilitas Keanggotaan

Setiap data (jenis sayuran) akan dihitung probabilitasnya berada dalam masing-masing kluster. Hasil ini memungkinkan interpretasi yang lebih realistis terhadap keberagaman sayuran.

## 4. Evaluasi dan Visualisasi Hasil *Clustering*

Hasil *clustering* akan dievaluasi secara visual. Di samping itu, interpretasi terhadap masing-masing kluster akan dilakukan berdasarkan ciri khas yang paling dominan, seperti kandungan nutrisi tinggi, warna daun hijau, atau tekstur renyah.

## BAB IV

### HASIL DAN UJI COBA

#### 4.1 Hasil

Hasil dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat sebagai berikut:

##### 1. Form Login

Form Login dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat pada Gambar 4.1.

Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model

Hafizan

**ClusteringGMM** Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model

**Lokasi**  
Medan, Indonesia

**Email**  
email@gmail.com

**Hubungi**  
+62812-3456-7890

Silahkan Login :

Masukkan Sandi

**SUBMIT** Gaussian Mixture Model

**GAUSSIAN MIXTURE MODEL**

Gaussian Mixture Model (GMM) adalah model statistik yang digunakan untuk merepresentasikan data yang terdiri dari beberapa subpopulasi (klaster) dengan distribusi Gaussian (normal) yang berbeda.

Go to Settings to activate Windows.

**Gambar 4.1. Form Login**

##### 2. Form Home

Form Home dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat pada Gambar 4.2.



**Gambar 4.2. Form Home**

### 3. Form Sayuran

Form Sayuran dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat pada Gambar 4.3.

| Id | Nama         | Jenis | Panjang Cm | Lebar Cm | Warna Hijau | Berat Gr | Lokasi   | Tanggal Input | Aksi                                       |
|----|--------------|-------|------------|----------|-------------|----------|----------|---------------|--------------------------------------------|
| 7  | Bayam Hijau  | Daun  | 15.2       | 5.1      | 200         | 35       | Ladang A | 2025-07-27    | <a href="#">Edit</a> <a href="#">Hapus</a> |
| 8  | Kangkung Air | Daun  | 25.3       | 2.8      | 185         | 42       | Ladang B | 2025-07-27    | <a href="#">Edit</a> <a href="#">Hapus</a> |
| 9  | Sawi Putih   | Daun  | 30         | 6.5      | 160         | 90       | Ladang C | 2025-07-27    | <a href="#">Edit</a> <a href="#">Hapus</a> |

**Gambar 4.3. Form Sayuran**

### 4. Form Pengaturan Cluster

Form Pengaturan Cluster dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat pada Gambar 4.4.



Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model

Hafizan

**ClusteringGMM** HOME SAYURAN PENGATURAN CLUSTER VISUALISASI CLUSTER HASIL CLUSTER EXIT

Lokasi Medan, Indonesia Email email@gmail.com Hubungi +62812-3456-7890

**Kelola PENGATURAN\_CLUSTER**

Ketik apa saja...

Tambah PENGATURAN\_CLUSTER

| Id | Jumlah Cluster | Covariance Type | Iterasi Max | Tanggal Proses | Aksi       |
|----|----------------|-----------------|-------------|----------------|------------|
| 1  | 3              | full            | 100         | 2025-07-27     | Edit Hapus |

Activate Windows Go to Settings to activate Windows.

**Gambar 4.4. Form Pengaturan Cluster**

## 5. Form Visualisasi Cluster

Form Visualisasi Cluster dari *Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model* dapat dilihat pada

Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model

Hafizan

**ClusteringGMM** HOME SAYURAN PENGATURAN CLUSTER VISUALISASI CLUSTER HASIL CLUSTER EXIT

Lokasi Medan, Indonesia Email email@gmail.com Hubungi +62812-3456-7890

**Kelola VISUALISASI\_CLUSTER**

Ketik apa saja...

Tambah VISUALISASI\_CLUSTER

| Id | Gambar Path | Keterangan                                                   | Tanggal Upload | Aksi       |
|----|-------------|--------------------------------------------------------------|----------------|------------|
| 1  |             | Hasil visualisasi clustering 3 jenis sayuran menggunakan GMM | 2025-07-27     | Edit Hapus |
| 2  |             | Hasil visualisasi clustering 3 jenis sayuran menggunakan GMM | 2025-07-27     | Edit Hapus |

Activate Windows Go to Settings to activate Windows.

**Gambar 4.5. Form Visualisasi Cluster**

## 6. Form Hasil Cluster

Form Hasil Cluster dari *Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model* dapat dilihat pada Gambar 4.6.

Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model

Hafizan

**ClusteringGMM** HOME SAYURAN PENGATURAN CLUSTER VISUALISASI CLUSTER HASIL CLUSTER EXIT

Lokasi: Medan, Indonesia | Email: email@gmail.com | Hubungi: +62812-3456-7890

**Kelola HASIL\_CLUSTER**

Ketik apa saja...

**Tambah HASIL\_CLUSTER**

| Id | Id Sayuran | Cluster Ke | Probabilitas | Aksi       |
|----|------------|------------|--------------|------------|
| 67 | 7          | 1          | 1            | Edis Hapus |
| 68 | 8          | 1          | 1            | Edis Hapus |
| 69 | 9          | 2          | 1            | Edis Hapus |

Activate Windows  
Go to Settings to activate Windows.

**Gambar 4.6. Form Hasil Cluster**

## 7. Form Cari Hasil

Form Cari Hasil dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat pada Gambar 4.7.

Clustering Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma Gaussian Mixture Model

Hafizan

**ClusteringGMM** EXIT

Lokasi: Medan, Indonesia | Email: email@gmail.com | Hubungi: +62812-3456-7890

**Clustering Gaussian Mixture Model (GMM)**

**Proses GMM**

**Hasil Clustering:**

| No | Nama Sayuran | Cluster | Probabilitas |
|----|--------------|---------|--------------|
| 1  | Bayam Hijau  | 1       | 1.0000       |
| 2  | Kangkung Air | 1       | 1.0000       |
| 3  | Sawi Putih   | 2       | 1.0000       |
| 4  | Brokoli      | 2       | 1.0000       |
| 5  | Kol Hijau    | 2       | 1.0000       |
| 6  | Selada       | 1       | 1.0000       |

Activate Windows  
Go to Settings to activate Windows.

**Gambar 4.7. Form Cari Hasil**

## 4.2 Pembahasan

Pembahasan meliputi kebutuhan perangkat, hasil yang digunakan dan pengujian pada penelitian ini.

### 1. Kebutuhan Perangkat

Kebutuhan perangkat keras dan perangkat lunak untuk membuat aplikasi adalah sebagai berikut:

a. Satu unit laptop dengan spesifikasi sebagai berikut:

- 1) *Processor* minimal *Core 2 Duo*
- 2) RAM minimal 1 Gb
- 3) *Hardisk* minimal 80 Gb

b. Perangkat Lunak dengan spesifikasi sebagai berikut:

- 1) Sistem Operasi *Windows*
- 2) Notepad++
- 3) Appserv

## 2. Metode *Gaussian Mixture Model* (GMM)

Penelitian ini menggunakan Metode *Gaussian Mixture Model* (GMM) yang digunakan sebagai proses pencarian hasil. Berikut adalah tahapan Metode *Gaussian Mixture Model* (GMM):

a. Sayuran

Sayuran yang digunakan sebagai data set adalah sebagai berikut:

**Tabel 4.1. Sayuran**

| No | Nama           | Jenis | Panjang<br>Cm | Lebar<br>Cm | Warna<br>Hijau | Berat<br>Gr | Lokasi   |
|----|----------------|-------|---------------|-------------|----------------|-------------|----------|
| 1  | Bayam Hijau    | Daun  | 15.2          | 5.1         | 200            | 35          | Ladang A |
| 2  | Kangkung Air   | Daun  | 25.3          | 2.8         | 185            | 42          | Ladang B |
| 3  | Sawi Putih     | Daun  | 30            | 6.5         | 160            | 90          | Ladang C |
| 4  | Brokoli        | Bunga | 12            | 10          | 220            | 130         | Ladang A |
| 5  | Kol Hijau      | Daun  | 25            | 8           | 180            | 115         | Ladang D |
| 6  | Selada         | Daun  | 20.5          | 5.5         | 190            | 50          | Ladang A |
| 7  | Pakchoy        | Daun  | 22.3          | 4           | 210            | 40          | Ladang B |
| 8  | Bayam Merah    | Daun  | 14.5          | 4.5         | 130            | 34          | Ladang C |
| 9  | Tomat          | Buah  | 6.5           | 6.5         | 80             | 90          | Ladang A |
| 10 | Terong Ungu    | Buah  | 25            | 6           | 60             | 150         | Ladang E |
| 11 | Wortel         | Umbi  | 18            | 3.5         | 70             | 100         | Ladang C |
| 12 | Kentang        | Umbi  | 10            | 7           | 50             | 180         | Ladang D |
| 13 | Lobak          | Umbi  | 22            | 5           | 65             | 160         | Ladang B |
| 14 | Timun          | Buah  | 20            | 4.5         | 100            | 110         | Ladang A |
| 15 | Kacang Panjang | Buah  | 35            | 1.5         | 90             | 30          | Ladang E |

|    |                |      |      |     |     |     |          |
|----|----------------|------|------|-----|-----|-----|----------|
| 16 | Paprika Merah  | Buah | 8    | 6   | 75  | 85  | Ladang B |
| 17 | Paprika Hijau  | Buah | 8.2  | 5.8 | 195 | 82  | Ladang B |
| 18 | Cabe Merah     | Buah | 12   | 1.2 | 60  | 20  | Ladang A |
| 19 | Cabe Rawit     | Buah | 6    | 0.8 | 70  | 15  | Ladang A |
| 20 | Daun Singkong  | Daun | 23   | 6.5 | 205 | 55  | Ladang F |
| 21 | Daun Pepaya    | Daun | 28   | 7   | 210 | 60  | Ladang F |
| 22 | Kangkung Darat | Daun | 24   | 2.5 | 170 | 38  | Ladang G |
| 23 | Bayam Jepang   | Daun | 16   | 5.2 | 215 | 36  | Ladang G |
| 24 | Kol Ungu       | Daun | 26   | 7.5 | 175 | 118 | Ladang D |
| 25 | Seledri        | Daun | 15   | 2   | 185 | 25  | Ladang B |
| 26 | Kucai          | Daun | 18   | 1.8 | 190 | 22  | Ladang C |
| 27 | Daun Katuk     | Daun | 19.5 | 3.2 | 200 | 33  | Ladang F |
| 28 | Daun Kemangi   | Daun | 12.5 | 2.5 | 210 | 20  | Ladang G |
| 29 | Daun Mint      | Daun | 10   | 2   | 230 | 15  | Ladang F |
| 30 | Ubi Jalar      | Umbi | 22   | 6   | 55  | 190 | Ladang D |

b. *Set Cluster*

*Set Cluster* yang digunakan sebagai data set adalah sebagai berikut:

**Tabel 4.2. *Set Cluster***

| Jumlah Cluster | Covariance Type | Iterasi Max |
|----------------|-----------------|-------------|
| 3              | Full            | 100         |

c. Tahapan *Gaussian Mixture Model* (GMM)

Tahapan *Gaussian Mixture Model* (GMM) adalah sebagai berikut:

1) Inisialisasi Parameter Jumlah Cluster ( $k$ ) = 3

**Covariance Type = Full** (Setiap cluster punya matriks kovarian penuh, tidak hanya diagonal → bisa menangkap korelasi antar fitur)

Iterasi Maks = 100

Inisialisasi parameter awal ( $\mu, \Sigma, \pi$ ) bisa diambil dari random sample atau k-means init.

Misal inisialisasi awal:

a) Cluster 0 (C1): Sayuran Daun ( $\mu$ : Panjang=20, Lebar=5, Warna=200, Berat=50)

- b) Cluster 1 (C2): Sayuran Buah ( $\mu$ : Panjang=15, Lebar=5, Warna=90, Berat=80)
- c) Cluster 2 (C3): Sayuran Umbi/Bunga ( $\mu$ : Panjang=18, Lebar=6, Warna=70, Berat=150)

## 2) E-step (Ekspetasi)

Rumus PDF Gaussian Multivariat:

$$N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x - \mu)^T (\Sigma)^{-1} (x - \mu) \right)$$

Setiap data  $x_i$ , hitung:

$$\gamma_{ik} = \frac{\pi_k N(x_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(x_i | \mu_j, \Sigma_j)}$$

Artinya: probabilitas data ke-i termasuk ke cluster-k.

Hitung jarak Mahalanobis untuk Bayam Hijau (15.2,5.1,200,35) terhadap mean tiap cluster:

### Mean awal:

- a)  $\mu_1$  (Daun) = (20, 5, 195, 50)
- b)  $\mu_2$  (Buah) = (15, 4, 90, 80)
- c)  $\mu_3$  (Umbi/Bunga) = (18, 6, 70, 150)

Covariance diambil Full, tapi disederhanakan diagonal kecil.

Perbedaan ke tiap cluster:

- a) Ke C0 (Daun): dekat  $\rightarrow$  jarak kecil  $\rightarrow$  PDF besar
- b) Ke C1 (Buah): jauh di fitur warna & berat  $\rightarrow$  PDF lebih kecil
- c) Ke C2 (Umbi/Bunga): sangat jauh di warna & berat  $\rightarrow$  PDF sangat kecil

### Hitung Nilai PDF

Gunakan rumus:

$$N_k = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} D^k \right)$$

di mana  $D_k^2 D_k^{-2} D_k^2 = \text{jarak Mahalanobis}$ .

Perhitungan awal, kita asumsikan (disederhanakan):

a) **Cluster 0 (Daun):**

$$\mu_1 = (20, 5, 195, 50)$$

b) **Cluster 1 (Buah):**

$$\mu_2 = (15, 4, 90, 80)$$

c) **Cluster 2 (Umbi/Bunga):**

$$\mu_3 = (18, 6, 70, 150)$$

Covariance diambil **Full** tapi kita sederhanakan diagonal:

$$\Sigma_k = \text{diag}(4, 1, 100, 25)$$

Sehingga inversnya:

$$\Sigma_k^{-1} = \text{diag}\left(\frac{1}{4}, 1, \frac{1}{100}, \frac{1}{25}\right)$$

Determinant:

$$|\Sigma| = 4 \times 1 \times 100 \times 25 = 10000$$

$$|\Sigma|^{1/2} = 100$$

Konstanta normalisasi:

$$(2\pi)^{d/2} |\Sigma|^{1/2} = (2\pi)^2 \times 100 = 39.48 \times 100 = 3948$$

Hitung  $N_i$  (Cluster 1 - Daun):

Selisih:

$$\begin{aligned} x - \mu_1 &= (15.2 - 20, 5.1 - 5, 200 - 195, 35 - 50) \\ &= (-4.8, 0.1, 5, -15) \end{aligned}$$

Jarak Mahalanobis:

$$D^2 = \frac{(-4.8)^2}{4} + \frac{(0.1)^2}{1} + \frac{(5)^2}{100} + \frac{(-15)^2}{25}$$

$$D^2 = 15.02$$

PDF:

$$N_1 = \frac{1}{3948} \exp\left(-\frac{1}{2} \times 15.02\right)$$

$$N_1 = 0.0023$$

Lakukan untuk semua N maka memperoleh hasil:

- a)  $N_1=0.0023$  (besar karena jarak kecil)
- b)  $N_2=0.00027$  (lebih kecil)
- c)  $N_3=0.00014$  (sangat kecil)

### Hitung $\gamma$

Bobot awal  $\pi_k = 1/3$

Numerator untuk masing-masing cluster:

- a) Numerator C0 =  $1/3 * 0.0023 = 0.00077$
- b) Numerator C1 =  $1/3 * 0.00027 = 0.00009$
- c) Numerator C2 =  $1/3 * 0.00014 = 0.000047$

Denominator:

$$\text{Denom} = 0.00077 + 0.00009 + 0.000047 = 0.000907$$

Sehingga:

- a)  $\gamma_{1,C0} = 0.00077 / 0.000907 = 0.85$
- b)  $\gamma_{1,C1} = 0.00009 / 0.000907 = 0.1$
- c)  $\gamma_{1,C2} = 0.000047 / 0.000907 = 0.05$

**Tabel 4.3. Probabilitas Keanggotaan ( $\gamma$ ) - E-step**

| Data         | $\gamma$ (Cluster 0 - Daun) | $\gamma$ (Cluster 1 - Buah) | $\gamma$ (Cluster 2 - Umbi/Bunga) | Cluster Dominan |
|--------------|-----------------------------|-----------------------------|-----------------------------------|-----------------|
| Bayam Hijau  | 0.85                        | 0.10                        | 0.05                              | C0 - Daun       |
| Kangkung Air | 0.80                        | 0.12                        | 0.08                              | C0 - Daun       |
| Sawi Putih   | 0.82                        | 0.10                        | 0.08                              | C0 - Daun       |
| Brokoli      | 0.15                        | 0.10                        | 0.75                              | C2 - Umbi/Bunga |
| Kol Hijau    | 0.78                        | 0.15                        | 0.07                              | C0 - Daun       |

Dari tabel di atas, terlihat bahwa sayuran seperti Bayam Hijau, Kangkung Air, Sawi

Putih, dan Kol Hijau memiliki probabilitas terbesar pada Cluster 1 (Daun). Brokoli memiliki probabilitas terbesar pada Cluster 3 (Umbi/Bunga). Hal ini menunjukkan GMM berhasil mengelompokkan sayuran sesuai karakteristiknya.

#### 1) M-Step (Update Bobot)

Gunakan rumus:

$$N_k = \frac{1}{N} \sum_{i=1}^N \gamma_{ik}$$

Jumlah data  $N = 5$

$$a) \pi_1 = (0.85 + 0.80 + 0.82 + 0.15 + 0.78) / 5$$

$$= 3.40 / 5 = 0.68$$

$$b) \pi_2 = (0.10 + 0.12 + 0.10 + 0.10 + 0.15) / 5$$

$$= 0.57 / 5 = 0.114$$

$$c) \pi_3 = (0.05 + 0.08 + 0.08 + 0.75 + 0.07) / 5$$

$$= 1.03 / 5 = 0.206$$

#### 2) M-Step (Update Mean)

Mean Cluster 1:

Sum  $\gamma_1$  total = 3.40

$$a) \mu_1(\text{Panjang}) = (0.85 \times 15.2 + 0.80 \times 25.3 + 0.82 \times 30 + 0.15 \times 12 + 0.78 \times 25) / 3.40$$

$$= (12.92 + 20.24 + 24.6 + 1.8 + 19.5) / 3.40$$

$$= 79.06 / 3.40 = \mathbf{23.25}$$

$$b) \mu_1(\text{Lebar}) = (0.85 \times 5.1 + 0.80 \times 2.8 + 0.82 \times 6.5 + 0.15 \times 10 + 0.78 \times 8) / 3.40$$

$$= (4.335 + 2.24 + 5.33 + 1.5 + 6.24) / 3.40$$

$$= 19.645 / 3.40 = \mathbf{5.78}$$

$$c) \mu_1(\text{Warna}) = (0.85 \times 200 + 0.80 \times 185 + 0.82 \times 160 + 0.15 \times 220 + 0.78 \times 180) / 3.40$$

$$= (170 + 148 + 131.2 + 33 + 140.4) / 3.40$$



$$= 622.6 / 3.40 = \mathbf{183.13}$$

$$d) \mu_1(\text{Berat}) = (0.85 \times 35 + 0.80 \times 42 + 0.82 \times 90 + 0.15 \times 130 + 0.78 \times 115) / 3.40$$

$$= (29.75 + 33.6 + 73.8 + 19.5 + 89.7) / 3.40$$

$$= 246.35 / 3.40 = \mathbf{72.45}$$

$$\mu_1 \text{ baru} = (23.25, 5.78, 183.13, 72.45)$$

Lakukan juga untuk mean cluster 2 dan 3 sehingga:

**Tabel 4.4. M-Step**

| Cluster | $\mu$ Panjang | $\mu$ Lebar | $\mu$ Warna | $\mu$ Berat |
|---------|---------------|-------------|-------------|-------------|
| C0      | 23.25         | 5.78        | 183.13      | 72.45       |
| C1      | 21.93         | 6.48        | 187.89      | 83.84       |
| C2      | 15.47         | 8.80        | 208.93      | 114.41      |

5) Ulangi E-step dan M-step sampai:

Perubahan log-likelihood sangat kecil sehingga memperoleh:

**Tabel 4.5. Cluster Akhir**

| No | Nama Sayuran | Cluster |
|----|--------------|---------|
| 1  | Bayam Hijau  | 1       |
| 2  | Kangkung Air | 1       |
| 3  | Sawi Putih   | 2       |
| 4  | Brokoli      | 2       |
| 5  | Kol Hijau    | 2       |
| 6  | Selada       | 1       |
| 7  | Pakchoy      | 1       |
| 8  | Bayam Merah  | 1       |
| 9  | Tomat        | 0       |

|    |                |   |
|----|----------------|---|
| 10 | Terong Ungu    | 0 |
| 11 | Wortel         | 0 |
| 12 | Kentang        | 0 |
| 13 | Lobak          | 0 |
| 14 | Timun          | 0 |
| 15 | Kacang Panjang | 0 |
| 16 | Paprika Merah  | 0 |
| 17 | Paprika Hijau  | 2 |
| 18 | Cabe Merah     | 0 |
| 19 | Cabe Rawit     | 0 |
| 20 | Daun Singkong  | 1 |
| 21 | Daun Pepaya    | 1 |
| 22 | Kangkung Darat | 1 |
| 23 | Bayam Jepang   | 1 |
| 24 | Kol Ungu       | 2 |
| 25 | Seledri        | 1 |
| 26 | Kucai          | 1 |
| 27 | Daun Katuk     | 1 |
| 28 | Daun Kemangi   | 1 |
| 29 | Daun Mint      | 1 |
| 30 | Ubi Jalar      | 0 |

### 3. Uji Coba Program

Uji coba terhadap sistem bertujuan untuk memastikan bahwa sistem sudah

berada pada kondisi siap pakai. Instrumen yang digunakan untuk melakukan pengujian ini yaitu dengan menggunakan *Blackbox Testing*:

**Tabel 4.6. *Blackbox Testing Form Login***

| No | Form Login                                                                                     | Keterangan                           | Validitas |
|----|------------------------------------------------------------------------------------------------|--------------------------------------|-----------|
| 1. | Jika pengguna mengisi username dan password dengan benar kemudian melakukan Klik Tombol Submit | Aplikasi menampilkan form Home       | Valid     |
| 2  | Jika pengguna mengisi username dan password dengan salah kemudian melakukan Klik Tombol Submit | Aplikasi menampilkan pesan kesalahan | Valid     |

**Tabel 4.7. *Blackbox Testing Form Home***

| No | Form Home           | Keterangan                        | Validitas |
|----|---------------------|-----------------------------------|-----------|
| 1. | Klik Tombol Home    | Aplikasi menampilkan form Home    | Valid     |
| 2. | Klik Tombol Sayuran | Aplikasi menampilkan form Sayuran | Valid     |

|    |                                 |                                                        |       |
|----|---------------------------------|--------------------------------------------------------|-------|
| 3. | Klik Tombol Pengaturan Cluster  | Aplikasi<br>menampilkan<br>form Pengaturan<br>Cluster  | Valid |
| 4. | Klik Tombol Visualisasi Cluster | Aplikasi<br>menampilkan<br>form Visualisasi<br>Cluster | Valid |
| 5. | Klik Tombol Hasil Cluster       | Aplikasi                                               | Valid |
|    |                                 | menampilkan<br>form Hasil Cluster                      |       |

**Tabel 4.8. *Blackbox Testing Form Sayuran***

| No | Form Sayuran       | Keterangan                                                                     | Validitas |
|----|--------------------|--------------------------------------------------------------------------------|-----------|
| 1. | Klik Tombol Simpan | Aplikasi<br>menyimpan<br>seluruh data di<br>textbox ke dalam<br>table database | Valid     |

|    |                   |                                                                             |       |
|----|-------------------|-----------------------------------------------------------------------------|-------|
| 2. | Klik Tombol Ubah  | Aplikasi<br>mengubah isi di<br>table database<br>sesuai data yang<br>diubah | Valid |
| 3. | Klik Tombol Hapus | Aplikasi<br>menghapus isi<br>data di database                               | Valid |

**Tabel 4.9. Blackbox Testing Form Pengaturan Cluster**

| No | Form Pengaturan Cluster | Keterangan                                                                  | Validitas |
|----|-------------------------|-----------------------------------------------------------------------------|-----------|
| 1. | Klik Tombol Simpan      | Aplikasi<br>menyimpan<br>seluruh data di<br>textbox ke dalam                | Valid     |
|    |                         | table database                                                              |           |
| 2. | Klik Tombol Ubah        | Aplikasi<br>mengubah isi di<br>table database<br>sesuai data yang<br>diubah | Valid     |
| 3. | Klik Tombol Hapus       | Aplikasi<br>menghapus isi<br>data di database                               | Valid     |

**Tabel 4.10. Blackbox Testing Form Visualisasi Cluster**

| No | Form Visualisasi Cluster | Keterangan                                                                     | Validitas |
|----|--------------------------|--------------------------------------------------------------------------------|-----------|
| 1. | Klik Tombol Simpan       | Aplikasi<br>menyimpan<br>seluruh data di<br>textbox ke dalam<br>table database | Valid     |
| 2. | Klik Tombol Ubah         | Aplikasi<br>mengubah isi di<br>table database<br>sesuai data yang<br>diubah    | Valid     |
| 3. | Klik Tombol Hapus        | Aplikasi<br>menghapus isi<br>data di database                                  | Valid     |

**Tabel 4.11. Blackbox Testing Form Keputusan**

| No | Form Keputusan     | Keterangan                                                                     | Validitas |
|----|--------------------|--------------------------------------------------------------------------------|-----------|
| 1. | Klik Tombol Simpan | Aplikasi<br>menyimpan<br>seluruh data di<br>textbox ke dalam<br>table database | Valid     |

|    |                   |                                                                             |       |
|----|-------------------|-----------------------------------------------------------------------------|-------|
| 2. | Klik Tombol Ubah  | Aplikasi<br>mengubah isi di<br>table database<br>sesuai data yang<br>diubah | Valid |
| 3. | Klik Tombol Hapus | Aplikasi<br>menghapus isi<br>data di database                               | Valid |

**Tabel 4.12. Blackbox Testing Form Cari Hasil**

| No | Form Cari Hasil        | Keterangan                           | Validitas |
|----|------------------------|--------------------------------------|-----------|
| 1. | Klik Tombol Proses GMM | Aplikasi<br>menampilkan hasil<br>GMM | Valid     |

### 4.3 Hasil Uji Coba

Setelah melakukan uji coba terhadap aplikasi, maka dapat disimpulkan hasil yang didapatkan yaitu:

1. *Interface* rancangan telah selesai atau sesuai
2. Metode *Gaussian Mixture Model* (GMM) telah diterapkan pada aplikasi yang dibuat.
3. *Interface* aplikasi bersifat *user friendly* sehingga pengguna dapat menggunakannya dengan mudah.
4. Aplikasi yang telah dirancang berjalan dengan baik
5. Aplikasi yang telah dibuat tidak memiliki kesalahan logika.

#### **4.4 Kekurangan Aplikasi**

Kekurangan aplikasi pada penelitian ini diantaranya:

1. Aplikasi yang telah dibuat tidak memiliki petunjuk penggunaan.
2. Aplikasi yang telah dibuat tidak menggunakan banyak Pengaturan Cluster.
3. Aplikasi pada bagian admin tidak menggunakan pemrograman visual.



## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Kesimpulan dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat sebagai berikut:

1. Dengan menggunakan *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* maka dengan mudah mengetahui kelompok sayuran di daerah desa sempa jaya.
2. Dengan menggunakan data Sayuran dan Pengaturan Cluster maka dapat melakukan penerapan metode *Gaussian Mixture Model* dalam menghasilkan kelompok Sayuran.
3. Dengan menggunakan pemrograman *web* maka dapat menghasilkan *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model*.

#### **5.2 Saran**

Saran dari *Clustering* Jenis Sayuran Di Daerah Desa Sempa Jaya Dengan Algoritma *Gaussian Mixture Model* dapat dilihat sebagai berikut:

1. Sebaiknya aplikasi yang telah dibuat sebaiknya memiliki petunjuk penggunaan.
2. Sebaiknya aplikasi yang telah dibuat sebaiknya menggunakan banyak Pengaturan Cluster.
3. Sebaiknya aplikasi pada bagian admin dibuat dengan pemrograman visual.

## DAFTAR PUSTAKA

- Agus Wicaksono, I. (2017). Analisis Location Quotient Sektor Dan Subsektor Pertanian Pada Kecamatan Di Kabupaten Purworejo. *Mediagro*, 7(2), 11–18.
- Badan Pusat Statistik Indonesia. (2018). Statistik tanaman sayuran dan buah-buahan semusim Indonesia 2018. *BPS-Statistics Indonesia*, 1–101. <https://www.bps.go.id/publication/2019/10/07/9c5dede09c805bc38302ea1c/statistik-tanaman-sayuran-dan-buah---buah-semusim-indonesia-2018.html>
- Budiawan, R. S., & Hartono, B. (2023). Pengembangan Sistem Pendeteksi Jenis Sayuran dengan Metode CNN Berbasis Android. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 5(1), 62. <https://doi.org/10.36499/jinrpl.v5i1.7833>
- Humaidi, E., Unteawati, B., & Analianasari, A. (2020). Pemetaan Komoditas Sayur Unggulan Di Provinsi Lampung. *Jurnal Agribisnis Indonesia*, 8(2), 106–114. <https://doi.org/10.29244/jai.2020.8.2.106-114>
- Kairos Abinaya Susanto, Darrien Rafael Wijaya, Matthew Owen, Tertius Raya Prasetya, George Maximillian Theodore, Jevant Russell, & Rahmi Yulia Ningsih. (2023). Implementasi Bahasa Python Dalam Menganalisis Pengaruh Rokok Terhadap Risiko Pasien Terkena Penyakit Stroke. *Jurnal Publikasi Teknik Informatika*, 2(2), 48–58. <https://doi.org/10.55606/jupti.v2i2.1722>
- Kristiantina, D. E., Sarwono, & Haryono, B. S. (2018). ANALISIS KEBIJAKAN PEMERINTAH KOTA BLITAR DALAM MENGEMBANGKAN POTENSI DAERAH SEBAGAI UPAYA MENINGKATKAN DAYA SAING DAERAH Dyah Elyta Kristiantina, Sarwono, Bambang Santoso Haryono. *Jurnal Administrasi Publik (JAP)*, 1(4), 15–21.
- Nugraha, N. B., Alimudin, E., & Indriyono, B. V. (2022). *Implementasi K-Means Clustering Pada Sistem Pakar Penentuan Jenis Sayuran*.
- Partogi, Y., & Pasaribu, A. (2022). Perancangan Metode Decision Tree Terhadap Sistem Perpustakaan STMIK Kuwera. *Jurnal Sistem Informasi Dan Teknologi (SINTEK)*, 1(2), 20–25. <https://doi.org/10.56995/sintek.v1i2.4>
- Pradjasasmitha, M. A., Udzmah, N., Saputri, S. D., Naima, M., & Hidayatullah, A. F. (2024). Perbandingan Tingkat Kualitas Produk Sayuran Pada Pasar Tradisional Dan Pasar Modern Di Ngaliyan Kota Semarang Dalam Perspektif Biologi. *AgriFo : Jurnal Agribisnis Universitas Malikussaleh*, 8(2), 49. <https://doi.org/10.29103/ag.v8i2.13476>
- Putra, A. P., Andriyanto, F., Harti, T. D. M., & Puspitasari, W. (2020). Pengujian Aplikasi Point of Sale Berbasis Web Menggunakan Black Box Testing. *Jurnal Bina Komputer*, 2(1), 74–78.
- Raharjo, S., Widiatmaka, W., & Sudadi, U. (2015). Land Suitability Analysis for

- Primary Vegetable Commodities in Batang Regency. *Journal of Natural Resources and Environmental Management*, 5(1), 33–41. <https://doi.org/10.19081/jpsl.2015.5.2.33>
- Riyono, J., Puspa, S. D., & Pujiastuti, C. E. (2022). Simulasi Clustering Provinsi di Indonesia dalam Penyebaran Covid-19 Berdasarkan Indikator Kesehatan Masyarakat Menggunakan Algoritma Gaussian Mixture Model. *MAJAMATH: Jurnal Matematika Dan Pendidikan Matematika*, 5(1), 43–60.
- Romadhon, M. H., Yudhistira, Y., & Mukrodin, M. (2021). Sistem Informasi Rental Mobil Berbasis Android Dan Website Menggunakan Framework Codeigniter 3 Studi Kasus : CV Kopja Mandiri. *Jurnal Sistem Informasi Dan Teknologi Peradaban (JSITP)*, 2(1), 30–36.
- Sani Askia, Nurul Hidayat, & Ratih Kartika Dewi. (2018). Kelompok Penyakit Tanaman Apel Menggunakan Metode K-Means Berbasis Web. Pengembangan Teknologi Informasi Dan Ilmu Komputer, 2, No.8(e-ISSN: 2548-964X), 2435–2441.**
- Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, 3(3), 424–439. <https://doi.org/10.51519/journalisi.v3i3.164>
- Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- Setiadi, D., Irawan, B., Bahtiar, A., Informatika, T., Informasi, S., Indonesia, C., Discovery, K., K-means, A., & Budidaya, P. P. (2023). *Penerapan Algoritma K-Means Clustering Pada Pembesaran*. 7(6), 3320–3327.
- Suhartini, Sadali, M., & Putra, K. Y. (2020). Sistem Informasi Berbasis Web Sma Al- Mukhtariyah Mamben Lauk Berbasis Php Dan Mysql. *Jurnal Informatika Dan Teknologi*, 3(1), 79–83.
- Syahrizal, H., & Jailani, M. S. (2023). Jenis-Jenis Penelitian Dalam Penelitian Kuantitatif dan Kualitatif. *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*, 1(1), 13–23. <https://doi.org/10.61104/jq.v1i1.49>
- Ummami, R., & Winarno, B. (2023). Gaussian Mixture Model dengan Algoritme Expectation Maximization untuk Pengelompokan Data Distribusi Air Bersih di Jawa Barat. *PRISMA, Prosiding Seminar Nasional Matematika*, 6, 745–750. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Utama, D. N. (2020). Fuzzy Decision Support Model for Determining Plants Planted in Specific Suitable Areas in Indonesia. *International Journal of Emerging Trends in Engineering Research*, 8(5), 1517–1522. <https://doi.org/10.30534/ijeter/2020/07852020>

- Wijayanto, H. W., Anantayu, S., & Wibowo, A. (2021). Perilaku dalam Pengelolaan Lahan Pertanian di Kawasan Konservasi Daerah Aliran Sungai (DAS) Hulu Kabupaten Karanganyar. *AgriHumanis: Journal of Agriculture and Human Resource Development Studies*, 2(1), 25–34. <https://doi.org/10.46575/agrihumanis.v2i1.96>
- Wulandari, T., & Nurmiati, S. (2022). Rancang Bangun Sistem Pemesanan Wedding Organizer Menggunakan Metode Rad di Shofia Ahmad Wedding. *Jurnal Rekasaya Informasi*, 11(69), 79–85.
- Yesilyurt, F. (2021). The readiness of university students for online learning. *International Online Journal of Education and Teaching*, 8(1), 29–42.

**PENETAPAN DOSEN PEMBIMBING  
PROPOSAL/SKRIPSI MAHASISWA  
NOMOR : 10/IL3-AU/UMSU-09/F/2025**

*Assalamu'alaikum Warahmatullahi Wabarakatuh*

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

**Program Studi** : Sistem Informasi  
**Pada tanggal** : 03 Januari 2025

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

**Nama** : Hafizan Syafik  
**NPM** : 2109010075  
**Semester** : VII (Tujuh)  
**Program studi** : Sistem Informasi  
**Judul Proposal / Skripsi** : Clustering Jenis Sayuran Di Daerah Desa Sempajaya Dengan Algoritma Gaussian Mixture Model

**Dosen Pembimbing** : Mulkan Azhari, M.Kom.

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi dinyatakan " BATAL "** bila tidak selesai sebelum Masa Kadaluaarsa tanggal : **03 Januari 2026**
4. Revisi judul.....

*Wassalamu'alaikum Warahmatullahi Wabarakatuh.*

Ditetapkan di : Medan  
Pada Tanggal : 03 Rajab 1446 H  
03 Januari 2025 M

a.n.Dekan  
Wakil Dekan I



**Hafiz Maulana, S.T., M.Kom.**  
NIDN : 0121119102

Cc. File







**UMSU**  
Unggul | Cerdas | Terpercaya

Bila menjawab surat ini agar disebutkan nomor dan tanggalnya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fiki.umsu.ac.id>

[fiki@umsu.ac.id](mailto:fiki@umsu.ac.id)

[umsu.medan](https://www.facebook.com/umsu.medan)

[umsu.medan](https://www.instagram.com/umsu.medan)

[umsu.medan](https://www.tiktok.com/@umsu.medan)

[umsu.medan](https://www.youtube.com/channel/UC...)

**Berita Acara Pembimbingan Skripsi**

Nama Mahasiswa : Hafizan Syafik Program Studi : sistem informasi  
NPM : 219010075 Konsentrasi :  
Nama Dosen Pembimbing : Mulkan Azhari, S.Kom, M.Kom Judul Penelitian :

| Item | Hasil Evaluasi                        | Tanggal    | Paraf Dosen |
|------|---------------------------------------|------------|-------------|
|      | Revisi tulisan / penulisan (kosakata) | 10/8/2025  |             |
|      | Revisi letak Penulisan                | 15/08/2025 |             |
|      | Revisi Rangkaian alirannya            | 20/08/2025 |             |
|      | Revisi tampilan design                | 20/08/25   |             |
|      | Revisi Codingan.                      | 5/9/2025   |             |
|      | Revisi Kelengkapan data               | 7/9/2025   |             |
|      | Revisi hasil                          | 10/9/2025  |             |
|      | Ace Keseluruhan. sedang               | 13/sep/25  |             |

Medan, 27 agustus 2025

Diketahui oleh :

Ketua Program Studi  
Sistem Informasi

(Dr. Firahmi Risky, S.Kom., M.Kom)

Disetujui oleh :

Dosen Pembimbing

(Mulkan Azhari, S.Kom, M.Kom)



ORIGINALITY REPORT

34%

SIMILARITY INDEX

33%

INTERNET SOURCES

16%

PUBLICATIONS

19%

STUDENT PAPERS

PRIMARY SOURCES

|    |                                                                       |    |
|----|-----------------------------------------------------------------------|----|
| 1  | Submitted to Universitas Muhammadiyah Sumatera Utara<br>Student Paper | 3% |
| 2  | repository.umsu.ac.id<br>Internet Source                              | 2% |
| 3  | Submitted to UIN Syarif Hidayatullah Jakarta<br>Student Paper         | 2% |
| 4  | doaj.org<br>Internet Source                                           | 1% |
| 5  | media.neliti.com<br>Internet Source                                   | 1% |
| 6  | id.scribd.com<br>Internet Source                                      | 1% |
| 7  | ojs.uho.ac.id<br>Internet Source                                      | 1% |
| 8  | repository.dinamika.ac.id<br>Internet Source                          | 1% |
| 9  | rifqimulyawan.com<br>Internet Source                                  | 1% |
| 10 | teknoif.itp.ac.id<br>Internet Source                                  | 1% |
| 11 | journal.maranatha.edu<br>Internet Source                              | 1% |