

**PENGELOMPOKAN PRODUK PESTISIDA BERDASARKAN POLA  
KONSUMSI RETAILER MENGGUNAKAN ALGORITMA  
HAC DAN K-MEDOIDS**

**SKRIPSI**

**DISUSUN OLEH**

**DINDA ANANTYA**

**2209020148**



**UMSU**

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2026**

**PENGELOMPOKAN PRODUK PESTISIDA BERDASARKAN POLA  
KONSUMSI RETAILER MENGGUNAKAN ALGORITMA  
HAC DAN K-MEDOIDS**

**SKRIPSI**

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

**DINDA ANANTYA**

**2209020148**

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

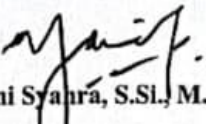
**MEDAN**

**2026**


**LEMBAR PENGESAHAN**

Judul Skripsi : Pengelompokan Produk Pestisida Berdasarkan Pola  
Konsumsi Retailer Menggunakan Algoritma HAC dan K-  
Medoids  
Nama Mahasiswa : Dinda Anantya  
NPM : 2209020148  
Program Studi : Teknologi Informasi

Menyetujui  
Komisi Pembimbing

  
(Yohanni Syahra, S.Si., M.kom)  
NIDN. 0129108201

Ketua Program Studi

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

Dekan

  
UNIVERSITAS MUHAMMADIYAH SUKOHARJO  
FAKULTAS TEKNOLOGI KOMPUTER DAN TEKNOLOGI INFORMASI  
(Dr. Fekher Rizmi, S.Kom., M.Kom.)  
NIDN. 0127099201

**PERNYATAAN ORISINALITAS**

**PENGELOMPOKAN PRODUK PESTISIDA BERDASARKAN POLA  
KONSUMSI RETAILER MENGGUNAKAN ALGORITMA**

**HAC DAN K-MEDOIDS**

**SKRIPSI**

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

Medan, Februari 2026

Yang membuat pernyataan



Dinda Anantya

NPM. 2209020148

## RIWAYAT HIDUP

### DATA PRIBADI

Nama Lengkap : Dinda Anantya  
Tempat dan Tanggal Lahir : Sukajadi, 18 April 2004  
Alamat Rumah : Desa Celawan Dusun IX Kec. Pantai Cermin  
Telepon/Faks/HP : 082161334713  
E-mail : dindaanantya1804@gmail.com  
Instansi Tempat Kerja : -  
Alamat Kantor : -

### DATA PENDIDIKAN

SD : Sukaramai No 101956 TAMAT: 2016  
SMP : SMPN 1 Pantai Cermin TAMAT: 2019  
SMA : SMAN 1 Perbaungan TAMAT: 2022

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

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

Nama : Dinda Anantya  
NPM : 2209020148  
Program Studi : Teknologi 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:

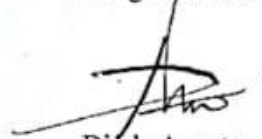
**PENGELOMPOKAN PRODUK PESTISIDA BERDASARKAN POLA  
KONSUMSI RETAILER MENGGUNAKAN ALGORITMA  
HAC DAN K-MEDOIDS**

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, 11 Maret 2026

Yang membuat pernyataan



Dinda Anantya

NPM. 2209020148

## KATA PENGANTAR



Penulis senantiasa memanjatkan puji dan syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, serta kekuatan yang diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Perjalanan dalam menyusun karya ini bukanlah proses yang singkat maupun mudah. Di dalamnya terdapat banyak tantangan, kelelahan, keraguan, serta proses belajar yang penuh makna. Namun berkat doa, dukungan, dan bimbingan dari berbagai pihak, penulis dapat melalui setiap tahapnya hingga akhirnya skripsi ini dapat terselesaikan. Oleh karena itu, dengan penuh rasa hormat dan ketulusan hati, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Agussani, M.Ap., selaku Rektor Universitas Muhammadiyah Sumatera Utara (UMSU), yang telah memberikan kesempatan dan fasilitas kepada penulis untuk menempuh pendidikan dan menimba ilmu di lingkungan universitas ini.
2. Bapak Assoc. Prof. Dr. Al-Khowarizmi, M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU, yang telah memberikan arahan serta dukungan selama penulis menjalani masa perkuliahan.
3. Ibu Fatma Sari Hutagalung, M.Kom., selaku Ketua Program Studi Teknologi Informasi, atas perhatian, bantuan, serta kemudahan administrasi yang diberikan kepada penulis selama proses studi hingga penyusunan skripsi ini.

4. Bapak Okvi Nugroho, S.kom., M.Kom., selaku Sekretaris Program Studi Teknologi Informasi, yang telah memberikan dukungan serta motivasi sehingga penulis dapat menjalani proses akademik dengan baik..
5. Ibu Yohanni Syahra, S.Si., M.Kom., selaku Dosen Pembimbing yang dengan penuh kesabaran, ketulusan, serta dedikasi telah memberikan arahan, ilmu, kritik, dan saran yang sangat berharga bagi penulis dalam menyelesaikan penelitian ini. Setiap bimbingan yang diberikan menjadi pelajaran penting yang tidak hanya membantu dalam penyusunan skripsi ini, tetapi juga dalam membentuk cara berpikir penulis sebagai seorang akademisi.
6. Dengan penuh rasa syukur dan hormat, penulis mengucapkan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta yang selalu memberikan doa, dukungan, kasih sayang, serta semangat tanpa henti. Segala pengorbanan, perhatian, dan motivasi yang diberikan menjadi kekuatan terbesar bagi penulis dalam menyelesaikan karya ini. Tanpa doa dan dukungan dari orang tua, penulis tidak akan mampu mencapai tahap ini. Semoga segala kebaikan dan kasih sayang yang telah diberikan senantiasa mendapatkan balasan yang terbaik.
7. Sahabat-sahabat terbaik penulis, Aida Fadhila, dan Dwi Nisyatul Wardah, Farhan Fanalty yang selalu hadir dalam perjalanan ini sebagai teman berbagi cerita, tempat berkeluh kesah, serta sumber semangat ketika rasa lelah dan putus asa mulai menghampiri. Terima kasih atas tawa, dukungan, dan kebersamaan yang membuat perjalanan ini terasa lebih ringan untuk dijalani. Semua pihak yang terlibat langsung ataupun tidak langsung yang

tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

8. Terakhir, penulis juga ingin menyampaikan terima kasih kepada diri sendiri. Terima kasih karena telah bertahan melalui hari-hari yang penuh tekanan, tetap berusaha ketika rasa lelah datang, dan tidak menyerah meskipun prosesnya terasa berat. Terima kasih karena telah berani memulai, berjuang sampai sejauh ini, dan membuktikan bahwa setiap usaha yang dilakukan dengan sungguh-sungguh akan menemukan jalannya. Skripsi ini bukan hanya tentang menyelesaikan sebuah kewajiban akademik, tetapi juga tentang perjalanan belajar, bertumbuh, dan menjadi pribadi yang lebih kuat.

Seluruh pihak yang telah membantu penulis, baik secara langsung maupun tidak langsung, yang tidak dapat disebutkan satu per satu. Setiap bantuan, doa, dan kebaikan yang diberikan memiliki arti yang sangat besar dalam perjalanan penyelesaian skripsi ini.

**PENGELOMPOKAN PRODUK PESTISIDA BERDASARKAN POLA  
KONSUMSI RETAILER MENGGUNAKAN ALGORITMA  
HAC DAN K-MEDOIDS**

**ABSTRAK**

Sektor pertanian sangat bergantung pada efektivitas input produksi pestisida, namun CV. Aneka Jaya Tunas Agro saat ini menghadapi tantangan operasional karena pengambilan keputusan distribusi yang cenderung subjektif. Penelitian ini bertujuan untuk mengelompokkan pola konsumsi retailer guna menghasilkan insight strategis dalam manajemen stok dan optimalisasi distribusi. Dataset yang digunakan merupakan transaksi penjualan tahun 2023–2024 yang dibersihkan menjadi 1.488 data valid dan diagregasi menjadi 134 profil retailer. Menggunakan metodologi Knowledge Discovery in Databases (KDD), dilakukan perbandingan performa antara algoritma Hierarchical Agglomerative Clustering (HAC) dengan metode Ward dan K-Medoids (PAM) menggunakan metrik Silhouette Score dan Davies-Bouldin Index (DBI). Hasil pengujian menunjukkan bahwa algoritma HAC memiliki performa lebih unggul dengan nilai Silhouette Score sebesar 0,5195 dan DBI sebesar 0,7207, dibandingkan K-Medoids yang menghasilkan Silhouette Score 0,2914 dan DBI 1,2987. Berdasarkan hasil klasterisasi, retailer dibagi menjadi tiga kategori utama, yaitu retailer besar (10 unit), retailer menengah (51 unit), dan retailer kecil (73 unit), yang masing-masing diberikan rekomendasi strategi bisnis spesifik. Penelitian ini menyimpulkan bahwa algoritma HAC lebih efektif dalam memetakan pola konsumsi retailer untuk mendukung pengambilan keputusan berbasis data (data-driven).

**Kata Kunci:** Klasterisasi, Pola Konsumsi, Pestisida, Hierarchical Agglomerative Clustering (HAC), K-Medoids.

# **CLUSTERING PESTICIDE PRODUCTS BASED ON RETAILER CONSUMPTION PATTERNS USING HAC AND K-MEDOIDS ALGORITHMS**

## ***ABSTRACT***

*The agricultural sector relies heavily on the effectiveness of pesticide production inputs; however, CV. Aneka Jaya Tunas Agro currently faces operational challenges due to subjective distribution decision-making. This study aims to cluster retailer consumption patterns to generate strategic insights for stock management and distribution optimization. The dataset consists of 2023–2024 sales transactions, which were cleaned into 1,488 valid records and aggregated into 134 retailer profiles. Using the Knowledge Discovery in Databases (KDD) methodology, the performance of the Hierarchical Agglomerative Clustering (HAC) algorithm with Ward's method was compared against K-Medoids (PAM) using the Silhouette Score and Davies-Bouldin Index (DBI) metrics. Test results indicate that the HAC algorithm performs better, achieving a Silhouette Score of 0.5195 and a DBI of 0.7207, compared to K-Medoids with a Silhouette Score of 0.2914 and a DBI of 1.2987. Based on the clustering results, retailers were categorized into three main segments: large retailers (10 units), medium retailers (51 units), and small retailers (73 units), each provided with specific business strategy recommendations. This study concludes that the HAC algorithm is more effective in mapping retailer consumption patterns to support data-driven decision-making.*

*Keywords: Clustering, Consumption Patterns, Pesticide, Hierarchical Agglomerative Clustering (HAC), K-Medoids.*

## DAFTAR ISI

HALAMAN PENGANTAR .....	ii
LEMBAR PENGESAHAN .....	iii
PERNYATAAN ORISINALITAS .....	<b>Error! Bookmark not defined.</b>
RIWAYAT HIDUP.....	v
KATA PENGANTAR .....	vii
ABSTRAK .....	i
<i>ABSTRACT</i> .....	ii
DAFTAR ISI.....	iii
DAFTAR TABEL.....	v
DAFTAR GAMBAR .....	vi
BAB I.....	1
PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	4
BAB II.....	5
LANDASAN TEORI.....	5
2.1. Produk Pestisida dan <i>Retailer</i> .....	5
2.1.1 <i>Retail</i> dan <i>Retailer</i> .....	5
2.1.2 Produk Pestisida .....	5
2.1.3 Pola Konsumsi <i>Retailer</i> .....	6
2.2 Knowledge Discovery in Databases (KDD).....	8
2.2.1. Tahapan Knowledge Discovery in Databases (KDD) .....	9
2.3. Data Mining.....	10
2.4. Clustering .....	11
2.5. Algoritma Clustering.....	12
2.5.1. Hierarchical Agglomerative Clustering (HAC).....	12
2.5.2. K-Medoids dan Algoritma PAM .....	13

2.6. Metrik Evaluasi Kualitas Cluster.....	14
2.6.1. Silhouette Score .....	15
2.6.2. Davies-Bouldin Index (DBI) .....	15
2.7. Ringkasan Penelitian Terdahulu.....	16
BAB III.....	18
METEDOLOGI PENELITIAN .....	18
3.1. Jenis Penelitian .....	18
3.2. Teknik Pengumpulan Data .....	18
3.3. Arsitektur Penelitian.....	19
3.3. Tahapan Penelitian .....	22
3.3.1. Preprocessing Data .....	22
3.3.2. Penentuan Variabel.....	22
3.4. Proses Algoritma .....	22
3.4.1. Algoritma Hierarchical Agglomerative Clustering (HAC) .....	23
3.4.2 Algoritma K-Medoids.....	26
3.5. Evaluasi Kualitas Cluster .....	26
3.5.1. Silhouette Score .....	27
3.5.2. Davies-Bouldin Index.....	28
BAB IV .....	29
HASIL DAN PEMBAHASAN.....	29
4.1 Deskripsi Data .....	29
4.2 Hasil Pra-Pemrosesan dan Transformasi Data .....	31
4.2.1. Data Cleaning (Pembersihan Data).....	31
4.2.2 Pra-Pemrosesan.....	33
4.2.3. Tampilan Halaman Clustering.....	42
4.2.4. Tampilan Halaman Dokumentasi <i>Clustering</i> (Visualisasi Dendrogram HAC).....	47
BAB V KESIMPULAN DAN SARAN.....	52
5.1 Kesimpulan.....	52
5.2. Saran.....	53
DAFRAT PUSTAKA .....	54

## DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian Terdahulu .....	16
Tabel 3.1 Dataset Transaksi Penjualan Pestisida pada Toko Aneka Jaya Agro .....	19
Tabel 4.1 Dataset Sebelum Diolah .....	29
Tabel 4.2 Fitur yang digunakan .....	30
Tabel 4.3 Atribut yang Digunakan .....	30
Tabel 4.4 Statistik Ringkasan Data Cleaning .....	33
Tabel 4.5 Contoh data pada retailer Ade .....	35
Tabel 4.6 hasil perhitungan manual sesuai dengan diweb .....	35
Tabel 4.7 hasil perhitungan manual sesuai dengan diweb .....	37
Tabel 4.8 Clustering Data Scaled Ade .....	41

## DAFTAR GAMBAR

Gambar 3.1 Arsitektur Penelitian.....	20
Gambar 4.1 Halaman Pembersihan data .....	32
Gambar 4.2 Halaman Hasil Prapemrosesan.....	33
Gambar 4.3 code hasil bagian Retailer Aggregated Data .....	34
Gambar 4.4 Halaman hasil Clustering data Original diWeb.....	36
Gambar 4.5 Code Hasil Clustering data Original .....	36
Gambar 4.6 Tampilan Hasil PreProcessing Meta Data dari Web.....	38
Gambar 4.7 Code Hasil PreProcessing Meta Data.....	39
Gambar 4.8 Tampilan Hasil Clustering Data Scaled dari Web .....	40
Gambar 4.9 Tampilan Halaman Hasil Analisis Clustering .....	42
Gambar 4.10 Hasil Clustering Distribusi HAC dan K-Medoids .....	44
Gambar 4.11 Hasil Matrik Evaluasi Silhouette Score dan Davies-Bouldin Index	47
Gambar 4.12 Visualisasi dendrogram Hac.....	48
Gambar 4.13 Sebaran Data per Cluster.....	49
Gambar 4.14 Visualisasi Metrik Evaluasi.....	50
Gambar 4.15 Distribusi Cluster.....	51
Gambar 4.16 Tampilan Rekomendasi dan Insight Hasil Clustering.....	52

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sektor pertanian merupakan pilar ekonomi nasional yang sangat bergantung pada efektivitas input produksi, salah satunya adalah pestisida. Dalam rantai distribusi, *retailer* memegang peran vital sebagai jembatan utama antara distributor dan petani. Pengelolaan rantai pasok yang efisien idealnya didasarkan pada data konsumsi *retailer* yang akurat agar ketersediaan stok dapat dijamin tepat waktu dan tepat jumlah ((John, Shobayo and Ogunleye, 2023).

Namun, kenyataan di lapangan pada CV. Aneka Jaya Tunas Agro menunjukkan bahwa data transaksi yang terkumpul dari ratusan transaksi setiap bulan sering kali hanya menjadi catatan administratif yang pasif. Padahal, volume data transaksi yang besar sebenarnya menyimpan pola perilaku konsumsi yang sangat berharga untuk memahami dinamika pasar (Noval et al., 2025). Saat ini, pengambilan keputusan distribusi di tingkat distributor masih cenderung bersifat subjektif atau asumtif, yang berisiko menyebabkan masalah operasional seperti *overstock* atau *stockout* pada produk-produk tertentu.

Meskipun populer untuk segmentasi, algoritma K-Means sering terkendala sensitivitas terhadap outlier dan ketergantungan pada centroid imajiner (Pireddu *et al.*, 2024). Menanggapi keterbatasan tersebut, penelitian ini menerapkan pendekatan yang lebih robust melalui Hierarchical Agglomerative Clustering (HAC) yang mampu memetakan struktur hierarki data secara sistematis (Oti and Olusola, 2024), serta K-Medoids (PAM) yang memberikan akurasi lebih baik karena menggunakan titik data nyata sebagai pusat kluster (Pireddu *et al.*, 2024).

Oleh karena itu, penelitian berjudul "Pengelompokan Produk Pestisida Berdasarkan Pola Konsumsi Retailer Menggunakan Algoritma HAC dan K-Medoids" menjadi sangat krusial. Urgensi penelitian ini terletak pada kebutuhan mendesak bagi CV. Aneka Jaya Tunas Agro untuk mentransformasi data mentah menjadi *insight* bisnis yang strategis. Dengan membandingkan algoritma HAC dan *K-Medoids*, penelitian ini bertujuan untuk menemukan model segmentasi yang paling akurat, yang nantinya digunakan untuk memetakan *retailer* ke dalam kategori kebutuhan (kecil, menengah, besar). Hasil segmentasi ini akan menjadi landasan empiris bagi distributor dalam merancang program loyalitas, optimasi stok, serta strategi distribusi yang berbasis bukti data (*data-driven*), guna meningkatkan efisiensi operasional secara berkelanjutan.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, permasalahan dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana mengimplementasikan algoritma *Hierarchical Agglomerative Clustering* (HAC) dan *K-Medoids* untuk mengelompokkan pola konsumsi *retailer* pada Toko Aneka Jaya Tunas Agro?
2. Bagaimana perbandingan performa antara algoritma *Hierarchical Agglomerative Clustering* (HAC) dan *K-Medoids* dalam melakukan segmentasi produk pestisida dilihat dari nilai *Silhouette Score* dan *Davies-Bouldin Index* (DBI)?
3. Bagaimana menghasilkan rekomendasi strategi bisnis yang efektif bagi pihak distributor berdasarkan hasil pengelompokan pola konsumsi *retailer*?

### 1.3 Batasan Masalah

Untuk menjaga fokus dan kedalaman penelitian, batasan masalah ditetapkan sebagai berikut:

1. Data yang digunakan merupakan data transaksi penjualan pestisida tahun 2023–2024 yang bersumber dari basis data internal Toko Aneka Jaya Tunas Agro.
2. Variabel utama dalam analisis klasterisasi dibatasi pada jenis pestisida, jenis kemasan, dan volume pembelian produk.
3. Algoritma yang diterapkan adalah *Hierarchical Agglomerative Clustering* (HAC) dengan *Ward's Linkage* serta *K-Medoids* dengan metode *Partitioning Around Medoids* (PAM).
4. Evaluasi kualitas kluster hanya menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index*.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun sistem pendukung keputusan yang mampu melakukan pengelompokan pola konsumsi *retailer* secara otomatis dan sistematis.
2. Menganalisis serta membandingkan akurasi performa antara algoritma HAC dan *K-Medoids* dalam melakukan segmentasi pasar produk pestisida.
3. Memberikan wawasan (*insight*) strategis mengenai karakteristik *retailer* untuk mendukung pengambilan keputusan dalam manajemen stok dan optimalisasi distribusi produk

## 1.5. Manfaat Penelitian

### 1. Manfaat bagi Mitra

Penelitian ini membantu CV. Aneka Jaya Tunas Agro mengoptimalkan distribusi dan manajemen stok melalui segmentasi *retailer* yang objektif, sehingga operasional lebih efisien serta strategi pemasaran menjadi lebih tepat sasaran berdasarkan pola konsumsi riil.

### 2. Manfaat bagi Universitas

Bagi mitra, yaitu distributor produk pestisida, hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih terstruktur mengenai pola konsumsi *retailer* terhadap produk pestisida. Informasi pengelompokan produk berdasarkan kemiripan pola konsumsi tersebut dapat digunakan sebagai bahan pertimbangan dalam perencanaan distribusi dan pengelolaan produk secara lebih efektif.

### 3. Manfaat bagi Mahasiswa

Penulis memperoleh peningkatan kompetensi teknis dalam pengolahan data menggunakan Python dan algoritma clustering, mengasah kemampuan analisis dalam pemecahan masalah bisnis.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1. Produk Pestisida dan *Retailer***

##### **2.1.1 *Retail* dan *Retailer***

*Retail* merupakan kegiatan penyaluran produk dari distributor atau produsen kepada konsumen akhir yang berperan sebagai penghubung utama dalam rantai pasok. Selain melakukan penjualan, *retailer* juga berperan dalam pengambilan keputusan terkait jenis produk, jumlah persediaan, dan waktu pembelian berdasarkan kebutuhan konsumen dan kondisi pasar.

Dalam konteks industri *retail*, data transaksi yang dihasilkan dari aktivitas pembelian pelanggan memiliki nilai strategis karena mencerminkan perilaku konsumsi yang nyata. Penelitian oleh (John, Shobayo and Ogunleye, 2023) menunjukkan bahwa *retailer* modern dihadapkan pada *volume* data transaksi yang besar, sehingga diperlukan pendekatan berbasis data untuk memahami pola pembelian dan karakteristik pasar secara lebih sistematis. Dengan demikian, *retailer* tidak hanya berfungsi sebagai perantara distribusi, tetapi juga sebagai sumber data penting dalam pengelolaan dan pengambilan keputusan berbasis informasi.

##### **2.1.2 Produk Pestisida**

Pestisida merupakan bahan kimia yang digunakan untuk mengendalikan organisme pengganggu tanaman guna menjaga kualitas dan produktivitas hasil pertanian. Berdasarkan fungsinya, pestisida diklasifikasikan menjadi

*insektisida*, fungisida, dan herbisida, di mana masing-masing memiliki karakteristik penggunaan, dosis, dan sasaran yang berbeda.

Dalam pasar retail, produk pestisida tersedia dalam berbagai merek, jenis, kemasan, dan volume, yang disesuaikan dengan kebutuhan pengguna. Penelitian (Chilipweli, Nyanza and Ngowi, 2025) menjelaskan bahwa distribusi dan konsumsi pestisida dalam rantai nilai pertanian dipengaruhi oleh faktor ekonomi, ketersediaan produk, serta kepercayaan pengguna terhadap merek tertentu. Variasi produk pestisida di tingkat retailer mencerminkan perbedaan kebutuhan pasar dan menjadi indikator penting dalam memahami pola konsumsi produk pestisida secara keseluruhan.

### **2.1.3 Pola Konsumsi Retailer**

Pola konsumsi retailer mencerminkan perilaku pembelian berulang yang dipengaruhi oleh jenis, variasi, dan volume produk berdasarkan karakteristik serta permintaan pasar tiap retailer. Oleh karena itu, pola ini perlu dianalisis melalui data transaksi guna memahami kecenderungan perilaku tersebut.

Pemanfaatan data transaksi penjualan memungkinkan pemetaan karakteristik pembelian retailer secara lebih objektif dan terstruktur. Melalui data tersebut, perbedaan kebutuhan antar retailer dapat diidentifikasi berdasarkan intensitas pembelian, keragaman produk, dan skala konsumsi. (John, Shobayo and Ogunleye, 2023) menjelaskan bahwa data transaksi memiliki peran penting dalam menggambarkan perbedaan karakteristik pembelian antar kelompok, sehingga pola konsumsi dapat dipetakan secara sistematis. Informasi mengenai pola konsumsi ini menjadi dasar penting dalam proses pengelompokan, karena retailer dengan karakteristik pembelian yang

serupa cenderung memiliki kebutuhan yang relatif sama dalam konteks distribusi produk.

Analisis pola konsumsi dalam sektor retail dapat dilakukan dengan memanfaatkan data transaksi yang tersimpan dalam basis data. Dengan meningkatnya volume data, pendekatan manual menjadi kurang efektif sehingga diperlukan metode berbasis data mining seperti clustering untuk mengidentifikasi perilaku konsumen. Clustering mampu mengelompokkan data tanpa label ke dalam kelompok yang memiliki karakteristik serupa, sehingga dapat membantu dalam memahami pola pembelian dan segmentasi pelanggan secara lebih sistematis (Wilbert *et al.*, 2023).

Pola konsumsi pelanggan dapat dianalisis melalui data transaksi yang mencerminkan perilaku pembelian, preferensi, serta kebiasaan konsumen. Dalam konteks e-commerce maupun retail, segmentasi pelanggan menjadi langkah penting untuk memahami kebutuhan yang beragam. Dengan melakukan segmentasi, pelanggan dapat dikelompokkan berdasarkan karakteristik tertentu sehingga memudahkan dalam penyusunan strategi pemasaran yang lebih tepat sasaran (Wu *et al.*, 2022).

Analisis pola konsumsi berbasis data transaksi juga dapat dilakukan dengan pendekatan clustering untuk mengelompokkan karakteristik pembelian. Penelitian oleh (Triantara, Raharja and Sarasvananda, 2025) menunjukkan bahwa segmentasi berbasis pola transaksi mampu mengidentifikasi perbedaan perilaku secara lebih objektif melalui pengelompokan data.

## 2.2 Knowledge Discovery in Databases (KDD)

*Knowledge Discovery in Databases* (KDD) didefinisikan sebagai sebuah metodologi sistematis untuk mengidentifikasi pola-pola di dalam data yang bersifat valid, baru, berpotensi bermanfaat, dan pada akhirnya dapat dipahami oleh manusia (Shu dan Ye, 2023). Konsep KDD bukan sekadar pencarian data sederhana, melainkan sebuah proses riset dialektis yang menggabungkan pendekatan deduktif dan induktif untuk mengekstraksi pengetahuan berharga dari basis data berukuran besar (Noval, *et al* 2025). Tujuan utamanya adalah mentransformasikan kumpulan fakta mentah menjadi wawasan strategis yang mendukung pengambilan keputusan organisasi, seperti manajemen stok atau efisiensi operasional (Noval, *et al* 2025).

Dalam konteks keilmuan, KDD sering kali dikaitkan erat dengan *data mining*, namun keduanya memiliki batasan fungsional yang berbeda. KDD merupakan kerangka kerja atau proses menyeluruh yang mencakup persiapan data hingga evaluasi hasil, sedangkan *data mining* adalah salah satu tahapan inti di dalam siklus tersebut (Shu dan Ye, 2023). Dengan kata lain, *data mining* berperan sebagai teknik atau alat eksekusi untuk menemukan pola tersembunyi, sementara KDD memastikan bahwa seluruh proses penemuan pengetahuan berjalan secara terstruktur dan sesuai dengan domain masalah yang diteliti (Noval, *et al* 2025).

Tahapan dalam KDD umumnya terdiri dari data selection, data preprocessing, data transformation, data mining, serta interpretation dan evaluation. Tahap prapemrosesan menjadi bagian penting karena bertujuan meningkatkan kualitas data dengan menangani data yang tidak lengkap, tidak konsisten, dan memiliki skala yang berbeda. Data yang telah melalui proses prapemrosesan dan transformasi

yang baik akan menghasilkan pola yang lebih akurat dan mudah diinterpretasikan pada tahap data mining (Luh *et al.*, 2022).

### 2.2.1. Tahapan Knowledge Discovery in Databases (KDD)

Proses KDD terdiri dari serangkaian tahapan iteratif yang harus dilakukan secara berurutan guna menjamin kualitas pengetahuan yang dihasilkan. Menurut (Noval, *et al* 2025), tahapan-tahapan tersebut adalah sebagai berikut:

1. Seleksi Data (*Data Selection*): Tahap awal ini melibatkan pemilihan atribut atau variabel yang relevan dari basis data operasional sesuai dengan tujuan penelitian (Noval, *et al* 2025). Data yang terpilih kemudian dipisahkan untuk menjadi dataset target agar tidak mengganggu integritas data asli (Shu dan Ye, 2023).
2. Prapemrosesan (*Preprocessing/Cleaning*): Langkah ini bertujuan untuk membersihkan dataset dari derau (*noise*), data yang tidak konsisten, serta menangani nilai yang kosong (*missing values*) (Shu dan Ye, 2023). Kualitas data pada tahap ini sangat menentukan akurasi hasil analisis akhir (Noval, *et al* 2025).
3. Transformasi (*Transformation*): Data yang telah bersih diubah ke dalam format yang sesuai untuk algoritma penambangan. Proses ini mencakup normalisasi skala (seperti penggunaan *MinMaxScaler*) dan agregasi data guna menyederhanakan representasi informasi (Noval, *et al* 2025).
4. Penambangan Data (*Data Mining*): Tahapan inti di mana teknik atau algoritma tertentu, seperti *K-Means Clustering*, diterapkan untuk mengekstraksi pola-pola menarik atau keteraturan tersembunyi dari data

yang telah ditransformasi (Noval, *et al* 2025).

5. Interpretasi dan Evaluasi (*Interpretation/Evaluation*): Hasil dari penambangan data divalidasi dan diterjemahkan menjadi informasi yang bermakna. Evaluasi dapat dilakukan secara kuantitatif menggunakan metrik seperti *Silhouette Score* untuk menguji validitas kluster sebelum hasil tersebut diimplementasikan dalam tindakan nyata (Noval, *et al* 2025).

### 2.3. Data Mining

*Data mining* merupakan disiplin ilmu yang berfokus pada proses penggalian informasi berharga atau penemuan pola baru dari koleksi data berskala besar (Rizky *et al.*, 2025). Secara teknis, *data mining* memanfaatkan teknik statistik, matematika, hingga kecerdasan buatan untuk mengidentifikasi hubungan antar-variabel yang sebelumnya tidak diketahui secara manual (Noval, *et al* 2025). Fungsi utama dari pendekatan ini adalah untuk menghasilkan model prediktif atau deskriptif yang mampu memberikan nilai tambah bagi organisasi dalam memahami tren masa depan berdasarkan data historis (Ahmad *et al.*, 2023).

Selain itu, (Tsiu *et al.*, 2025) menegaskan bahwa data mining berperan penting dalam meningkatkan kualitas pengambilan keputusan organisasi melalui pemanfaatan data transaksi yang sebelumnya hanya berfungsi sebagai arsip. Dengan pendekatan berbasis data mining, organisasi dapat memahami pola perilaku, efisiensi operasional, serta karakteristik pasar secara lebih terstruktur.

Pemanfaatan data mining pada data transaksi memungkinkan pengelompokan pola pembelian berdasarkan kemiripan karakteristik transaksi. (Kajornkasirat *et al.*, 2025) menunjukkan bahwa data transaksi ritel dapat dimanfaatkan untuk

mengelompokkan pelanggan atau produk berdasarkan pola pembelian menggunakan teknik clustering dan association rule mining.

#### **2.4. Clustering**

Clustering adalah algoritma unsupervised learning yang membagi data tak berlabel ke dalam kelompok-kelompok berdasarkan tingkat kemiripannya (Syahra *et al.*, 2025). Objek dalam satu kluster memiliki homogenitas internal yang tinggi, sedangkan antar kluster memiliki heterogenitas eksternal yang kuat (Oti and Olusola, 2024). Tujuan utama clustering adalah mengungkap struktur tersembunyi dalam (Rizky *et al.*, 2025). Algoritma ini bekerja dengan meminimalkan variansi di dalam kluster dan memaksimalkan pemisahan antar kluster.

Menurut (Pitafi, Anwar and Sharif, 2023) clustering digunakan untuk menemukan struktur alami dalam data dengan memanfaatkan ukuran kemiripan atau jarak antar data. Metode clustering secara umum dapat diklasifikasikan ke dalam beberapa 11 kategori utama, seperti partitional clustering dan hierarchical clustering, yang masing-masing memiliki mekanisme pembentukan kelompok yang berbeda.

Selain itu, (Farahnakian *et al.*, 2023) menjelaskan bahwa clustering banyak diterapkan dalam berbagai bidang karena kemampuannya dalam mengungkap pola tersembunyi pada data berukuran besar. Penerapan clustering memungkinkan pemahaman yang lebih baik terhadap struktur data dan hubungan antar objek, terutama ketika data tidak memiliki label kelas yang jelas.

Penerapan clustering dalam data transaksi retail telah banyak digunakan untuk melakukan segmentasi pelanggan berdasarkan perilaku pembelian. Teknik ini memungkinkan pengelompokan data berdasarkan kemiripan karakteristik seperti

frekuensi pembelian, jumlah transaksi, dan nilai transaksi. Selain itu, kualitas hasil clustering dapat dievaluasi menggunakan berbagai metrik validasi seperti Silhouette Score dan Davies-Bouldin Index untuk memastikan keakuratan pengelompokan (Wilbert *et al.*, 2023).

Clustering merupakan salah satu teknik data mining yang banyak digunakan dalam segmentasi pelanggan karena mampu mengelompokkan data berdasarkan tingkat kemiripan tanpa memerlukan label. Dalam penelitian terkait segmentasi pelanggan, clustering digunakan untuk mengidentifikasi kelompok pelanggan dengan karakteristik konsumsi yang berbeda sehingga dapat mendukung pengambilan keputusan bisnis secara lebih efektif (Wu *et al.*, 2022).

## **2.5. Algoritma Clustering**

### **2.5.1. Hierarchical Agglomerative Clustering (HAC)**

Hierarchical Agglomerative Clustering (HAC) menggunakan pendekatan bawah-ke-atas (bottom-up) di mana setiap titik data dianggap sebagai kluster tunggal yang kemudian digabungkan secara iteratif berdasarkan kesamaan (Oti and Olusola, 2024). Jarak antar kluster ditentukan melalui kriteria keterhubungan (linkage criteria) seperti Single Linkage, Complete Linkage, Average Linkage, dan Ward's Linkage yang meminimalkan peningkatan total varians (Oti and Olusola, 2024). HAC menghasilkan diagram pohon yang disebut dendrogram untuk memvisualisasikan hierarki pengelompokan.

Selain itu, (Dhulipala *et al.*, 2022) menegaskan bahwa HAC banyak digunakan dalam berbagai bidang karena sifatnya yang intuitif dan kemampuannya dalam menghasilkan cluster dengan kualitas yang baik

berdasarkan fungsi linkage tertentu, seperti single linkage, complete linkage, dan average linkage.

### **2.5.2. K-Medoids dan Algoritma PAM**

K-Medoids adalah algoritma partisi yang lebih tangguh terhadap outlier dibandingkan K-Means karena menggunakan objek nyata dari dataset (medoid) sebagai pusat kluster (Pireddu *et al.*, 2024). Implementasi paling populer adalah algoritma Partitioning Around Medoids (PAM), yang menukar medoid secara iteratif untuk meminimalkan total disimilaritas. Perkembangan terbaru seperti (Mathelin *et al.*, 2025) telah meningkatkan efisiensi waktu komputasi secara signifikan tanpa mengurangi kualitas kluster. Selain itu, metode KluSIM memungkinkan percepatan langkah swap melalui pemangkasan ruang pencarian (Teixeira *et al.*, 2024).

K-Medoids merupakan salah satu metode clustering partisi yang bertujuan membagi data ke dalam sejumlah kelompok berdasarkan tingkat kemiripan antar objek. Berbeda dengan K-Means yang menggunakan nilai rata-rata sebagai pusat cluster, K-Medoids menggunakan objek aktual dari data sebagai pusat cluster atau medoid. Pendekatan ini membuat hasil clustering lebih mudah diinterpretasikan karena setiap medoid merepresentasikan data nyata dalam kelompoknya (Tiwari *et al.*, 2023).

Menurut (Garcia-Morato *et al.*, 2024) K-Medoids memiliki keunggulan dalam fleksibilitas penggunaan fungsi jarak karena tidak terbatas pada jarak Euclidean. Hal ini memungkinkan algoritma ini diterapkan pada berbagai jenis data, termasuk data berdimensi tinggi, data non-numerik, maupun data dengan jarak khusus. Oleh karena itu, K-Medoids banyak digunakan dalam 13

pengelompokan data transaksi, data perilaku, serta data dengan potensi outlier yang tinggi.

Keterkaitan antara HAC dan K-Medoids dalam penelitian ini terletak pada komplementaritasnya. Peneliti menggunakan HAC untuk mengidentifikasi struktur hierarki dan kedekatan antar produk pestisida, sementara K-Medoids diterapkan untuk mendapatkan segmentasi yang lebih robust terhadap data transaksi ekstrem (outlier). Penggunaan medoid sebagai pusat kluster memberikan interpretasi yang lebih nyata bagi distributor pestisida dibandingkan centroid K-Means yang hanya berupa nilai rata-rata imajiner.

Algoritma K-Medoids merupakan salah satu metode clustering yang efektif dalam mengelompokkan data transaksi karena menggunakan objek nyata sebagai pusat kluster (medoid). Pendekatan ini membuat K-Medoids lebih tahan terhadap outlier dibandingkan dengan K-Means. Dalam penelitian pada sektor retail, algoritma ini terbukti mampu menghasilkan segmentasi pelanggan yang lebih stabil dan akurat berdasarkan pola transaksi, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan strategis (Agustin *et al.*, 2025).

## **2.6. Metrik Evaluasi Kualitas Cluster**

Evaluasi hasil clustering menjadi tahap penting untuk memastikan kualitas kluster yang terbentuk. Beberapa metrik yang umum digunakan adalah Silhouette Score yang mengukur tingkat kesamaan dalam kluster serta Davies-Bouldin Index yang menilai pemisahan antar kluster. Penggunaan metrik evaluasi ini membantu

dalam menentukan jumlah kluster optimal dan meningkatkan akurasi segmentasi data (Agustin *et al.*, 2025).

### 2.6.1. Silhouette Score

Silhouette Score mengukur seberapa baik sebuah objek ditempatkan dalam klusternya dibandingkan dengan kluster terdekat lainnya (Syahra *et al.*, 2025). Nilainya berkisar antara -1 hingga +1, di mana nilai mendekati +1 menunjukkan pemisahan yang tegas (John, Shobayo dan Ogunleye, 2023).

Rumus Silhouette Score  $s(i)$  adalah:

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

di mana  $a(i)$  adalah rata-rata jarak intra-kluster dan  $b(i)$  adalah rata-rata jarak inter-kluster terkecil (Syahra *et al.*, 2025)

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

Davies-Bouldin Index (DBI) menilai kualitas berdasarkan rasio dispersi intra-kluster terhadap pemisahan antar kluster (Syahra *et al.*, 2025). Nilai yang lebih kecil mendekati 0 mengindikasikan pengelompokan yang lebih baik karena kluster yang terbentuk bersifat padat dan terpisah jauh satu sama lain (Luh *et al.*, 2022).

## 2.7. Ringkasan Penelitian Terdahulu

Tabel 2.1 Ringkasan Penelitian Terdahulu

No	Judul dan Peneliti	Pembahasan	Metode	Kelebihan dan Kekurangan
1	Customer Segmentation Using RFM and K-Means Clustering to Support CRM in Retail Industry (Syahra e all 2025)	Melakukan segmentasi pelanggan ritel lokal menjadi 4 kelompok (Loyal, Potential, At-Risk, One-Time) untuk mendukung pengembangan strategi CRM yang terukur.	Analisis RFM, Algoritma K-Means, Elbow Method, dan normalisasi Min-Max Scaling.	Kelebihan: Menghasilkan wawasan bagi UKM dengan infrastruktur terbatas. Kekurangan: Berpotensi bias karena hanya memakai data transaksi statis.
2	Penerapan K-Means untuk Menentukan Pola Penjualan di Zahra Mart (Syafri et al., 2025)	Mengidentifikasi produk laris dan tidak laris berdasarkan data transaksi untuk mencegah penumpukan stok barang di gudang.	KDD (Knowledge Discovery in Database), Algoritma K-Means, dan Jarak Euclidean.	Kelebihan: Website memudahkan pemantauan stok secara real-time. Kekurangan: Sensitif terhadap outlier karena menggunakan K-Means.
3	An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market (John, et al 2023)	Eksplorasi berbagai algoritma untuk memahami perilaku pembelian pelanggan di pasar ritel online Inggris guna meningkatkan profitabilitas.	RFM Framework, K-Means, GMM, DBSCAN, (HAC), BIRCH, dan PCA.	Kelebihan: Perbandingan algoritma komprehensif; GMM paling akurat (Silhouette 0.80). Kekurangan: Komputasi tinggi pada dataset besar.

No	Judul dan Peneliti	Pembahasan	Metode	Kelebihan dan Kekurangan
4	Customer Segmentation with Association Rule Mining on Retail Transaction Data (Kajornkasirat et al., 2025)	Menggabungkan segmentasi pelanggan dan aturan asosiasi produk untuk meningkatkan efisiensi stok dan promosi bundling.	RFM Analysis, K-Means, Fuzzy C-Means, SONN, dan FP-Growth.	Kelebihan: Mengungkap hubungan antar produk dalam tiap segmen pelanggan. Kekurangan: Membutuhkan komputasi tinggi pada dataset besar.
5	Clustering Destinasi Wisata di Denpasar dengan K-Means dan DBSCAN (Dharmawan et al., 2025)	Mengelompokkan destinasi berdasarkan preferensi kunjungan & fasilitas untuk mendukung kebijakan pengembangan wilayah.	K-Means, DBSCAN, Elbow Method, evaluasi Silhouette Score.	Kelebihan: Dashboard interaktif memudahkan interpretasi data. Kekurangan: Hasil bergantung pada parameter epsilon dan min_samples.

## **BAB III**

### **METEDOLOGI PENELITIAN**

#### **3.1. Jenis Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif yang berfokus pada pengolahan dan analisis data numerik untuk menghasilkan informasi yang objektif. Dalam penelitian ini dilakukan pengujian atau percobaan menggunakan metode komputasi dengan menerapkan teknik data mining pada data pola konsumsi produk pestisida oleh *retailer*. Data yang telah dikumpulkan kemudian diolah dan dianalisis untuk menemukan pola atau karakteristik konsumsi produk berdasarkan kesamaan perilaku pembelian.

Fokus penelitian ini adalah menerapkan dua algoritma *clustering*, yaitu *Hierarchical Agglomerative Clustering* (HAC) dan K-Medoids, untuk mengelompokkan produk berdasarkan pola konsumsi retailer. Hasil pengelompokan dari kedua algoritma tersebut kemudian dibandingkan untuk melihat performa masing-masing metode dalam menghasilkan cluster yang efektif. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan gambaran pola konsumsi produk serta menjadi dasar pendukung dalam pengambilan keputusan terkait strategi pengelolaan dan distribusi produk.

#### **3.2. Teknik Pengumpulan Data**

Penelitian ini menggunakan teknik dokumentasi dengan mengumpulkan data transaksi penjualan pestisida tahun 2023–2024 yang diperoleh dari database internal Toko Aneka Jaya Tunas Agro. Data yang digunakan mencakup beberapa atribut penting seperti tahun dan bulan transaksi, nama retailer, wilayah/kabupaten, nama produk, jenis pestisida, kemasan, serta volume pembelian. Data tersebut

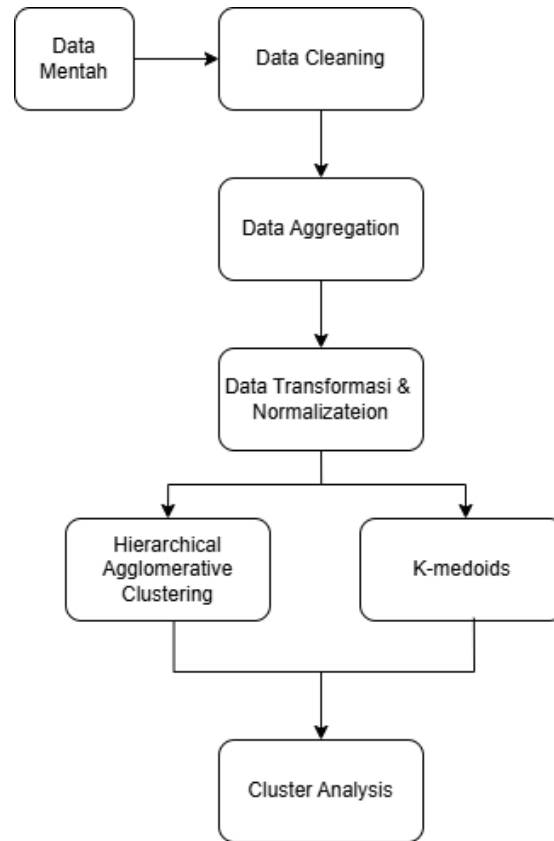
kemudian digunakan sebagai bahan utama dalam proses analisis untuk mengelompokkan pola konsumsi produk pestisida menggunakan algoritma Hierarchical Agglomerative Clustering (HAC) dan K-Medoids.

### 3.3. Arsitektur Penelitian

Tabel 3.1 Dataset Transaksi Penjualan Pestisida pada Toko Aneka Jaya Tunas Agro

Year	Motnh	Distributor	Area	Retailer Name(R1)	Address R1	Rregncy/Sistrict	Product	Jenis Pestisida	Packaging	Volume L/Kg
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	AHLI TANI 13 (BPK MAHMUD)	Panimbang, Pandeglang	Pandeglang	Adore 200 sc	Fungisida	250 ml	7.5
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	AHLI TANI 13 (BPK MAHMUD)	Panimbang, Pandeglang	Pandeglang	Avjunsu 18 ec	Insektisida	500 ml	20
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	AHLI TANI 13 (BPK MAHMUD)	Panimbang, Pandeglang	Pandeglang	Avjunsu 18 ec	Insektisida	100 ml	.0
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	PUTRA TANI (BPK IDIK)	Sukabumi	Sukabumi	Adore 200 sc	Fungisida	250 ml	10.0
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	PUTRA TANI (BPK IDIK)	Sukabumi	Sukabumi	Avjunsu 18 ec	Insektisida	500 ml	15.0
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	PUTRA TANI (BPK IDIK)	Sukabumi	Sukabumi	Avjunsu 18 ec	Insektisida	250 ml	50.0
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	PUTRA TANI (BPK IDIK)	Sukabumi	Sukabumi	Et-Trazol 250 sc	Zat Pengatur Tumbuh	1 L	20.0
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	MINI JAYA TANI (BPK TARMIN)	Bojen, Pandeglang	Pandeglang	Et- Upstar 480 SI	Herbisida	1 L	10.0
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	MASTANI (SONNY)	Serang	Serang	Avjunsu 18 ec	Insektisida	100 ml	100.0
30.02023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	Bina Tani (RIFKI)	Sawah Luhur, Serang	Serang	Adore 200 sc	Fungisida	250 ml	30.0
2023	01.Jan	CV. Aneka Jaya Tunas Agro	West Java	Bina Tani (RIFKI)	Sawah Luhur, Serang	Serang	Adore 200 sc	Fungisida	250 ml	5.0

Arsitektur penelitian adalah rangkaian proses sistematis yang dilakukan untuk menganalisis data guna menjawab pertanyaan penelitian. Alur pengolahan data dalam penelitian ini dimulai dari database mentah hingga didapatkan hasil akhir, dengan tahapan sebagai berikut:



Gambar 3.1 Arsitektur Penelitian

Flowchart tersebut menggambarkan alur pengolahan data yang dilakukan secara bertahap mulai dari data mentah hingga menghasilkan analisis cluster. Adapun tahapan-tahapannya adalah sebagai berikut:

1. Data Mentah

Tahap awal berupa data mentah yang diperoleh dari sumber data. Data ini masih belum terstruktur dan belum siap untuk dianalisis.

2. Data Cleaning

Pada tahap ini dilakukan pembersihan data, seperti menghapus data duplikat, menangani data yang kosong (missing value), serta memperbaiki inkonsistensi data agar kualitas data menjadi lebih baik.

3. Data Aggregation

Data yang telah dibersihkan kemudian diagregasi atau diringkas. Proses ini bertujuan untuk mengelompokkan data berdasarkan kebutuhan analisis, misalnya merangkum data penjualan per retailer sehingga lebih representatif.

#### 4. Data Transformasi & Normalization

Selanjutnya dilakukan transformasi dan normalisasi data. Transformasi bertujuan mengubah format data agar sesuai dengan kebutuhan algoritma, sedangkan normalisasi dilakukan untuk menyamakan skala data agar tidak terjadi bias dalam perhitungan.

#### 5. Proses Clustering

Pada tahap ini digunakan dua metode clustering, yaitu:

1. Hierarchical Agglomerative Clustering (HAC): metode pengelompokan yang membentuk cluster secara bertahap dari bawah ke atas berdasarkan kedekatan data.
2. K-Medoids: metode clustering yang menggunakan objek data sebagai pusat cluster (medoid) dan lebih tahan terhadap outlier.

#### 6. Cluster Analysis

Tahap akhir adalah analisis hasil clustering. Pada tahap ini dilakukan interpretasi terhadap setiap cluster yang terbentuk untuk mengetahui karakteristik masing-masing kelompok data, seperti kategori retailer berdasarkan pola penjualannya.

### 3.3. Tahapan Penelitian

#### 3.3.1. Preprocessing Data

Tahap ini memastikan kualitas data sebelum diolah agar hasil klusterisasi akurat. Tahapan preprocessing meliputi:

1. Data Cleaning: Menghapus data yang tidak lengkap atau memiliki duplikasi.
2. Data Integration: Menggabungkan data transaksi dari berbagai periode.
3. Data Transformation: Mengonversi data ke dalam format yang sesuai untuk analisis.
4. Data Normalization: Melakukan normalisasi dan standarisasi data numerik agar berada dalam skala yang sama, sehingga tidak ada variabel yang mendominasi proses clustering.

#### 3.3.2. Penentuan Variabel

Pada tahap penentuan kriteria dan parameter yang akan digunakan dalam proses clustering. Kriteria pengelompokan ditentukan berdasarkan beberapa variabel yang merepresentasikan pola konsumsi produk pestisida oleh *retailer*, yaitu jenis pestisida, *packaging*, dan *volume* pembelian. Variabel-variabel tersebut digunakan sebagai dasar dalam proses analisis untuk mengidentifikasi kesamaan karakteristik pembelian antar *retailer* sehingga dapat membentuk kelompok dengan pola konsumsi yang serupa.

### 3.4. Proses Algoritma

Penelitian ini menggunakan pendekatan *Unsupervised Learning* dalam memproses data transaksi penjualan pestisida. Berbeda dengan metode *supervised*

yang memerlukan label target, metode *unsupervised* bekerja dengan mengidentifikasi pola tersembunyi atau struktur kelompok secara mandiri berdasarkan karakteristik data yang ada.

Dalam penelitian ini, algoritma yang diimplementasikan adalah *Hierarchical Agglomerative Clustering* (HAC) dan *K-Medoids*. Keduanya dipilih karena kemampuannya dalam melakukan *clustering* (pengelompokan) data tanpa memerlukan label kategori sebelumnya, sehingga sangat efektif untuk menemukan segmentasi pola konsumsi *retailer* yang belum terdefinisi secara jelas.

### 3.4.1. Algoritma Hierarchical Agglomerative Clustering (HAC)

#### 1. Pengukuran Jarak

Dalam penelitian ini, pengukuran kedekatan antar data dilakukan menggunakan *Euclidean Distance*. Metode ini digunakan untuk menghitung jarak antara dua titik data dalam ruang multidimensi berdasarkan nilai setiap variabel yang digunakan dalam proses *clustering*. *Euclidean Distance* dipilih karena mampu mengukur kemiripan data numerik secara langsung dan sering digunakan pada algoritma *clustering*.

Secara matematis, *Euclidean Distance* antara dua data  $x$  dan  $y$  dengan  $n$  variabel dapat dihitung menggunakan persamaan berikut:

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

Keterangan:

1.  $d(x, y)$  = jarak antara data  $x$  dan  $y$
2.  $x_i$  = nilai variabel ke- $i$  pada data  $x$
3.  $y_i$  = nilai variabel ke- $i$  pada data  $y$
4.  $n$  = jumlah variabel yang digunakan

Dalam penelitian ini, perhitungan jarak dilakukan berdasarkan beberapa variabel yang merepresentasikan pola konsumsi *retailer*, yaitu jenis pestisida, *packaging*, serta *volume* pembelian. Semakin kecil nilai jarak antar data maka semakin tinggi tingkat kemiripan karakteristik antar *retailer*.

## 2. Metode Penggabungan

Pada algoritma HAC, proses penggabungan cluster dilakukan menggunakan metode *Ward*. Metode *Ward* bekerja dengan cara menggabungkan cluster yang menghasilkan peningkatan variansi terkecil dalam cluster. Tujuan utama metode ini adalah untuk menghasilkan cluster yang memiliki tingkat homogenitas yang tinggi.

Secara matematis, metode *Ward* meminimalkan nilai *Within Cluster Sum of Squares* (WCSS) yang dapat dituliskan sebagai berikut:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} (x - \mu_i)^2$$

Keterangan:

1.  $k$  = jumlah cluster

2.  $C_i$  = cluster ke-i
3.  $x$  = data dalam cluster
4.  $\mu_i$  = centroid cluster ke-i

Proses penggabungan cluster dilakukan secara bertahap hingga mencapai jumlah cluster yang telah ditentukan.

### 3. Visualisasi Dendrogram

Dalam metode *Hierarchical Agglomerative Clustering*, proses penggabungan cluster dapat divisualisasikan menggunakan dendrogram. Dendrogram merupakan diagram berbentuk pohon yang menggambarkan hubungan hierarki antar data berdasarkan tingkat kemiripan.

Pada dendrogram, setiap objek data pada awalnya berada pada posisi terpisah. Kemudian objek-objek yang memiliki jarak paling dekat akan digabungkan menjadi satu cluster. Proses ini terus berlangsung hingga semua data tergabung dalam satu cluster besar.

Sumbu vertikal pada dendrogram menunjukkan nilai jarak antar cluster, sedangkan sumbu horizontal menunjukkan objek data yang dianalisis. Jumlah cluster dapat ditentukan dengan cara memotong dendrogram pada ketinggian tertentu. Dalam penelitian ini, dendrogram dipotong sehingga menghasilkan tiga cluster utama yang merepresentasikan retailer dengan aktivitas pembelian tinggi, sedang, dan rendah.

### 3.4.2 Algoritma K-Medoids

Selain menggunakan HAC, penelitian ini juga menerapkan algoritma K-Medoids untuk membandingkan hasil pengelompokan data. K-Medoids merupakan metode clustering yang termasuk dalam kategori *partitioning clustering*, dimana data dibagi ke dalam sejumlah cluster yang telah ditentukan sebelumnya.

Pada penelitian ini digunakan metode PAM (*Partitioning Around Medoids*). Berbeda dengan K-Means yang menggunakan centroid sebagai pusat cluster, K-Medoids menggunakan medoid, yaitu titik data yang paling merepresentasikan cluster.

Rumus pemilihan medoid adalah :

$$m = \arg \min_{x_j \in C} \sum_{x_i \in C} d(x_i, x_j)$$

Keterangan:

1.  $m$  = medoid cluster
2.  $C$  = kumpulan data dalam cluster
3.  $d(x_i, x_j)$  = jarak antara dua data

Dalam penelitian ini, jumlah cluster yang digunakan adalah tiga cluster ( $k = 3$ ) dengan menggunakan Euclidean Distance sebagai metode pengukuran jarak.

### 3.5. Evaluasi Kualitas Cluster

Setelah proses clustering dilakukan menggunakan algoritma Hierarchical Agglomerative Clustering (HAC) dan K-Medoids, tahap selanjutnya adalah melakukan evaluasi terhadap kualitas cluster yang dihasilkan. Evaluasi ini bertujuan untuk mengetahui seberapa baik data dikelompokkan dalam cluster yang

terbentuk. Dalam penelitian ini digunakan dua metode evaluasi yaitu *Silhouette Score* dan *Davies-Bouldin Index*.

### 3.5.1. Silhouette Score

*Silhouette Score* digunakan untuk mengukur tingkat kesamaan suatu data terhadap cluster tempat data tersebut berada dibandingkan dengan cluster lainnya. Nilai *Silhouette Score* berada pada rentang -1 hingga 1.

1. Nilai mendekati 1 menunjukkan bahwa data berada pada cluster yang tepat.
2. Nilai mendekati 0 menunjukkan bahwa data berada di antara dua cluster.
3. Nilai mendekati -1 menunjukkan bahwa data kemungkinan salah ditempatkan dalam cluster.

Rumus *Silhouette Score* dapat dituliskan sebagai berikut:

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

Keterangan:

1.  $s(i)$  = nilai silhouette untuk data ke-i
2.  $a(i)$  = rata-rata jarak antara data ke-i dengan seluruh data dalam cluster yang sama
3.  $b(i)$  = rata-rata jarak terkecil antara data ke-i dengan data pada cluster lain

Nilai *Silhouette Score* yang semakin tinggi menunjukkan bahwa cluster yang terbentuk memiliki tingkat pemisahan yang lebih baik.

### 3.5.2. Davies-Bouldin Index

*Davies-Bouldin Index* (DBI) digunakan untuk mengevaluasi kualitas cluster berdasarkan tingkat kesamaan antar cluster dan jarak antar cluster. Metode ini mengukur rasio antara jarak dalam cluster (*intra-cluster distance*) dan jarak antar cluster (*inter-cluster distance*).

Rumus *Davies-Bouldin Index* adalah sebagai berikut:

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

Keterangan:

1.  $k$  = jumlah cluster
2.  $S_i$  = rata-rata jarak data dalam cluster ke pusat cluster
3.  $S_j$  = rata-rata jarak data dalam cluster lainnya
4.  $M_{ij}$  = jarak antara pusat cluster  $i$  dan cluster  $j$

Nilai *Davies-Bouldin Index* memiliki rentang nilai  $\geq 0$ , dimana semakin kecil nilai DBI maka semakin baik kualitas cluster yang dihasilkan. Nilai yang kecil menunjukkan bahwa cluster memiliki jarak antar cluster yang besar dan jarak dalam cluster yang kecil.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Deskripsi Data

Data yang digunakan dalam penelitian ini merupakan data transaksi penjualan produk pestisida yang diperoleh dari distributor Aneka Jaya Tunas Agro. Dataset ini berisi informasi mengenai penjualan produk pestisida kepada retailer pada periode tahun 2023. Data tersebut mencakup informasi mengenai wilayah distribusi, nama retailer, jenis produk pestisida, jenis pestisida, ukuran kemasan, serta volume penjualan produk. Dataset yang digunakan dalam penelitian ini berjumlah 1594 data transaksi dengan 11 atribut yang masih berupa data mentah sehingga perlu dilakukan proses seleksi dan pengolahan data sebelum digunakan dalam proses analisis clustering.

Tabel 4.1 Dataset Sebelum Diolah

Year	Month	DISTRIBUTOR	AREA	RETAILER NAME (R1)	ADDRESS R1	REGION/DISTRIC	PRODUCT	JENIS PESTISIDA	PACKAGING	VOLUME (L/KG)
2023	01 JAN	CV. Aneka Jaya Tunas Agro	West Java	AHLI TANI 13 (BPK MAHMUD)	Panimbang	PANDEGLANG	ADORE 200 SC	Fungisida	250 ML	7.5
2023	01 JAN	CV. Aneka Jaya Tunas Agro	West Java	PUTRA TANI (BPK IDIK)	Sukabumi	Sukabumi	Avzunsu 18 ec	Insektisida	500 ML	15
2024	01 JAN	CV. Aneka Jaya Tunas Agro	West Java	ZAKI TANI (BPK DEDEN)	Ciruas, Serang	SERANG	ET-PRORISO 10/50 OD	Herbisida	200 ML	6
2024	01 JAN	CV. Aneka Jaya Tunas Agro	West Java	HJ. LIM	Beberan	Beberan	AVJUNSU 18 EC	Insektisida	1 L	10.0

Setelah data kriminalitas diambil secara manual, selanjutnya dilakukan penghapusan variabel atau fitur yang tidak diperlukan, sehingga menghasilkan tabel berikut:

Tabel 4.2 Fitur yang digunakan

RETAILER NAME (R1)	JENIS PESTISIDA	PACKAGING	VOLUME (L/KG)
AHLI TANI 13 (BPK MAHMUD)	Fungisida	250 ML	7.5
PUTRA TANI (BPK IDIK)	Insetisida	500 ML	15
ZAKI TANI (BPK DEDEN)	Herbisida	200 ML	6

Selanjutnya dilakukan pemilihan atribut yang mengacu pada variabel-variabel yang relevan dalam proses analisis pengelompokan pola konsumsi produk pestisida oleh retailer. Atribut-atribut yang digunakan dalam penelitian ini dapat dilihat pada Tabel 4.3 berikut.

Tabel 4.3 Atribut yang Digunakan

Atribut	Keterangan
Nama Retailer	Nama retailer yang melakukan pembelian
Jenis Pestisida	Kategori jenis pestisida yang dijual
Packaging	Ukuran kemasan produk pestisida
Volume	Jumlah volume produk yang dibeli

Jumlah data yang digunakan dalam penelitian ini sebanyak 1594 data transaksi penjualan pestisida yang diperoleh dari distributor kepada retailer. Data tersebut memiliki beberapa atribut yang digunakan dalam analisis yaitu nama retailer, jenis pestisida, packaging, dan volume penjualan. Selanjutnya data transaksi tersebut nantinya akan dilakukan proses agregasi berdasarkan pola konsumsi retailer sehingga diperoleh 134 data yang merepresentasikan pola pembelian masing-masing retailer.

Karakteristik pola pembelian retailer dalam penelitian ini dapat dilihat berdasarkan beberapa kondisi berikut:

1. Volume pembelian tinggi dengan variasi produk banyak.
2. Volume pembelian sedang dengan variasi produk cukup.

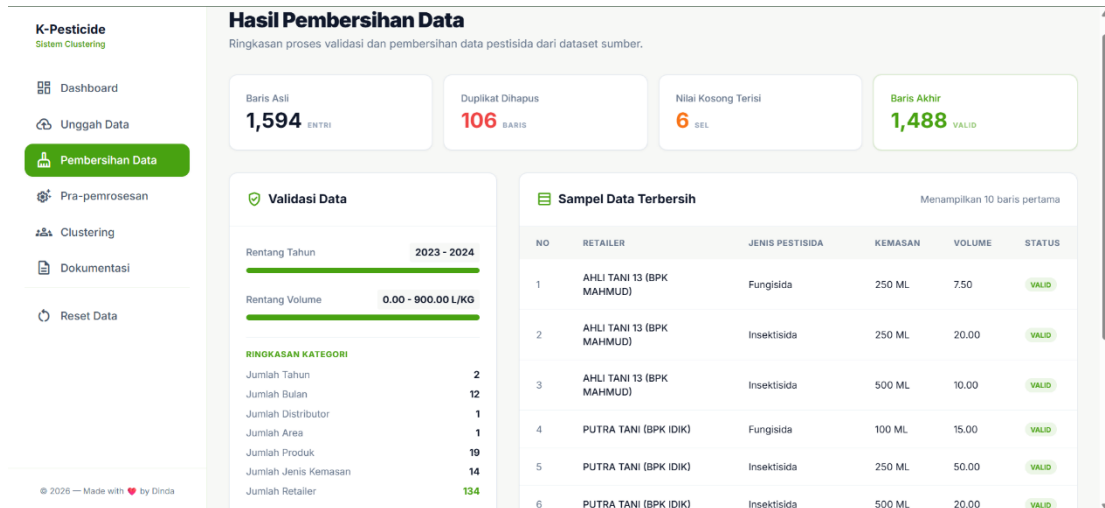
3. Volume pembelian rendah dengan variasi produk sedikit.

Berdasarkan kondisi tersebut, data transaksi penjualan pestisida yang dianalisis menunjukkan bahwa setiap retailer memiliki pola konsumsi yang berbeda terhadap produk pestisida. Perbedaan pola pembelian ini kemudian digunakan sebagai dasar dalam proses analisis untuk mengidentifikasi kelompok retailer berdasarkan tingkat aktivitas pembelian produk pestisida. Hasil analisis ini diharapkan dapat memberikan gambaran mengenai karakteristik retailer sehingga dapat membantu distributor dalam memahami tingkat permintaan produk serta mendukung pengambilan keputusan dalam pengelolaan distribusi dan persediaan produk pestisida.

## **4.2 Hasil Pra-Pemrosesan dan Transformasi Data**

### **4.2.1. Data Cleaning (Pembersihan Data)**

Tahap pertama dalam pemrosesan data adalah pembersihan data (*data cleaning*) untuk memastikan integritas dan kualitas dataset sebelum analisis. Proses ini meliputi penghapusan data duplikat, penanganan nilai yang hilang (*missing values*), serta perbaikan ketidakkonsistenan format data seperti perbedaan penulisan atau spasi. Selain itu, dilakukan juga deteksi dan penanganan nilai yang tidak wajar (*outlier*) agar tidak memengaruhi hasil analisis. Dengan proses ini, data menjadi lebih rapi, konsisten, dan akurat sehingga siap digunakan pada tahap pengolahan dan analisis selanjutnya.



Gambar 4.1 Halaman Pembersihan data

Sistem menjalankan tiga prosedur utama untuk memastikan kualitas data:

1. **Penanganan Duplikasi:** Sistem menggunakan fungsi `drop_duplicates(keep='first')` untuk menghapus data yang redundan. Logikanya, jika terdapat baris dengan nilai identik di semua kolom, hanya baris pertama yang dipertahankan untuk menjaga akurasi distribusi data.
2. **Imputasi *Missing Values*:** Untuk atribut alamat dan wilayah (`ADDRESS R1` dan `REGENCY/DISTRICT`) yang memiliki data kosong, sistem memberikan label 'Unknown' menggunakan fungsi `fillna()`. Langkah ini memastikan data tetap dapat diproses dalam statistik tanpa harus membuang seluruh *record*.
3. **Standarisasi Format:** Spasi yang tidak sengaja tertinggal di awal atau akhir teks pada kolom kategorikal (seperti nama produk atau retailer) dibersihkan menggunakan metode `.str.strip()`. Ini bertujuan agar data yang secara makna sama, tidak terbaca sebagai kategori yang berbeda oleh sistem.

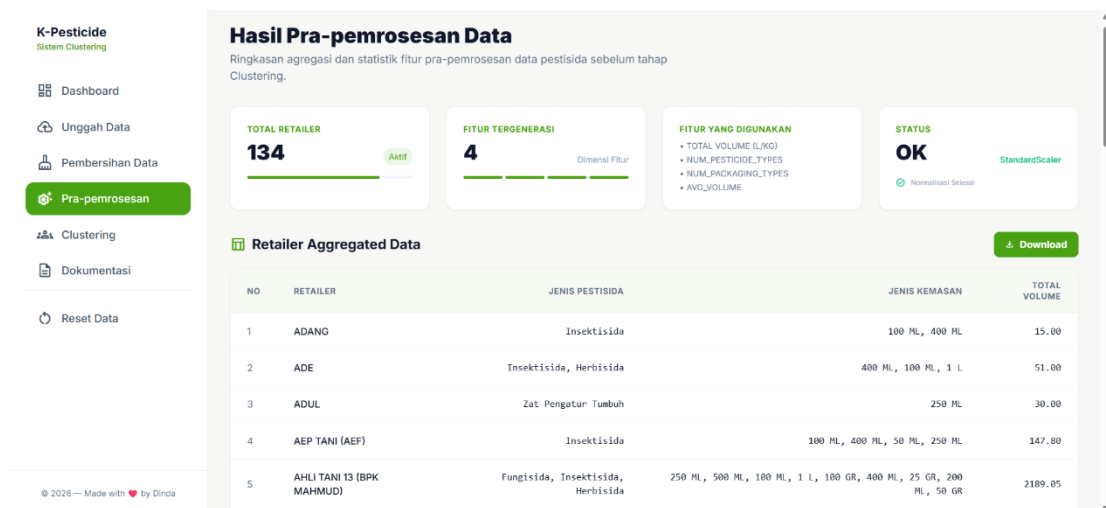
Berikut adalah ringkasan dampak dari proses pembersihan data yang dilakukan oleh sistem:

Tabel 4.4 Statistik Ringkasan Data Cleaning

Deskripsi Metrik	Jumlah
Total Data Sebelum (Original Rows)	1,594
Data Duplikat yang Dihapus	106
Missing Values yang Terisi	6
Total Data Sesudah (Final Rows)	1,488

Berdasarkan statistik di atas, terlihat bahwa proses cleaning berhasil menstandarisasi dataset. Penghapusan data duplikat sebesar 106 baris dan penanganan 6 *missing values* memastikan dataset telah bersih dari redundansi dan ketidakkonsistenan format. Dengan demikian, data dinyatakan layak untuk masuk ke tahap selanjutnya, yaitu Preprocessing dan Agregasi.

#### 4.2.2 Pra-Pemrosesan



Gambar 4 2 Halaman Hasil Prapemrosesan

Halaman Pra-pemrosesan Data menampilkan hasil pengolahan data sebelum proses clustering dilakukan. Informasi yang ditampilkan meliputi Retailer Aggregated Data yang berisi data ringkasan berdasarkan retailer, Clustering Data Original yaitu data dengan fitur yang digunakan dalam clustering namun masih dalam skala asli, PreProcessing Meta Data yang

menampilkan ringkasan informasi proses pra-pemrosesan, serta Clustering Data Scaled yaitu data yang telah dinormalisasi menggunakan metode StandardScaler sehingga siap digunakan pada proses clustering.

#### a. Hasil *Retailer Aggregated Data*

Hasil dari proses agregasi data retailer ditampilkan pada bagian *Retailer Aggregated Data* sebagai berikut :

```
# Group by retailer and aggregate
df_grouped = df_subset.groupby('RETAILER NAME (R1)').agg({
    'JENIS PESTISIDA': lambda x: ', '.join(x.dropna().unique()),
    'PACKAGING': lambda x: ', '.join(x.dropna().unique()),
    'VOLUME (L/KG)': 'sum'
}).reset_index()

df_grouped.columns = ['RETAILER NAME (R1)', 'JENIS PESTISIDA', 'PACKAGING', 'TOTAL VOLUME (L/KG)']
```

Gambar 4.3 code hasil bagian *Retailer Aggregated Data*

Proses agregasi data dilakukan dengan mengelompokkan data berdasarkan nama retailer, kemudian menghitung total volume pembelian serta menggabungkan jenis pestisida dan jenis kemasan yang dimiliki oleh retailer tersebut. Secara matematis proses ini dapat dituliskan sebagai berikut:

$$Total\ Volume_r = \sum_{i=1}^n Volume_i$$

Keterangan :

1.  $Total\ Volume_r$  = total volume pembelian retailer  $r$
2.  $Volume_i$  = volume pembelian pada transaksi ke- $i$
3.  $n$  = jumlah transaksi retailer

Sedangkan atribut kategori dihitung dengan menggabungkan nilai unik:

$$Jenis\ Pestisida_r = Unique(Jenis\ Pestisida_i)$$

$$Packaging_r = Unique(Jenis\ Packaging_i)$$

Misalnya retailer ADE memiliki beberapa transaksi sebagai berikut:

Tabel 4.5 Contoh data pada retailer Ade

Retailer	Jenis Pestisida	Packaging	Volume
ADE	Insektisida	400 ML	20
ADE	Herbisida	100 ML	15
ADE	Insektisida	1 L	16

Maka hasil agregasinya adalah:

1. Jenis Pestisida = Insektisida, Herbisida
2. Packaging = 400 ML, 100 ML, 1 L
3. Total Volume  $Total\ Volume_{ADE} = 20 + 15 + 16 = 51$

Sehingga pada Retailer Aggregated Data akan menjadi:

Tabel 4.6 hasil perhitungan manual sesuai dengan diweb

Retailer	Jenis Pestisida	Packaging	Total Volume
ADE	Insektisida,Herbisida	400ML,100 ML, 1 L	51

## b. Tampilan Hasil *Clustering* data Original

NO	RETAILER	TOTAL VOLUME	JENIS PESTISIDA	JENIS KEMASAN	AVG VOLUME
1	ADANG	15.00	1	2	7.50
2	ADE	51.00	2	3	8.50
3	ADUL	30.00	1	1	30.00
4	AEP TANI (AEF)	147.80	1	4	36.95
5	AHLI TANI 13 (BPK MAHMUD)	2189.05	3	9	81.08
6	AILA TANI (BPK MAKMUR)	522.75	3	8	21.78
7	ALAM MAKMUR (H. IBRAHIM)	186.30	3	7	5.06
8	AM-POWER (BPK ROBI)	5.00	1	1	5.00
9	ANGGITA JAYA (BPK SURYONO)	2438.90	3	9	90.33
10	ANGRI TANI (ADANG)	100.00	1	2	50.00

Gambar 4.4 Halaman hasil Clustering data Original diWeb

Tampilan ini menunjukkan halaman Clustering Data Original yang berisi data awal retailer sebelum dilakukan proses pengelompokan. Data yang ditampilkan meliputi nama retailer, total volume, jumlah jenis pestisida, jumlah jenis kemasan, serta rata-rata volume (*avg volume*). Informasi ini merupakan hasil dari tahap pra-pemrosesan yang telah diubah ke dalam bentuk numerik dan siap digunakan sebagai input dalam proses clustering untuk menentukan kelompok retailer berdasarkan karakteristik yang dimiliki.

Pada tahap ini dilakukan pembentukan fitur numerik yang akan digunakan dalam proses clustering. Fitur tersebut dihitung berdasarkan hasil agregasi data retailer dengan rumus sebagai berikut:

Jumlah jenis pestisida yang dimiliki oleh retailer dihitung dengan:

$$NumPesticideTypes_r = Count (Jenis Pesticida_i)$$

Jumlah jenis kemasan yang digunakan oleh retailer dihitung dengan:

```
# Create a copy for feature engineering
df_clustering = df_grouped.copy()

# Count unique pesticide types per retailer
df_clustering['NUM_PESTICIDE_TYPES'] = df_clustering['JENIS PESTISIDA'].apply(
    lambda x: len(str(x).split(','))
)

# Count unique packaging types per retailer
df_clustering['NUM_PACKAGING_TYPES'] = df_clustering['PACKAGING'].apply(
    lambda x: len(str(x).split(','))
)

# Average volume per transaction (simplified metric)
df_clustering['AVG_VOLUME'] = df_clustering['TOTAL VOLUME (L/KG)] / (
    df_clustering['NUM_PESTICIDE_TYPES'] * df_clustering['NUM_PACKAGING_TYPES']
)

# Fill any NaN with 0
df_clustering = df_clustering.fillna(0)
```

Gambar 4.5 Code Hasil Clustering data Original

$$NumPackagingTypes_r = Count (JPackaging_i)$$

Rata-rata volume dihitung dengan:

$$AvgVolume_r = \frac{TotalVolume_r}{NumPesticideTypes_r \times NumPackagingTypes_r}$$

Keterangan:

1.  $TotalVolume_r$  = total volume pembelian retailer
2.  $NumPesticideTypes_r$  = jumlah jenis pestisida
3.  $NumPackagingTypes_r$  = jumlah jenis kemasan

Misalnya retailer ADE memiliki data hasil agregasi sebagai berikut:

1. Jenis Pestisida = Insektisida, Herbisida
2. Packaging = 400 ML, 100 ML, 1 L
3. Total Volume = 51

Maka diperoleh:

Jumlah jenis pestisida:  $NumPesticideTypes_r = 2$

Jumlah jenis kemasan:  $NumPackagingTypes_r = 3$

Rata-rata volume:  $AvgVolume = \frac{51}{2 \times 3} = 8,5$

Sehingga pada *Clustering* Data Original akan terbentuk fitur seperti berikut:

Tabel 4 7 hasil perhitungan manual sesuai dengan diweb

Retailer	Total Volume	Nums Pesticide Types	Num Packaging Types	AvgVolume
ADE	51	2	3	8.5

### c. Tampilan Hasil PreProcessing Meta Data

NAMA FITUR	MIN	MAX	MEAN	STD DEV	SCALER MEAN	SCALER SCALE
TOTAL VOLUME (L/KG)	0.50	2972.00	219.84	541.41	219.84	539.39
NUM_PESTICIDE_TYPES	1	4	1.64	0.79	1.64	0.79
NUM_PACKAGING_TYPES	1	11	3.61	2.64	3.61	2.63

Gambar 4.6 Tampilan Hasil PreProcessing Meta Data dari Web

Tampilan ini menunjukkan bagian Preprocessing Meta Data yang berisi ringkasan statistik dari setiap fitur yang digunakan dalam proses clustering. Informasi yang ditampilkan meliputi nilai minimum (min), maksimum (max), rata-rata (mean), dan standar deviasi (std dev), serta parameter hasil normalisasi seperti *scaler mean* dan *scaler scale*. Data ini memberikan gambaran distribusi masing-masing fitur seperti total volume, jumlah jenis pestisida, jumlah jenis kemasan, dan rata-rata volume, sekaligus menunjukkan bahwa data telah dipersiapkan dan dinormalisasi sehingga siap digunakan dalam proses clustering.

```

metadata = pd.DataFrame({
    'Feature': clustering_features,
    'Original_Min': [X_features[col].min() for col in clustering_features],
    'Original_Max': [X_features[col].max() for col in clustering_features],
    'Original_Mean': [X_features[col].mean() for col in clustering_features],
    'Original_Std': [X_features[col].std() for col in clustering_features],
    'Scaler_Mean': scaler.mean_,
    'Scaler_Scale': scaler.scale_
})

```

Gambar 4.7 Code Hasil PreProcessing Meta Data

Pada tahap ini sistem menghitung statistik dasar dari setiap fitur yang digunakan dalam proses clustering, yaitu *TOTAL VOLUME (L/KG)*, *NUM\_PESTICIDE\_TYPES*, *NUM\_PACKAGING\_TYPES*, dan *AVG\_VOLUME*. Statistik yang dihitung meliputi nilai minimum (Min), maksimum (Max), rata-rata (Mean), dan standar deviasi (Std Dev) sebagai berikut :

Sebagai contoh pada fitur *TOTAL VOLUME (L/KG)*, sistem menghitung nilai minimum sebesar 0.50 dan maksimum sebesar 2972.00 dari seluruh data retailer.

1. Menghitung Nilai Rata-rata (Mean)

Rata-rata diperoleh dari jumlah seluruh data dibagi jumlah retailer

$$Mean_{(x)} = \frac{\sum x_i}{n}$$

Misalnya total seluruh volume retailer adalah :  $Total = 29458.56$

Jumlah Retailer :  $n = 134$

Maka :

$$Mean = \frac{29458.56}{134} = 219.84$$

Nilai ini sesuai dengan Mean pada tabel metadata.

## 2. Menghitung Parameter StandardScaler

StandardScaler menggunakan rumus normalisasi:  $Z = \frac{x - \mu}{\sigma}$

Keterangan:

1.  $x$  = nilai data
2.  $\mu$  = mean
3.  $\sigma$  = standar deviasi

Pada Fitur Total Volume Scaler Mean 219.84, dan Scaler Scale 539.39

Artinya: Scaler Mean = Mean fitur, Scaler Scale = standar deviasi yang digunakan untuk normalisasi.

Contoh Normalisasi Data :

Misalnya retailer Ade memiliki TOTAL VOULME = 51

Maka nilai setelah normalisasi:  $Z = \frac{51 - 219.84}{539.39} = -0.31$

Nilai -0.31 merupakan nilai baru dari fitur TOTAL VOLUME setelah proses normalisasi.

#### d. Tampilan Hasil *Clustering Data Scaled*



NO	RETAILER	TOTAL VOLUME	TOTAL JENIS PESTISIDA	TOTAL JENIS KEMASAN	AVG VOLUME
1	ADANG	-0.380	-0.816	-0.614	-0.390
2	ADE	-0.213	0.456	-0.233	-0.351
3	ADUL	-0.352	-0.816	-0.995	0.485
4	AEP TANI (AEF)	-0.134	-0.816	0.148	0.755
5	AHLI TANI 13 (BPK MAHMUDI)	3.651	1.728	2.052	2.471
6	AILA TANI (BPK MAKMUR)	0.562	1.728	1.671	0.165
7	ALAM MAKMUR (H. IBRAHIM)	-0.210	1.728	1.290	-0.485
8	AM-POWER (BPK ROBI)	-0.398	-0.816	-0.995	-0.487
9	ANGGITA JAYA (BPK SURYONO)	4.114	1.728	2.052	2.831
10	ANGRI TANI (ADANG)	-0.222	-0.816	-0.614	1.262

Gambar 4.8 Tampilan Hasil *Clustering Data Scaled* dari Web

Pada tahap ini dilakukan proses normalisasi data menggunakan metode *StandardScaler* agar seluruh fitur memiliki skala yang sebanding sebelum digunakan dalam proses clustering. Normalisasi diperlukan karena setiap fitur memiliki rentang nilai yang berbeda, sehingga tanpa normalisasi fitur dengan nilai besar dapat mendominasi perhitungan jarak pada algoritma clustering.

*StandardScaler* menggunakan rumus normalisasi:  $Z = \frac{x - \mu}{\sigma}$

Keterangan:

1.  $x$  = nilai data
2.  $\mu$  = mean
3.  $\sigma$  = standar deviasi

a. Retailer Ade Normalisasi Total Volume

Dik : TOTAL VOULME = 51

Maka nilai Total Volume:  $Z = \frac{51 - 219.84}{539.39} = -0.31$

b. Retailer Ade Normalisasi Total Jenis Pestisida

Perhitungan

Dik : Jumlah Jenis Pestisida = 2

Maka nilai Num\_Pesticide\_Types:  $Z = \frac{2-1.64}{0.79} = 0,46$

c. Retailer Ade Normalisai Total Jenis Kemasan

Dik : Jumlah kemasan =3

Maka nilai Num\_Pesticide\_Types:  $Z = \frac{3-3.61}{2.63} = -0.23$

d. Retailer Ade Normalisasi AVG\_VOLUME

Dari perhitungan sebelumnya:

Total Volume = 51

Jumlah Pestisida = 2

Jumlah Kemasan = 3

$$Avg\_Volume = \frac{51}{2 \times 3} = 8.5$$

Maka nilai Avg\_Volume :  $Z = \frac{8.5-17.53}{25.72} = -0.35$

Hasil Clustering Data Scaled

Tabel 4.8 Clustering Data Scaled Ade

Retailer	Total Volume	Nums Pesticide Types	Num Packaging Types	AvgVolume
ADE	-0.31	0.46	-0.23	-0.35

### 4.2.3. Tampilan Halaman Clustering

**K-Pesticide Sistem Clustering**

**Hasil Analisis Clustering**  
Perbandingan performa algoritma HAC (Ward) dan K-Medoids (PAM)

Detail Hasil Clustering (Sampel) Menampilkan 20 retailer pertama

No	Nama Retailer	HAC Cluster	K-Medoids Cluster	Total Volume	Pesticide Types	Match
1	ADANG	2	0	15.00	1	⊖
2	ADE	0	2	51.00	2	⊖
3	ADUL	2	2	30.00	1	⊕
4	AEP TANI (AEF)	2	2	147.80	1	⊕
5	AHLI TANI 13 (BPK MAHMUD)	1	1	2189.05	3	⊕
6	AILA TANI (BPK MAKMUR)	0	1	522.75	3	⊖
7	ALAM MAKMUR (H. IBRAHIM)	0	1	106.30	3	⊖
8	AM-POWER (BPK ROBI)	2	0	5.00	1	⊖
9	ANGGITA JAYA (BPK SURYONO)	1	1	2438.90	3	⊕
10	ANGRI TANI (ADANG)	2	2	100.00	1	⊕

Gambar 4.9 Tampilan Halaman Hasil Analisis Clustering

Pada tahap ini sistem menampilkan hasil pengelompokan retailer yang diperoleh dari proses clustering menggunakan algoritma Hierarchical Agglomerative Clustering (HAC) dan K-Medoids. Hasil clustering ditampilkan dalam bentuk tabel yang memuat informasi nama retailer, cluster hasil HAC, cluster hasil K-Medoids, total volume pembelian, serta jumlah jenis pestisida yang dimiliki oleh masing-masing retailer.

Proses penentuan cluster dilakukan dengan menghitung jarak antar data menggunakan Euclidean Distance pada fitur yang telah dinormalisasi sebelumnya. Secara matematis perhitungan jarak antar dua data dapat dituliskan sebagai berikut:

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

Keterangan:

1.  $x$  = data retailer pertama
2.  $y$  = data retailer kedua

3.  $n$  = jumlah fitur yang digunakan dalam clustering

Sebagai contoh pada data retailer ADE yang memiliki:

Total Volume = 51.00

Jumlah Jenis Pestisida = 2

Nilai tersebut digunakan bersama fitur lainnya untuk menghitung jarak dengan retailer lain. Berdasarkan hasil perhitungan tersebut, algoritma HAC menempatkan retailer ADE pada Cluster 0, sedangkan algoritma K-Medoids menemukannya pada Cluster 2.

Contoh lainnya pada retailer ADUL dengan Total Volume 30.00 dan 1 jenis pestisida, kedua algoritma yaitu HAC dan K-Medoids menempatkan retailer tersebut pada Cluster 2, sehingga hasil pengelompokannya sama.

Detail Hasil Clustering (Sampel) menampilkan sebagian data retailer beserta cluster yang diperoleh dari masing-masing algoritma, sehingga memudahkan dalam melihat bagaimana setiap retailer dikelompokkan berdasarkan karakteristik pembelian pestisida.

#### a. Tampilan Hasil Cluster Distribusi HAC dan K-Medoids

The screenshot shows a dashboard for 'K-Pesticide Sistem Clustering'. On the left is a navigation menu with options: Dashboard, Unggah Data, Pembersihan Data, Pra-pemrosesan, Clustering (highlighted), Dokumentasi, and Reset Data. The main content area displays a list of retailers and two summary tables for HAC and K-Medoids clustering results.

ID	Retailer	HAC	K-Medoids	Total Volume	Total Pesticide
17	BANYU TANI (BPK SARWAN)	2	0	5.00	1
18	BAROKAH TANI (PIAN)	2	0	0.50	1
19	BERKAH JAYA TANI (BPK ENDANG)	2	0	5.00	1
20	BERKAH TANI (BPK DEDEN)	2	2	32.00	1

Distribusi Kluster HAC (Ward)					Distribusi Kluster K-Medoids (PAM)				
Kluster	Jumlah Retailer	Persentase	Avg Volume	Avg Pesticide	Kluster	Jumlah Retailer	Persentase	Avg Volume	Avg Pesticide
Kluster 0	51	38.1%	147.71	2.35	Kluster 0	52	38.8%	7.17	1.12
Kluster 1	10	7.5%	1966.46	2.70	Kluster 1	40	29.8%	589.23	2.62
Kluster 2	73	54.5%	30.97	1.00	Kluster 2	42	31.3%	131.35	1.36
<b>Total</b>	<b>134</b>	<b>100%</b>	-	-	<b>Total</b>	<b>134</b>	<b>100%</b>	-	-

Gambar 4.10 Hasil Clustering Distribusi HAC dan K-Medoids

Tampilan Distribusi Kluster HAC (Ward) dan Distribusi Kluster K-Medoids (PAM) menunjukkan ringkasan hasil pengelompokan retailer berdasarkan jumlah anggota pada setiap cluster, persentase distribusi cluster,

serta rata-rata nilai fitur seperti Total Volume dan Jumlah Jenis Pestisida. Nilai-nilai tersebut diperoleh melalui proses perhitungan yang ditampilkan pada langkah berikut :

1. Perhitungan Jumlah Retailer dalam Cluster

Jumlah retailer dalam cluster dihitung dengan menghitung banyaknya data yang memiliki label cluster yang sama.

$$Jumlah\ Retailer\ cluster = \sum_{i=1}^n I (label_i = cluster)$$

Keterangan :

1.  $label_i$  = label cluster data ke-i
2.  $I$  = fungsi indikator

Contoh pada HAC Cluster 0 : *Jumlah Retailer cluster* == 51

2. Perhitungan Persentase Cluster

Persentase cluster dihitung dengan membagi jumlah retailer dalam cluster dengan total retailer.

$$Jumlah\ Retailer\ cluster = \frac{Jumlah\ Retailer\ Cluster}{Total\ Retailer} \times 100\ %$$

Contoh pada HAC Cluster 0:

Total retailer = 134

$$Persentase = \frac{51}{134} \times 100 = 38.1\%$$

3. Perhitungan Rata-rata Total Volume

Rata-rata total volume dihitung dengan menjumlahkan seluruh total volume dalam cluster kemudian dibagi jumlah retailer dalam cluster.

$$Avg Volume = \frac{\sum_{i=1}^n Volume_i}{Jumlah Retailer cluster}$$

Diketahui:

Jumlah retailer dalam Cluster 0 = 51

Misalnya sebagian nilai volume dalam cluster tersebut:

$$15 + 51 + 30 + 147.80 + 522.75 + \dots + 2438.90$$

Total volume seluruh retailer dalam Cluster 0 dari sistem adalah:

$$\sum Volume = 7533.21$$

Maka rata-rata volume dihitung sebagai berikut:

$$Avg Volume = \frac{7533.21}{51} = 147.71$$

Hasil tersebut sesuai dengan nilai Avg Volume = 147.71 yang ditampilkan pada tabel Distribusi Kluster HAC.

#### 4. Perhitungan Rata-rata Jenis Pestisida

Rata-rata jumlah jenis pestisida dihitung dengan menjumlahkan seluruh nilai NUM\_PESTICIDE\_TYPES dalam cluster kemudian dibagi jumlah retailer.

$$Avg Volume = \frac{\sum_{i=1}^n PesticideTypes_i}{Jumlah Retailer}$$

Contoh (HAC Cluster 0)

Total jumlah jenis pestisida dalam cluster:

$$\sum PesticideTypes = 119.85$$

Jumlah retailer: 51

Maka:

$$Avg Volume = \frac{119.85}{51} = 2.35$$

### b. Tampilan Hasil Matrik Evaluasi

ALGORITMA	JUMLAH KLASTER	SILHOUETTE SCORE	DAVIES-BOULDIN INDEX	METODE
HAC (Hierarchical Agglomerative)	3	0.5195	0.7207	ward (euclidean)
K-Medoids (PAM)	3	0.2914	1.2987	PAM (euclidean)

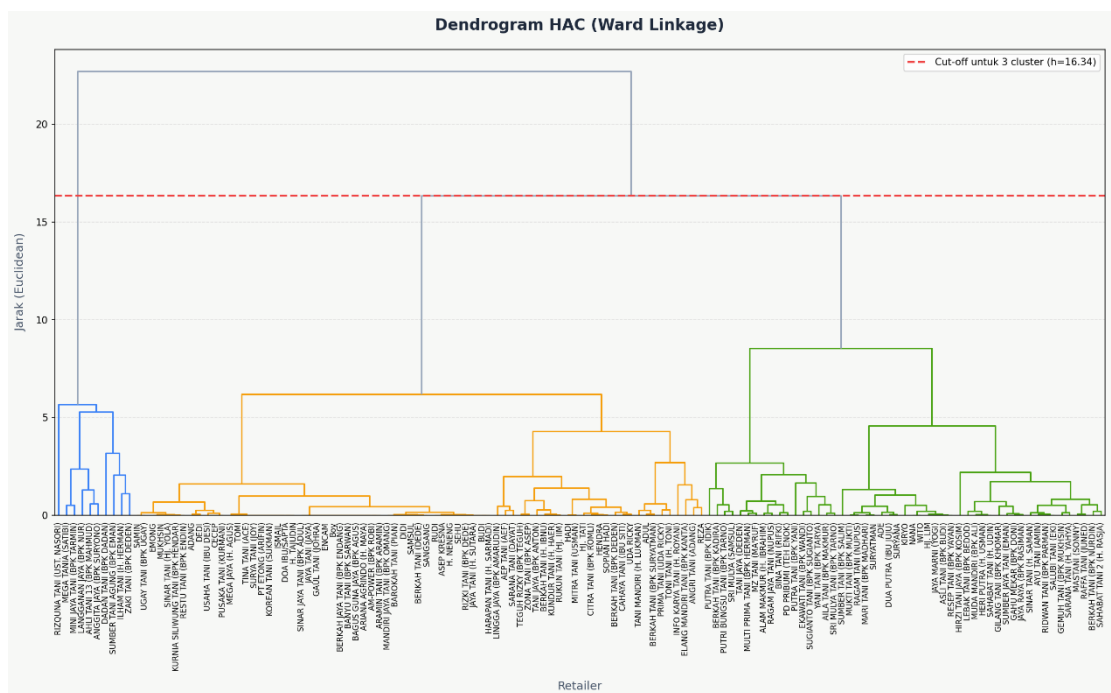
Gambar 4.11 Hasil Matrik Evaluasi Silhouette Score dan Davies-Bouldin Index

Bagian Perbandingan Metrik Evaluasi menampilkan hasil pengukuran kualitas clustering yang dihasilkan oleh dua algoritma yang digunakan, yaitu Hierarchical Agglomerative Clustering (HAC) dengan metode Ward dan K-Medoids (PAM). Evaluasi dilakukan menggunakan dua metrik yaitu Silhouette Score dan Davies-Bouldin Index untuk melihat seberapa baik struktur cluster yang terbentuk dari data retailer. Berdasarkan hasil pengujian, algoritma HAC menghasilkan nilai Silhouette Score sebesar 0.5195, sedangkan K-Medoids menghasilkan nilai sebesar 0.2914. Nilai Silhouette Score yang lebih tinggi menunjukkan bahwa data dalam cluster memiliki tingkat kemiripan yang lebih baik dengan anggota cluster yang sama dan memiliki perbedaan yang cukup jelas dengan cluster lainnya.

Selain itu, evaluasi juga dilakukan menggunakan Davies-Bouldin Index (DBI) untuk mengukur tingkat pemisahan antar cluster. Hasil pengujian

menunjukkan bahwa algoritma HAC memperoleh nilai DBI sebesar 0.7207, sedangkan K-Medoids memperoleh nilai sebesar 1.2987. Pada metrik ini, nilai yang lebih kecil menunjukkan kualitas cluster yang lebih baik karena jarak antar cluster lebih jelas dan tidak saling tumpang tindih. Berdasarkan kedua metrik evaluasi tersebut, dapat disimpulkan bahwa algoritma Hierarchical Agglomerative Clustering (HAC) memberikan hasil clustering yang lebih baik dibandingkan dengan algoritma K-Medoids pada data penelitian ini

#### 4.2.4. Tampilan Halaman Dokumentasi *Clustering* (Visualisasi Dendrogram HAC)



Gambar 4.12 Visualisasi dendrogram Hac

Diagram Dendrogram HAC (Ward Linkage) menampilkan proses penggabungan data retailer secara bertahap berdasarkan tingkat kemiripan menggunakan jarak Euclidean. Pada dendrogram tersebut, setiap cabang menunjukkan proses penggabungan cluster hingga seluruh data tergabung menjadi satu struktur hierarki. Garis putus-putus merah menunjukkan batas

pemotongan (cut-off) pada ketinggian sekitar 16.34 yang digunakan untuk membentuk 3 cluster utama. Berdasarkan pemotongan tersebut, data retailer terbagi menjadi tiga kelompok yang memiliki karakteristik kemiripan tertentu berdasarkan fitur yang digunakan dalam proses clustering.

a. Visualisasai Sebaran Data per Cluster



Gambar 4.13 Sebaran Data per Cluster

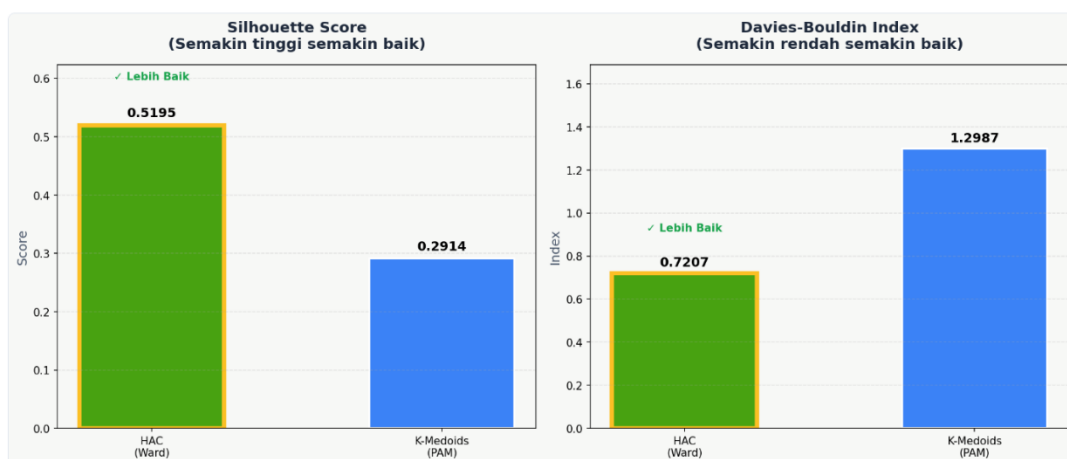
Visualisasi pada gambar menunjukkan hasil pengelompokan data retailer menggunakan algoritma Hierarchical Agglomerative Clustering (HAC) dengan metode Ward dan K-Medoids (PAM). Data yang semula memiliki empat dimensi direduksi menjadi dua dimensi menggunakan Principal Component Analysis (PCA) agar dapat divisualisasikan dalam bentuk scatter plot.

Pada scatter plot HAC, setiap titik merepresentasikan satu retailer yang dikelompokkan ke dalam tiga cluster yang ditandai dengan perbedaan warna. Metode HAC mengelompokkan data berdasarkan kedekatan jarak antar objek secara hierarki, sehingga tidak memiliki titik pusat cluster. Hasil visualisasi menunjukkan bahwa cluster yang terbentuk memiliki pemisahan yang cukup baik, meskipun terdapat beberapa titik yang berada berdekatan antar cluster.

Sementara itu, pada scatter plot K-Medoids, selain menampilkan sebaran data berdasarkan cluster, juga ditampilkan titik medoid yang ditandai dengan simbol khusus (X). Medoid merupakan titik data aktual yang paling representatif dalam suatu cluster, yang diperoleh dengan meminimalkan total jarak terhadap anggota cluster lainnya. Setiap cluster memiliki satu medoid, sehingga pada hasil clustering dengan tiga cluster terdapat tiga titik medoid.

Secara visual, medoid berada di pusat distribusi masing-masing cluster dan dikelilingi oleh anggota cluster yang memiliki karakteristik serupa. Hal ini menunjukkan bahwa K-Medoids tidak hanya mampu mengelompokkan data, tetapi juga memberikan representasi pusat cluster yang lebih interpretable dibandingkan HAC.

#### b. Visualisasi Metrik Evaluasi

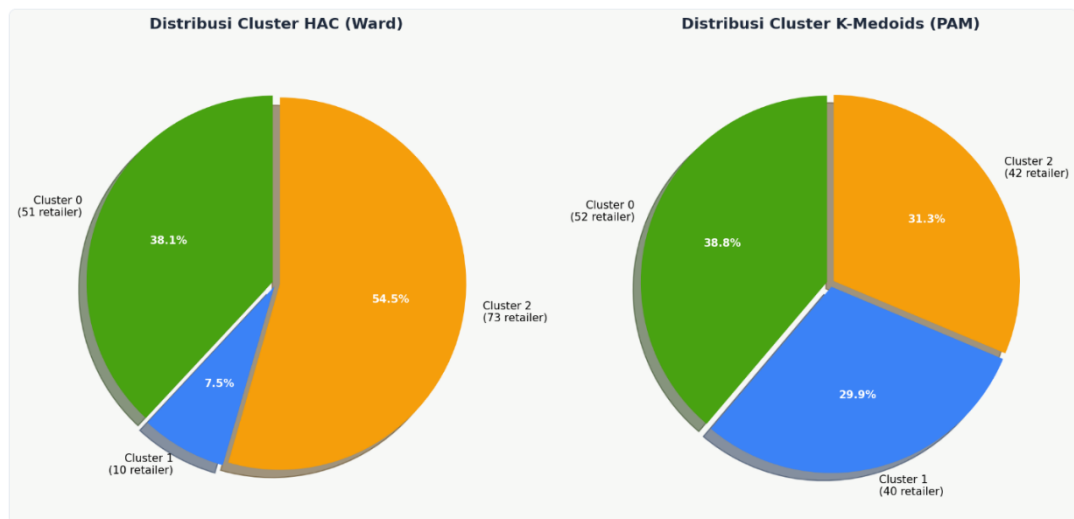


Gambar 4.14 Visualisasi Metrik Evaluasi

Grafik Silhouette Score dan Davies-Bouldin Index menunjukkan perbandingan kualitas hasil clustering antara algoritma Hierarchical Agglomerative Clustering (HAC) dan K-Medoids. Pada grafik Silhouette Score terlihat bahwa algoritma HAC memperoleh nilai sebesar 0.5195, lebih tinggi

dibandingkan K-Medoids sebesar 0.2914, yang menunjukkan bahwa struktur cluster yang dihasilkan HAC lebih baik. Sementara itu pada grafik Davies-Bouldin Index, algoritma HAC memperoleh nilai 0.7207 yang lebih rendah dibandingkan K-Medoids sebesar 1.2987, sehingga menunjukkan bahwa pemisahan antar cluster pada HAC lebih optimal dibandingkan K-Medoids.

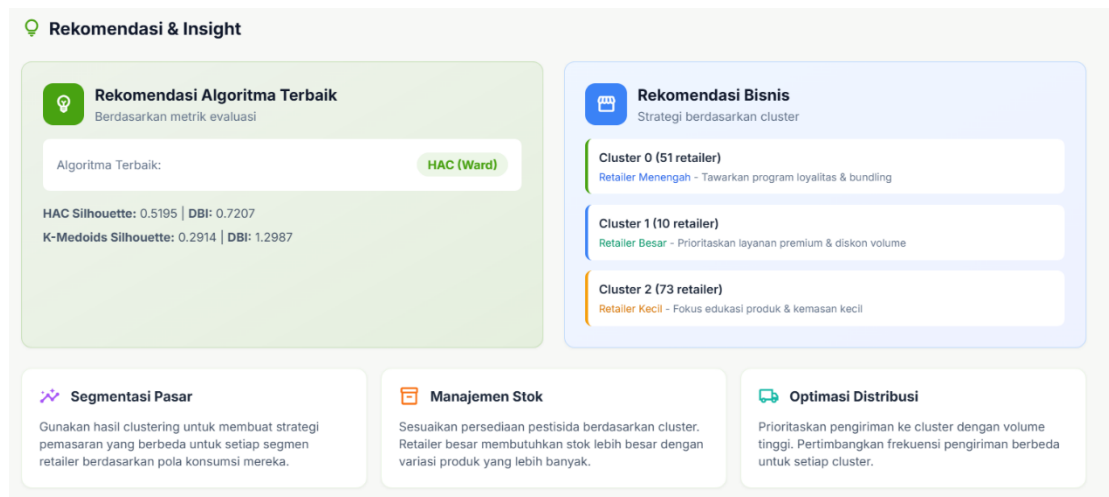
### c. Visualisasi Distribusi Cluster



Gambar 4 14 Distribusi Cluster

Diagram Distribusi Cluster HAC (Ward) dan Distribusi Cluster K-Medoids (PAM) menampilkan perbandingan jumlah retailer pada setiap cluster yang terbentuk dari kedua algoritma clustering. Pada metode HAC, sebagian besar retailer berada pada Cluster 2 sebanyak 73 retailer (54.5%), diikuti Cluster 0 sebanyak 51 retailer (38.1%), dan Cluster 1 sebanyak 10 retailer (7.5%). Sementara pada metode K-Medoids, distribusi data lebih merata yaitu Cluster 0 sebanyak 52 retailer (38.8%), Cluster 1 sebanyak 40 retailer (29.9%), dan Cluster 2 sebanyak 42 retailer (31.3%). Perbedaan distribusi ini menunjukkan bahwa masing-masing algoritma menghasilkan pola pengelompokan retailer yang berbeda berdasarkan karakteristik data yang dianalisis.

#### d. Tampilan Rekomendasi dan Insight Hasil Clustering



Gambar 4 15 Tampilan Rekomendasi dan Insight Hasil Clustering

Gambar ini menampilkan halaman Rekomendasi & Insight yang berisi ringkasan hasil analisis clustering. Sistem merekomendasikan algoritma HAC (Ward) sebagai metode terbaik berdasarkan nilai Silhouette Score 0.5195 dan Davies-Bouldin Index 0.7207, serta menampilkan rekomendasi strategi bisnis berdasarkan hasil pengelompokan retailer ke dalam tiga cluster.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai pengelompokan produk pestisida berdasarkan pola konsumsi retailer menggunakan algoritma Hierarchical Agglomerative Clustering (HAC) dan K-Medoids, dapat ditarik kesimpulan sebagai berikut:

1. Pra-Pemrosesan Data: Proses pembersihan berhasil mentransformasi dataset mentah sejumlah 1.594 transaksi menjadi 1.488 data valid. Setelah agregasi, diperoleh 134 profil retailer yang merepresentasikan karakteristik pembelian mereka.
2. Perbandingan Algoritma: Setelah membandingkan *Hierarchical Agglomerative Clustering* (HAC) dengan *K-Medoids* (PAM), algoritma HAC menggunakan metode *Ward* terbukti memberikan performa yang jauh lebih unggul. Hal ini terlihat dari nilai *Silhouette Score* HAC yang mencapai 0,5195 (lebih tinggi dari *K-Medoids* yang hanya 0,2914) serta nilai *Davies-Bouldin Index* (DBI) HAC yang lebih rendah yaitu 0,7207 (dibandingkan *K-Medoids* sebesar 1,2987). Secara teknis, HAC mampu menghasilkan klaster yang lebih padat dan terpisah dengan lebih optimal untuk data transaksi ini.
3. Segmentasi Retailer: Hasil pengelompokan membagi retailer ke dalam tiga klaster utama yang memberikan insight strategis bagi distributor:

- Cluster 0 (51 retailer): Dikategorikan sebagai Retailer Menengah, disarankan untuk mendapatkan penawaran program loyalitas dan bundling.
- Cluster 1 (10 retailer): Dikategorikan sebagai Retailer Besar, disarankan untuk mendapatkan prioritas layanan premium dan diskon volume.
- Cluster 2 (73 retailer): Dikategorikan sebagai Retailer Kecil, disarankan untuk difokuskan pada edukasi produk dan kemasan kecil.

## 5.2. Saran

Demi pengembangan riset ini ke depannya, penulis menyampaikan beberapa saran sebagai berikut:

1. **Penambahan Variabel Eksternal:** Peneliti selanjutnya disarankan untuk memasukkan faktor eksternal lain, seperti tren musiman, harga kompetitor, atau detail lokasi geografis, agar pemetaan karakteristik *retailer* menjadi lebih presisi.
2. **Optimalisasi Parameter:** Penting untuk mengeksplorasi teknik *clustering* lainnya atau menggunakan metode tambahan seperti *Elbow Method* untuk menentukan jumlah klaster secara lebih dinamis dan objektif.
3. **Sistem *Real-time*:** Diharapkan adanya pengembangan sistem agar mampu memproses data secara *real-time* atau berkala, sehingga distributor dapat lebih responsif dalam menyesuaikan stok sesuai perubahan pola konsumsi *retailer*.
4. **Implementasi Strategi:** Bagi CV. Aneka Jaya Tunas Agro, rekomendasi strategi bisnis yang dihasilkan dalam penelitian ini sangat disarankan untuk segera diimplementasikan secara bertahap guna meningkatkan efisiensi operasional dan optimalisasi manajemen persediaan produk.

## DAFRAT PUSTAKA

Agustin, E.W. *et al.* (2025) ‘Optimization of Customer Segmentation in the Retail Industry Using the K-Medoid Algorithm’, *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), pp. 766–775. Available at: <https://doi.org/10.57152/malcom.v5i3.1977>.

Ahmad, S.F. *et al.* (2023) ‘Data Mining Methods and Obstacles : A Comprehensive Analysis’, 06(01).

Chilipweli, P.M., Nyanza, E.C. and Ngowi, A.V. (2025) ‘Patterns , Practices , and Socio-Environmental Dynamics of Pesticide Use in the Horticultural Value Chain : Insights from Smallholder Farmers and Agro-Input Sellers in Iringa and’, pp. 1–24.

Dhulipala, L. *et al.* (2022) ‘Hierarchical Agglomerative Graph Clustering in Poly-Logarithmic Depth’, *Advances in Neural Information Processing Systems*, 35(NeurIPS).

Farahnakian, Farshad *et al.* (2023) ‘A Comprehensive Study of Clustering-Based Techniques for Detecting Abnormal Vessel Behavior’, *Remote Sensing*, 15(6), pp. 1–34. Available at: <https://doi.org/10.3390/rs15061477>.

Garcia-Morato, E. *et al.* (2024) ‘A general framework for distributed approximate similarity search with arbitrary distances’. Available at: <http://arxiv.org/abs/2405.13795>.

John, J.M., Shobayo, O. and Ogunleye, B. (2023) ‘An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market’, *Analytics*, 2(4), pp. 809–823. Available at: <https://doi.org/10.3390/analytics2040042>.

Kajornkasirat, S. *et al.* (2025) ‘Customer segmentation using association rule mining on retail transaction data’, *IAES International Journal of Artificial Intelligence*, 14(3), pp. 1919–1929. Available at: <https://doi.org/10.11591/ijai.v14.i3.pp1919-1929>.

Luh, N. *et al.* (2022) ‘Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means ( Studi Kasus : STMIK Primakara )’, 16(2), pp. 105–112.

Mathelin, A. de *et al.* (2025) ‘OneBatchPAM: A Fast and Frugal K-Medoids Algorithm’, *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(15), pp. 16172–16180. Available at: <https://doi.org/10.1609/aaai.v39i15.33776>.

Oti, E.U. and Olusola, M.O. (2024) ‘Overview of Agglomerative Hierarchical Clustering Methods’, *British Journal of Computer, Networking and Information Technology*, 7(2), pp. 14–23. Available at: <https://doi.org/10.52589/bjcnit-cv9poogw>.

Pireddu, A. *et al.* (2024) ‘A Review of Data Mining Strategies by Data Type, with a Focus on Construction Processes and Health and Safety Management’, *International Journal of Environmental Research and Public Health*, 21(7). Available at: <https://doi.org/10.3390/ijerph21070831>.

Pitafi, S., Anwar, T. and Sharif, Z. (2023) 'A Taxonomy of Machine Learning Clustering Algorithms, Challenges, and Future Realms', *Applied Sciences (Switzerland)*, 13(6). Available at: <https://doi.org/10.3390/app13063529>.

Rizky, F. *et al.* (2025) 'IMPLEMENTATION OF DATA MINING TO CATEGORIZE DAILY SALES TRANSACTION DATA USING K-MEANS', 6(1), pp. 2160–2169.

Syakra, Y. *et al.* (2025) 'Customer Segmentation Using RFM and K-Means Clustering to Support CRM in Retail Industry', *PeerJ Computer Science*, 9(3), pp. 1120–1131. Available at: <https://doi.org/10.33395/sinkron.v9i3.14907>.

Teixeira, L.R. *et al.* (2024) 'KluSIM: Speeding up K-Medoids Clustering over Dimensional Data with Metric Access Method', *International Conference on Enterprise Information Systems, ICEIS - Proceedings*, 1(Iceis), pp. 73–84. Available at: <https://doi.org/10.5220/0012599900003690>.

Tiwari, M. *et al.* (2023) 'BanditPAM++: Faster k-medoids Clustering', *Advances in Neural Information Processing Systems*, 36(NeurIPS).

Triantara, I.K.Y., Raharja, M.A. and Sarasvananda, I.B.G. (2025) 'Segmentasi Pelanggan Berbasis RFMT Menggunakan K-Means dan Hierarchical Clustering I Komang Yosua Triantara a1 , Made Agung Raharja a2, Ida Bagus Gede Sarasvananda a3', *JNATIA*, 3(4).

Tsiu, S. V. *et al.* (2025) 'Applications and Competitive Advantages of Data Mining and Business Intelligence in SMEs Performance: A Systematic Review', *Businesses*, 5(2), p. 22. Available at: <https://doi.org/10.3390/businesses5020022>.

Wilbert, H.J. *et al.* (2023) 'Using Clustering for Customer Segmentation from Retail Data'. Available at: <https://doi.org/10.20944/preprints202308.0580.v1>.

Wu, Z. *et al.* (2022) 'Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm', *Computational Intelligence and Neuroscience*, 2022. Available at: <https://doi.org/10.1155/2022/9930613>.