



BERITA ACARA SIDANG PENDADARAN TUGAS AKHIR

S/UBL/FTI/0081/VII/26

Pada hari ini, Senin 13 Juli 2026 telah dilaksanakan Ujian Sidang Pendadaran Tugas Akhir sebagai berikut:

Judul: PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS POLA PENJUALAN BERDASARKAN WAKTU, PRODUK, DAN WILAYAH PADA GWMM SHOP

Nama : Febri Ardana
NIM : 2212500199
Dosen Pembimbing : Muhammad Ainur Rony, S.Kom., M.T.I.

Berdasarkan penilaian pada Presentasi, Penulisan, Penguasaan Materi, Penguasaan Program / Prototipe, Kompleksitas maka Mahasiswa tersebut di atas dinyatakan:

LULUS

dengan nilai angka : **86** huruf : **A**

Mahasiswa tersebut di atas wajib menyerahkan hasil perbaikan tulisan Tugas Akhir dalam bentuk terjilid sesuai dengan Panduan Perbaikan Tugas Akhir, selambat-lambatnya Senin 27 Juli 2026.

Panitia Penguji:

- | | |
|-------------|-------------------------------------|
| 1 Ketua | Lauw Li Hin, S.Kom., M.Kom. |
| 2 Anggota | Agnes Aryasanti, S.Kom., M.Kom. |
| 3 Moderator | Muhammad Ainur Rony, S.Kom., M.T.I. |

Keterangan:

Nilai Huruf: A:85-100 A-:80-84,99 B+:75-79,99 B:70-74,99 B-:65-69,99 C:60-64,99 D:40-59,99 E-:0-39,99



UNIVERSITAS BUDI LUHUR
FAKULTAS TEKNOLOGI INFORMASI

Kartu Bimbingan Tugas Akhir

NIM: 2212500199

Nama: Febri Ardana

Pembimbing: Muhammad Ainur Rony, S.Kom., M.T.I.

No.	Tanggal	Materi
1	23-03-2026	Pendahuluan latar belakang
2	31-03-2026	Bab 1 finalisasi
3	06-04-2026	Bab 2 finalisasi
4	14-04-2026	Bab 3
5	21-04-2026	Bab 3 progress
6	12-05-2026	Bab 3 finalisasi
7	25-05-2026	Bab 4
8	23-06-2026	bab 5



LEMBAR PENGESAHAN

Nama : Febri Ardana
Nomor Induk Mahasiswa : 2212500199
Program Studi : Sistem Informasi
Bidang Peminatan : Data Science
Jenjang Studi : Strata 1
Judul : PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK
ANALISIS POLA PENJUALAN BERDASARKAN WAKTU, PRODUK,
DAN WILAYAH PADA GWMM SHOP



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 13 Juli 2026

Tim Penguji:

Ketua : Lauw Li Hin, S.Kom., M.Kom.
Anggota : Agnes Aryasanti, S.Kom., M.Kom.
Pembimbing : Muhammad Ainur Rony, S.Kom., M.T.I.
Ketua Program Studi : Bima Cahya Putra, S.Kom., M.Kom.

**Penerapan Algoritma K-Means Clustering untuk Analisis Pola
Penjualan Berdasarkan Waktu, Produk, dan Wilayah pada
GWMM SHOP**

**Diajukan untuk memenuhi salah satu persyaratan memperoleh
gelar Sarjana Komputer (S. Kom)**

TUGAS AKHIR



**Oleh:
Febri Ardana
2212500199**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2026**



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

LEMBAR PENGESAHAN

Nama	: Febri Ardana
Nomor Induk Mahasiswa	: 2212500199
Program Studi	: Sistem Informasi
Peminatan	: Data Science
Jenjang Studi	: Strata 1
Judul	: PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS POLA PENJUALAN BERDASARKAN WAKTU, PRODUK, DAN WILAYAH PADA GWMM SHOP



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 13 Juli 2026

Tim Penguji:

Ketua	: Lauw Li Hin, S.Kom., M.Kom.
Anggota	: Agnes Aryasanti, S.Kom., M.Kom.
Pembimbing	: Muhammad Ainur Rony, S.Kom., M.T.I.
Ketua Program Studi	: Bima Cahya Putra, S.Kom., M.Kom.

ABSTRAK
PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK ANALISIS
POLA PENJUALAN BERDASARKAN WAKTU, PRODUK, DAN
WILAYAH PADA GWMM SHOP

Oleh : Febri Ardana (NIM : 2212500199)

Pesatnya pertumbuhan *e-commerce* di Indonesia mendorong semakin banyak pelaku UMKM memanfaatkan platform *marketplace* seperti Shopee sebagai saluran penjualan utama. Namun sebagian besar UMKM masih mengelola bisnisnya secara konvensional tanpa memanfaatkan data transaksi yang sebenarnya sudah mereka miliki. GWMM SHOP, sebuah toko parfum yang berjualan di Shopee, mengalami permasalahan serupa karena belum memiliki sistem analisis yang mampu mengidentifikasi pola penjualan secara spesifik berdasarkan waktu, jenis produk, dan wilayah pembeli. Akibatnya, keputusan terkait manajemen stok, waktu promosi, dan target pemasaran masih dilakukan secara spekulatif dan belum berbasis data. Penelitian ini bertujuan menerapkan metode *data mining* dengan algoritma K-Means *Clustering* untuk mengelompokkan data transaksi penjualan GWMM SHOP berdasarkan tiga dimensi utama, yaitu waktu (bulan), produk, dan wilayah (provinsi). Data yang digunakan merupakan data *primer* berupa riwayat transaksi penjualan yang diekspor langsung dari *Seller Center Shopee* milik GWMM SHOP pada periode Januari 2024 hingga Desember 2025, dengan jumlah awal 4.163 baris data yang menyusut menjadi 3.854 baris setelah melalui tahap data cleaning. Proses analisis mengikuti metodologi CRISP-DM, mulai dari pemahaman data, persiapan data seperti label *encoding* dan normalisasi *Min-Max*, pemodelan, hingga evaluasi menggunakan *Elbow Method* dan *Davies-Bouldin Index* (DBI). Hasil komparasi antara algoritma K-Means dan K-Medoids menunjukkan bahwa K-Means dengan 4 klaster memberikan nilai DBI paling rendah, yaitu sebesar 1,002, sehingga dipilih sebagai model terbaik dalam penelitian ini. Keempat klaster yang terbentuk memiliki karakteristik yang berbeda. Klaster 0 berisi 1.203 transaksi yang terkonsentrasi pada bulan Juni hingga September dengan rata-rata total harga Rp162.596 dan wilayah dominan Sumatera Utara. Klaster 1 merupakan klaster dengan jumlah transaksi terbanyak, yaitu 1.449 transaksi, dengan rata-rata total harga tertinggi sebesar Rp725.445, terjadi pada bulan Oktober hingga Desember bertepatan dengan momen Harbolnas, Natal, dan Tahun Baru, dengan wilayah dominan DKI Jakarta, Jawa Barat, dan Jawa Timur. Klaster 2 berisi 652 transaksi pada bulan Mei hingga Juli dengan rata-rata total harga Rp234.784 dan wilayah dominan Papua, Nusa Tenggara Timur, serta Kalimantan Timur. Klaster 3 berisi 550 transaksi dengan nilai transaksi paling rendah, yaitu rata-rata Rp93.116, terkonsentrasi pada bulan Januari hingga Mei dengan wilayah dominan Bali, Gorontalo, dan Bengkulu. Hasil penelitian ini menghasilkan rekomendasi bagi GWMM SHOP dalam pengambilan keputusan operasional berbasis pola penjualan, mencakup strategi manajemen stok menjelang akhir tahun, penentuan waktu promosi, dan perluasan target pemasaran ke wilayah potensial di luar Pulau Jawa.

Kata Kunci: Data Mining, K-Means Clustering, Pola Penjualan, UMKM, E-Commerce.

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK.....	ii
PERNYATAAN TIDAK PLAGIAT	iii
PERNYATAAN PENGGUNAAN AI GENERATIF	iv
KATA PENGANTAR	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
DAFTAR ISI.....	ix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Masalah.....	2
1.3 Tujuan dan Manfaat Penelitian	3
1.4 Sistematika Penulisan	3
BAB II STUDI PUSTAKA.....	5
2.1 Teori – Teori Umum.....	5
2.1.1 Pengertian <i>Data Mining</i>	5
2.1.2 CRISP – DM	5
2.1.3 <i>K-Means</i>	6
2.1.4 Metode Elbow	7
2.1.5 <i>Davies Bouldin Index</i> (DBI)	7
2.2 Teori Khusus	8
2.2.1 <i>E-Commerce</i>	8
2.2.2 Klusterisasi.....	8
2.2.3 Transformasi Data.....	9
2.2.4 Normalisasi Data.....	9
2.3 <i>Tools</i> Yang Digunakan	9
2.3.1 Google Collab	9
2.3.2 RapidMiner	10
2.3.3 Microsoft Excel.....	10
2.4 Studi Literatur	10
BAB III METODOLOGI PENELITIAN	12
3.1 Tahapan Penelitian.....	12
3.1.1 Studi Literatur	12
3.1.2 Identifikasi Masalah.....	12
3.1.3 Analisa Masalah.....	13
3.1.4 Kesimpulan dan Saran	14
3.2 Perbedaan dari Penelitian Sebelumnya.....	14
3.3 <i>Data Preprocessing</i>	15
3.4 Penentuan Data Latih dan Data Uji	16
3.5 Pengembangan Model.....	16
3.6 Teknik Pengujian	16
BAB IV ANALISIS DAN PEMBAHASAN.....	17
4.1 Pengumpulan Data.....	17

4.2	<i>Data Preprocessing / Pra-Pemrosesan Data</i>	19
4.2.1	<i>Data Cleaning</i>	19
4.2.2	<i>Data Integration</i>	21
4.2.3	Transformasi Data	21
4.3	Pemodelan	29
4.3.1	Hasil Komparasi Model	29
4.3.2	Penyajian Model Terbaik	30
4.3.3	Perhitungan Algoritma Terpilih	34
4.4	Pengujian	39
BAB V PENUTUP		41
5.1	Kesimpulan	41
5.2	Saran	42
DAFTAR PUSTAKA		43
LAMPIRAN		45

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, pertumbuhan *e-commerce* di Indonesia berlangsung cukup pesat. Berdasarkan Survei *E-Commerce* yang dirilis Badan Pusat Statistik (BPS) pada tahun 2025, jumlah usaha *e-commerce* di Indonesia pada akhir tahun 2023 mencapai 3,82 juta usaha yang tersebar di 38 provinsi, dengan nilai transaksi sebesar Rp1.100,87 triliun atau naik 27,40 persen dibandingkan tahun sebelumnya. Kondisi ini mendorong semakin banyak pelaku UMKM untuk mengandalkan platform *marketplace* seperti Shopee sebagai saluran penjualan utama mereka. Salah satu sektor yang ikut terdampak adalah industri parfum, yang permintaannya dipengaruhi oleh faktor musiman, momen tertentu seperti Lebaran dan Natal, serta preferensi konsumen yang berbeda di setiap wilayah. Keragaman variasi produk seperti merek, volume, jenis aroma, dan segmen harga membuat analisis pola penjualan semakin kompleks dan tidak mungkin dikerjakan secara manual.

GWMM SHOP adalah salah satu UMKM yang bergerak di bidang penjualan parfum melalui platform Shopee. Selama beroperasi, toko ini telah menghasilkan data transaksi dalam jumlah yang cukup besar, mulai dari waktu pembelian, jenis produk yang terjual, hingga wilayah pengiriman pembeli. Namun, data tersebut belum pernah dimanfaatkan secara serius untuk mendukung pengambilan keputusan bisnis. Pengelolaan toko masih dilakukan secara konvensional, dan keputusan terkait stok, waktu promosi, serta target pasar masih bergantung pada perkiraan dan intuisi pemilik usaha. Akibatnya, potensi yang tersimpan dalam data transaksi historis belum pernah diolah secara sistematis.

Dari kondisi tersebut muncul tiga masalah yang konkret. Pertama, GWMM SHOP belum memahami produk mana yang paling banyak terjual sehingga pengadaan stok masih bersifat spekulatif. Kedua, toko ini belum memahami kapan lonjakan transaksi terjadi sehingga momen puncak seperti *harbolnas* atau akhir tahun sering belum terantisipasi dengan baik. Ketiga, GWMM SHOP belum memahami provinsi mana yang paling aktif bertransaksi sehingga strategi pemasaran belum dapat diarahkan ke wilayah yang paling potensial. Ketiga masalah ini bermuara pada satu persoalan yang sama, yaitu data transaksi yang sudah ada belum pernah diolah secara terstruktur. Dalam persaingan *marketplace* yang semakin ketat, kemampuan membaca data bukan lagi nilai tambah, melainkan kebutuhan dasar.

Pendekatan yang relevan untuk menjawab permasalahan ini adalah data mining, khususnya teknik *clustering*. *Clustering* memungkinkan pengelompokan data transaksi ke dalam kelompok-kelompok yang memiliki karakteristik serupa tanpa memerlukan label kategori sebelumnya, sehingga cocok digunakan pada data penjualan yang belum terstruktur. Algoritma K-Means dipilih karena efisien secara komputasi, mudah diimplementasikan, dan menghasilkan keluaran yang mudah diinterpretasikan untuk keperluan analisis bisnis. Penentuan jumlah kluster yang optimal dilakukan

menggunakan Elbow Method, sedangkan kualitas hasil clustering diukur menggunakan metrik Davies-Bouldin Index (DBI). Dengan pendekatan ini, pola penjualan berdasarkan dimensi waktu, produk, dan wilayah dapat diidentifikasi secara sistematis dari data transaksi GWMM SHOP.

Beberapa penelitian terdahulu telah membuktikan efektivitas K-Means dalam konteks analisis penjualan. (Dermawan et al., 2024) pada Toko Daun Indah di Shopee berhasil menemukan pola pengelompokan produk berdasarkan tingkat penjualan menggunakan K-Means dengan kerangka CRISP-DM yang divalidasi menggunakan Davies-Bouldin Index. (Riadi et al., 2023) juga berhasil mengelompokkan produk parfum ke dalam beberapa kluster berdasarkan karakteristik penjualan sehingga manajemen toko dapat menyusun strategi pengadaan yang lebih tepat. (Dona et al., 2025) menunjukkan bahwa pengelompokan produk kosmetik menggunakan K-Means dapat mencegah kehabisan stok sekaligus menjadi acuan perencanaan persediaan yang lebih akurat. Meski demikian, penelitian tersebut umumnya hanya berfokus pada satu dimensi analisis dan belum mengintegrasikan tiga dimensi sekaligus, yaitu waktu, produk, dan wilayah, dalam satu kerangka clustering yang komprehensif pada objek UMKM parfum berbasis Shopee.

1.2 Masalah

a. Identifikasi Masalah

Berdasarkan latar belakang dari penelitian ini maka dapat diidentifikasi permasalahan sebagai berikut:

- 1) Data transaksi penjualan pada GWMM SHOP belum dianalisis secara mendalam.
- 2) Belum tersedia pengelompokan pola penjualan berdasarkan waktu, produk, dan wilayah yang menyebabkan momen penting sering terlewat, pengelolaan stok berdasarkan produk yang paling laris, serta tidak mengetahui provinsi yang sering melakukan transaksi.
- 3) Pengambilan keputusan masih dilakukan secara spekulatif yang dapat menyebabkan kerugian bagi perusahaan.

b. Batasan Masalah

Batasan atau ruang lingkup masalah dalam penelitian ini yaitu:

- 1) Data yang digunakan dalam penelitian adalah data transaksi penjualan shopee pada GWMM SHOP selama 2 tahun terakhir dari tahun 2024 sampai 2025.
- 2) Analisis yang digunakan hanya berfokus pada 3 atribut utama yaitu produk, waktu (bulan), dan wilayah (provinsi).
- 3) Metode data mining yang digunakan pada penelitian ini menggunakan algoritma K-Means.
- 4) *Clustering* yang digunakan adalah sangat tinggi, tinggi, sedang dan rendah.
- 5) Penelitian bertujuan untuk mengelompokkan dan menganalisis pola penjualan, bukan melakukan prediksi (*forecasting*).

- 6) Hasil analisis yang diperoleh digunakan sebagai bahan rekomendasi dalam pengambilan keputusan, tanpa membahas implemementasi sistem secara *real-time*.

c. Rumusan Masalah

- 1) Bagaimana penerapan algoritma K-Means dalam mengelompokkan data penjualan pada GWMM SHOP?
- 2) Bagaimana pola penjualan yang terbentuk berdasarkan waktu, produk, dan wilayah?
- 3) Bagaimana hasil clustering dapat digunakan untuk mendukung pengambilan keputusan bisnis?

1.3 Tujuan dan Manfaat Penelitian

a. Tujuan

- 1) Mengolah data transaksi penjualan agar dapat menghasilkan informasi yang berguna.
- 2) Menerapkan algoritma K-Means *Clustering* pada data penjualan.
- 3) Mengelompokkan data penjualan berdasarkan pola tertentu.
- 4) Mengidentifikasi pola penjualan berdasarkan waktu, produk, dan wilayah.

b. Manfaat

Penelitian ini diharapkan bermanfaat bagi tiga pihak. Bagi GWMM SHOP, hasil penelitian memberikan gambaran berbasis data mengenai produk terlaris, waktu terjadinya lonjakan transaksi, dan provinsi dengan aktivitas pembelian tertinggi. Data tersebut dapat digunakan perusahaan sebagai dasar dalam mengambil keputusan operasional dan pemasaran yang lebih tepat sasaran. Bagi dunia akademik, penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang membahas penerapan algoritma *clustering* pada analisis penjualan UMKM berbasis *platform digital*. Sementara itu, bagi peneliti, penelitian ini menjadi kesempatan untuk mempraktikkan ilmu *data mining* dan algoritma K-Means yang selama ini hanya dipelajari secara teori di perkuliahan, kali ini diterapkan langsung pada permasalahan bisnis yang nyata dan terukur.

1.4 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II STUDI PUSTAKA

Bab ini menjelaskan tentang teori-teori yang mendasari penelitian, mulai dari konsep data mining, *tools* yang digunakan, algoritma K-Means *Clustering*, hingga tinjauan penelitian terdahulu yang relevan.

BAB III METODOLOGI PENELITIAN

Menjelaskan metodologi penelitian, termasuk kerangka kerja CRISP-DM yang digunakan, metode pengumpulan data, variabel penelitian, dan alur analisis.

BAB IV ANALISIS DAN PEMBAHASAN

Menyajikan hasil analisis dan pembahasan secara menyeluruh, mulai dari pengumpulan data, preprocessing, pemodelan K-Means, pengujian dengan Davies-Bouldin Index, hingga interpretasi setiap klaster yang terbentuk.

BAB V KESIMPULAN

Berisi kesimpulan yang menjawab seluruh rumusan masalah secara spesifik, serta saran untuk penelitian dan pengembangan lebih lanjut.

BAB II

STUDI PUSTAKA

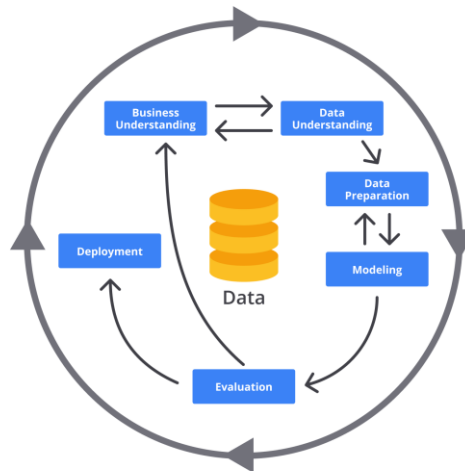
2.1 Teori – Teori Umum

2.1.1 Pengertian *Data Mining*

Menurut (Ilmu, 2024) *Data mining* secara sederhana dapat dipahami sebagai proses menggali informasi yang berguna dari kumpulan data berukuran besar yang belum terstruktur, dengan tujuan menemukan pola atau hubungan yang sebelumnya tidak terlihat. Dalam konteks bisnis, pendekatan ini berguna karena mampu mengolah volume data transaksi yang besar menjadi informasi yang bisa langsung dipakai untuk mengambil keputusan. Secara umum, proses *data mining* mencakup beberapa tahapan mulai dari seleksi data, *preprocessing*, transformasi, penerapan algoritma, sampai interpretasi hasil.

2.1.2 CRISP – DM

Menurut (Dermawan et al., 2024) CRISP-DM (*Cross Industry Standard Process for Data Mining*) adalah kerangka kerja analisis data yang sudah banyak dipakai di berbagai industri karena alurnya yang runtut dan tidak terlalu kaku untuk disesuaikan. Kerangka ini membawa peneliti melewati proses dari identifikasi masalah bisnis hingga penerapan hasil analisis ke dalam solusi yang bisa langsung digunakan, termasuk dalam konteks UMKM. Seperti terlihat pada Gambar 2.1, CRISP-DM mencakup 6 tahapan proses yang saling terhubung satu sama lain.



Gambar 2.1 CRISP-DM

a. *Business Understanding*

Tahap awal ini berfokus pada pemahaman tujuan serta kebutuhan proyek dari sudut pandang bisnis secara menyeluruh. Peneliti harus mampu menerjemahkan masalah nyata pemilik usaha menjadi tujuan teknis data mining yang tepat.

- b. *Data Understanding*
Tahap ini melibatkan pengumpulan data awal serta pemeriksaan kualitas data untuk mengenali karakteristik dasar dari informasi tersedia. Peneliti melakukan eksplorasi guna menemukan wawasan pertama atau mendeteksi adanya masalah seperti data tidak valid. Untuk menentukan variabel yang paling relevan untuk diolah pada tahap selanjutnya.
- c. *Data Preparation*
Persiapan data mencakup aktivitas pembersihan serta transformasi data mentah menjadi format yang siap untuk digunakan dalam proses modeling.
- d. *Modeling*
Tahap ini dilakukan dengan menerapkan algoritma *clustering* pada dataset yang telah disiapkan sebelumnya. Peneliti menentukan parameter optimal untuk mendapatkan hasil pengelompokan data yang paling akurat sesuai dengan pengujian model.
- e. *Evaluation*
Setelah model berhasil dibuat maka tahap evaluasi dilakukan untuk mengukur kualitas hasil pengelompokan menggunakan metrik pengujian. Peneliti meninjau kembali apakah hasil analisis sudah sesuai dengan tujuan bisnis yang telah ditetapkan pada tahap pertama.
- f. *Deployment*
Tahap akhir ini gunanya menginterpretasikan hasil data mining dalam bentuk laporan atau rekomendasi strategi untuk digunakan dalam pengambilan keputusan.

2.1.3 *K-Means*

Menurut (Fatkhudin et al., 2023) *K-Means* adalah metode *clustering* dalam kategori *unsupervised machine learning* yang bekerja dengan cara mengelompokkan data ke dalam sejumlah k kelompok berdasarkan kemiripan karakteristik antar datanya. Metode ini banyak dipilih karena cara kerjanya yang tidak terlalu rumit dan tidak membutuhkan komputasi yang berat. Dalam ranah data mining, *K-Means* masuk ke dalam kelompok metode *partitioning*, yaitu pendekatan yang memecah dataset menjadi beberapa bagian tanpa ada irisan di antara kelompok-kelompok tersebut sehingga setiap titik data hanya akan berada di satu *cluster*. Berikut metode dalam algoritma *K-Means*:

- a. Menentukan nilai k optimal.
- b. Hitung jarak antara setiap data dengan masing-masing centroid menggunakan rumus Euclidean Distance.
- c. Menentukan data mana yang masuk ke *cluster* tertentu, menggunakan pengukuran jarak antara setiap titik data dengan centroid menggunakan rumus *Euclidean Distance*:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Keterangan:

x_i = data kriteria.

μ_j = *centroid* pada *cluster* ke-j.

- d. Masukkan setiap data ke dalam *cluster* yang *centroid*-nya paling dekat.
- e. Setelah semua data ditetapkan ke *cluster* masing-masing, *centroid* setiap *cluster* dihitung ulang menggunakan rata-rata dari seluruh data yang ada di dalamnya:

$$C_k = \frac{1}{n_k} \sum d_i$$

Keterangan:

C_k = Titik pusat (*centroid*) baru untuk *cluster* ke-k.

n_k = Jumlah total titik data yang tergabung ke dalam *cluster* ke-k.

d_i = Titik data ke-i yang menjadi anggota dari *cluster* ke-k.

- f. Ulangi langkah 3 sampai 5 sampai posisi *centroid* tidak lagi berubah.

2.1.4 Metode Elbow

Metode Elbow adalah salah satu pendekatan yang umum digunakan untuk menentukan jumlah kluster optimal dalam algoritma K-Means *Clustering*. Cara kerjanya cukup langsung: algoritma K-Means dijalankan berulang kali dengan nilai k yang berbeda, mulai dari k=1 hingga batas nilai k yang telah ditentukan, lalu pada setiap percobaan dihitung nilai *Within Cluster Sum of Squares* (WCSS) atau yang sering disebut inerti. WCSS sendiri mengukur total kuadrat jarak antara setiap titik data dengan *centroid* klasternya masing-masing (Vania et al., 2023), yang secara matematis dituliskan sebagai:

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

WCSS = total jumlah kuadrat jarak seluruh data terhadap *centroid*-nya.

K = jumlah *cluster* yang sedang diuji.

C_k = himpunan data yang termasuk dalam *cluster* ke-k.

x_i = titik data ke-i dalam *cluster* C_k

μ_k = nilai *centroid* dari *cluster* ke-k

$\|x_i - \mu_k\|^2$ = kuadrat jarak Euclidean antara data x_i dengan *centroid* μ_k

2.1.5 Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan metode evaluasi *cluster* yang digunakan untuk menentukan jumlah *cluster* yang optimal dengan didukung pengukuran jarak yang tepat. Hasil akhir DBI berupa nilai

numerik yang mencerminkan kualitas keseluruhan dari proses *clustering* yang telah dilakukan. Semakin kecil nilai DBI yang dihasilkan, semakin baik kualitas *cluster* tersebut karena menandakan bahwa data dalam satu *cluster* sangat rapat satu sama lain sekaligus terpisah jauh dari *cluster* lainnya. Prinsip inilah yang menjadikan DBI sebagai salah satu metode evaluasi internal yang objektif karena tidak memerlukan data label eksternal untuk menilai kualitas *clustering*, sehingga sangat cocok dipadukan dengan algoritma K-Means yang bersifat *unsupervised*. (Jollyta et al., 2023) Adapun rumus untuk mengetahui nilai DBI sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{d(c_i, c_j)} \right)$$

Keterangan:

DBI = nilai akhir *Davies Bouldin Index*.

K = jumlah *cluster* yang terbentuk.

S_i = nilai *scatter* dari *cluster* ke- i

S_j = nilai *scatter* dari *cluster* ke- j

$d(\mu_i, \mu_j)$ = jarak antara centroid *cluster* ke- i dan ke- j

$\max_{j \neq i}$ = nilai rasio tertinggi terhadap semua *cluster* lain selain *cluster* ke- i

2.2 Teori Khusus

2.2.1 E-Commerce

E-commerce atau perdagangan elektronik merupakan bentuk transaksi jual beli barang maupun jasa yang dilakukan melalui media elektronik, khususnya dengan memanfaatkan jaringan internet sebagai sarana utamanya. Dalam perkembangannya, *e-commerce* tidak hanya mencakup aktivitas pembelian dan penjualan produk secara online, tetapi juga melibatkan berbagai proses bisnis seperti pemasaran digital, manajemen logistik, hingga layanan purna jual yang semuanya terintegrasi dalam satu ekosistem digital (Dasopang, 2024). Kehadiran *e-commerce* telah mengubah cara pelaku usaha berinteraksi dengan konsumen secara signifikan, karena transaksi dapat dilakukan kapan saja dan di mana saja tanpa terbatas oleh jarak maupun waktu.

2.2.2 Klasterisasi

Klasterisasi atau *clustering* merupakan salah satu teknik dalam analisis data yang bertujuan untuk mengelompokkan sekumpulan objek atau data ke dalam beberapa kelompok (*cluster*) berdasarkan tingkat kemiripan atau kedekatan karakteristiknya. Data yang berada dalam satu *cluster* memiliki kemiripan yang tinggi satu sama lain, sedangkan data yang berada di *cluster* berbeda memiliki karakteristik yang jauh berbeda. Proses pengelompokan ini dilakukan secara otomatis tanpa menggunakan label kelas yang telah ditentukan sebelumnya, sehingga *clustering* termasuk ke dalam kategori *unsupervised learning* dalam bidang *machine*

learning dan *data mining* (Khalif et al., 2024). Dalam konteks *e-commerce*, teknik *clustering* sangat berguna untuk mengelompokkan pelanggan berdasarkan pola perilaku transaksinya, sehingga hasil pengelompokan tersebut dapat dimanfaatkan oleh pelaku usaha untuk menyusun strategi pemasaran yang lebih tepat sasaran.

2.2.3 Transformasi Data

Menurut (Gori et al., 2024) Transformasi data adalah salah satu tahap dalam *preprocessing* yang mengubah format data agar kompatibel dengan algoritma yang akan dipakai. Secara umum, proses *preprocessing* mencakup beberapa langkah seperti integrasi data, pelabelan, transformasi, dan pembersihan data. Semua langkah ini dikerjakan sebelum algoritma *machine learning* mulai berjalan. Ketika dataset memiliki kolom bertipe kategorikal, salah satu teknik yang sering dipilih adalah *Label Encoding*. Teknik ini bekerja dengan mengonversi setiap nilai kategori pada suatu kolom menjadi angka, sehingga algoritma *machine learning* bisa membaca dan memproses kolom tersebut.

2.2.4 Normalisasi Data

Normalisasi data adalah proses menyamakan skala nilai pada setiap atribut numerik agar tidak ada satu atribut yang nilainya jauh lebih besar dari atribut lain dan secara tidak sengaja mendominasi hasil perhitungan. Ini menjadi langkah yang perlu dilakukan sebelum menjalankan algoritma berbasis jarak seperti K-Means, karena algoritma tersebut mengandalkan perhitungan jarak Euclidean yang sangat terpengaruh oleh perbedaan skala antara atribut. (Ramadhan et al., 2026) Dalam konteks penelitian berbasis K-Means, normalisasi dilakukan pada atribut numerik menggunakan *Min-Max Scaling* agar setiap atribut berada dalam rentang yang setara, sehingga bias dalam perhitungan jarak dapat ditekan. Metode yang digunakan adalah *Min-Max Scaler*, yaitu teknik yang mengubah setiap nilai data ke dalam rentang 0 hingga 1 menggunakan rumus berikut:

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Keterangan:

x' = Nilai data baru setelah dinormalisasi.

x_i = Nilai data asli yang sedang diproses.

$\min(x)$ = Nilai minimum dari suatu atribut.

$\max(x)$ = Nilai maksimum dari suatu atribut.

2.3 Tools Yang Digunakan

2.3.1 Google Collab

Google Colaboratory, atau Google Colab, adalah platform berbasis *cloud* yang dikembangkan oleh Google dan dapat diakses gratis melalui *browser* tanpa perlu menginstal perangkat lunak apapun. Pengguna bisa langsung mulai menulis kode Python tanpa setup awal, berbagi *notebook* dengan orang lain, dan memanfaatkan sumber daya komputasi untuk

kebutuhan pemrosesan data yang cukup berat. Platform ini berjalan di atas lingkungan Jupyter *Notebook*, sehingga pengguna bisa menulis kode, menjalankannya langsung, dan menyimpan hasilnya dalam satu dokumen yang terhubung dengan Google Drive (Wilyani et al., 2024).

2.3.2 RapidMiner

RapidMiner adalah perangkat lunak *open source* untuk analisis data yang dirilis di bawah lisensi AGPL versi 3. Platform ini sudah dipakai di lebih dari 40 negara dan mencakup berbagai kebutuhan dalam proses data mining, mulai dari *preprocessing* hingga evaluasi model. Yang membedakan RapidMiner dari tools lain adalah pendekatannya yang berbasis antarmuka visual *drag-and-drop*, sehingga alur kerja analisis bisa dibangun tanpa harus menulis kode dari awal. Kemudahan ini yang menjadi salah satu pertimbangan penggunaannya dalam penelitian ini, terutama untuk mempercepat proses eksplorasi dan pengujian model (Ananda et al., 2023).

2.3.3 Microsoft Excel

Microsoft Excel adalah perangkat lunak *spreadsheet* yang dikembangkan oleh Microsoft dan menjadi bagian dari paket Microsoft Office. Excel digunakan untuk mengolah data secara otomatis, mulai dari perhitungan sederhana, penggunaan rumus dan fungsi, pembuatan tabel dan grafik, hingga manajemen data untuk mendukung proses pengambilan keputusan (Pratiwi et al., 2023).

2.4 Studi Literatur

Penelitian mengenai pengelompokan data dalam konteks pengelolaan stok produk telah banyak dilakukan oleh berbagai peneliti sebelumnya, khususnya dengan menggunakan algoritma K-Means *Clustering*. Penelitian oleh (Wahyudi et al., 2022) yang berjudul "Implementasi Metode K-Means *Clustering* Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos" menunjukkan bahwa algoritma K-Means dapat dimanfaatkan untuk mengelompokkan data stok produk pada toko online berskala UMKM yang bergerak di bidang perdagangan kaos. Penelitian ini menggunakan data transaksi penjualan internal toko dengan atribut kode produk, nama produk, dan jumlah stok yang bergerak, kemudian dievaluasi menggunakan Davies-Bouldin Index. Hasil penelitian menunjukkan bahwa algoritma K-Means berhasil membentuk pengelompokan produk berdasarkan tingkat perputaran stok, sehingga dapat digunakan sebagai dasar pengambilan keputusan operasional dalam manajemen inventaris toko.

Penelitian yang dilakukan oleh (Kamalludin et al., 2023) berjudul "Penerapan Algoritme Klasterisasi K-Means Kinerja Produk dengan Analisis Recency Frequency Monetary pada Cafe XYZ" menerapkan pendekatan RFM (*Recency, Frequency, Monetary*) sebagai dasar pembentukan fitur sebelum dilakukan proses *clustering* terhadap data transaksi pada Cafe XYZ. Dataset yang digunakan memiliki atribut tanggal transaksi, id pelanggan,

nama produk, frekuensi pembelian, dan nilai transaksi. Hasil penelitian ini menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan kinerja produk pada usaha skala kecil, sehingga pengelola Cafe XYZ dapat menyusun strategi pemasaran yang lebih terarah berdasarkan karakteristik produk pada masing-masing klaster yang terbentuk.

Penelitian lain oleh (Fitri et al., 2023) berjudul "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means pada PT XYZ" menjadikan wilayah sebagai variabel utama dalam proses *clustering* data penjualan pada sebuah perusahaan besar. Dataset yang digunakan mencakup atribut wilayah penjualan, jumlah transaksi per wilayah, dan total nilai penjualan, dengan evaluasi klasterisasi menggunakan Silhouette Score. Hasil penelitian ini menunjukkan bahwa klasterisasi berbasis wilayah mampu memetakan distribusi penjualan secara geografis, sehingga informasinya dapat digunakan oleh PT XYZ sebagai dasar pengambilan keputusan strategis dalam menentukan target pasar di setiap wilayah.

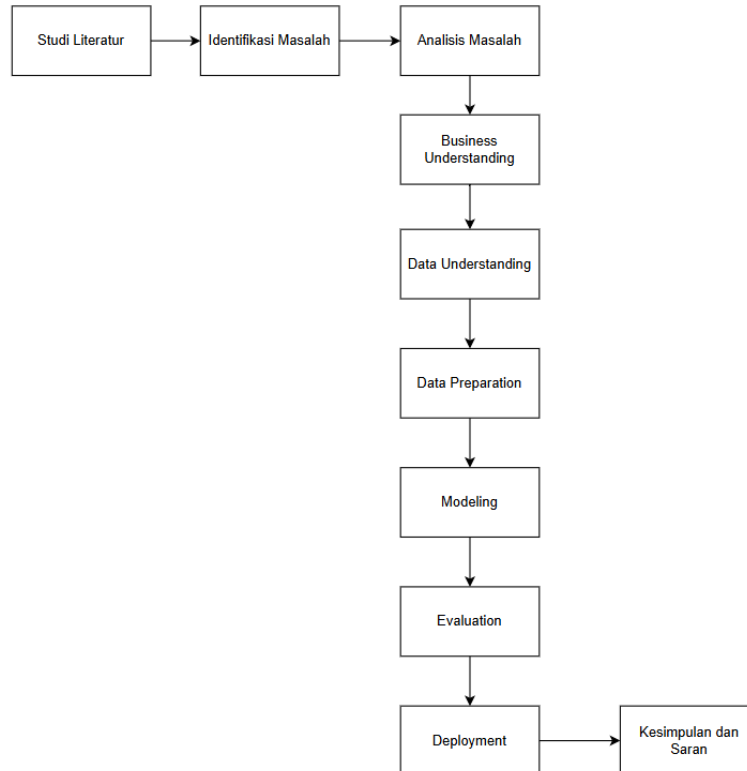
(Eriskiannisa Febrianty Luchia Awalina, 2023), dalam penelitiannya yang berjudul "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means *Clustering* pada Transaksi Online Retail", membahas penerapan algoritma K-Means dalam mensegmentasi pelanggan berdasarkan data transaksi pada platform online retail dengan atribut id pelanggan, tanggal pembelian, frekuensi transaksi, dan total nilai belanja. Hasil penelitian ini menunjukkan bahwa K-Means *Clustering* efektif diterapkan pada lingkungan *e-commerce*, di mana terbentuknya beberapa segmen pelanggan dengan karakteristik belanja yang berbeda dapat dimanfaatkan sebagai dasar penyusunan strategi pemasaran yang lebih terarah dan personal bagi setiap segmen.

Sebagai perbandingan terakhir, (Dermawan et al., 2024) melalui penelitiannya yang berjudul "Analisis Data Penjualan Menggunakan Algoritma K-Means *Clustering* Pada Toko Daun Indah di Shopee" menerapkan algoritma K-Means dengan kerangka metodologi CRISP-DM untuk menganalisis data penjualan sebuah toko yang beroperasi di platform Shopee. Dataset yang digunakan bersumber dari data ekspor Shopee Seller Center toko tersebut, dengan atribut nama produk, kategori produk, dan jumlah terjual, yang dievaluasi menggunakan Davies-Bouldin Index (DBI). Hasil penelitian ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan produk berdasarkan tingkat penjualannya, sehingga hasil klasterisasi tersebut dapat digunakan oleh pemilik Toko Daun Indah untuk menentukan strategi promosi dan pengelolaan stok produk yang lebih efektif di platform Shopee.

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini menggunakan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai kerangka kerja analisis data. CRISP-DM membagi proses penelitian ke dalam enam tahap yang berurutan, sebagaimana ditunjukkan pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

Berikut adalah penjelasan dari setiap tahapan yang digambarkan pada Gambar 3.1:

3.1.1 Studi Literatur

Studi literatur dilakukan untuk mengumpulkan referensi dari jurnal ilmiah, buku teks, dan laporan penelitian sebelumnya. Tahap ini digunakan untuk memahami teori *clustering*, cara kerja metode K-Means, dan konteks bisnis penjualan UMKM di *marketplace* digital. Selain itu, hasil studi literatur menjadi dasar dalam menyusun kerangka pemikiran dan mengidentifikasi celah penelitian yang membedakan penelitian ini dari penelitian sebelumnya.

3.1.2 Identifikasi Masalah

Berdasarkan studi literatur dan kondisi di lapangan, peneliti menemukan bahwa GWMM SHOP belum mengolah data transaksi penjualannya secara sistematis. Data transaksi yang tersedia di platform

Shopee belum pernah dianalisis untuk menemukan pola penjualan berdasarkan waktu, jenis produk, dan wilayah pengiriman. Dari kondisi ini, masalah penelitian dirumuskan sebagai kebutuhan untuk mengetahui pola penjualan yang bisa mendukung pengambilan keputusan bisnis secara berbasis data.

3.1.3 Analisa Masalah

Analisa masalah dilakukan melalui penerapan enam sub-tahap CRISP-DM:

- a. *Business Understanding*
Proses ini bertujuan memahami kebutuhan bisnis GWMM SHOP sebagai UMKM parfum yang berjualan di platform Shopee. Tujuan bisnis kemudian diterjemahkan ke dalam tujuan data mining, yaitu mengelompokkan pola penjualan berdasarkan waktu (bulan), produk, dan wilayah (provinsi) guna mendukung pengambilan keputusan operasional dan strategi pemasaran.
- b. *Data Understanding*
Mengeksplorasi data awal transaksi penjualan GWMM SHOP di Shopee selama periode 2024 hingga 2025. Eksplorasi dilakukan dengan memeriksa kualitas data, mengidentifikasi nilai yang hilang, dan memetakan atribut yang tersedia. Dari proses ini, ditetapkan tiga variabel utama yang akan digunakan dalam analisis, yaitu waktu transaksi (bulan), nama produk, dan wilayah pembeli (provinsi).
- c. *Data Preparation*
Melakukan seleksi variabel penting (bulan, produk, provinsi, dan jumlah transaksi), pembersihan data (menangani data hilang, duplikat, dan inkonsistensi), serta transformasi data berupa encoding kategorikal dan normalisasi numerik menggunakan metode Min-Max agar kompatibel dengan algoritma K-Means.
- d. *Modeling*
Menerapkan algoritma K-Means pada dataset yang telah disiapkan dan menentukan jumlah kluster optimal menggunakan Elbow Method. Setiap kluster yang terbentuk diidentifikasi karakteristiknya berdasarkan dimensi waktu, produk, dan wilayah.
- e. *Evaluation*
Mengevaluasi hasil klasterisasi menggunakan Davies-Bouldin Index (DBI) untuk mengukur kualitas pengelompokan secara kuantitatif. Nilai DBI yang lebih kecil menunjukkan hasil klasterisasi yang lebih baik. Selain itu dilakukan validasi visual untuk memastikan kluster yang terbentuk sesuai dengan karakteristik bisnis yang diharapkan.
- f. *Deployment*
Menyusun laporan hasil analisis dan memberikan rekomendasi strategi bisnis kepada GWMM SHOP berdasarkan karakteristik setiap kluster yang terbentuk, mencakup rekomendasi terkait manajemen stok, waktu promosi, dan target pemasaran berdasarkan wilayah.

3.1.4 Kesimpulan dan Saran

Bagian ini menyajikan temuan dari seluruh tahapan CRISP-DM, mencakup pola klaster yang terbentuk berdasarkan dimensi waktu, produk, dan wilayah. Dari temuan tersebut, dihasilkan rekomendasi praktis bagi GWMM SHOP terkait pengadaan stok, penentuan waktu promosi, dan penentuan wilayah target pemasaran yang paling potensial.

3.2 Perbedaan dari Penelitian Sebelumnya

Penelitian oleh (Wahyudi et al., 2022) menerapkan K-Means untuk mengelompokkan produk berdasarkan tingkat perputaran stok pada toko online berskala UMKM berjudul "Implementasi Metode K-Means *Clustering* Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos". Hasilnya membuktikan bahwa K-Means relevan digunakan pada konteks produk toko online skala kecil. Namun, analisisnya hanya mencakup satu dimensi, yaitu pergerakan stok, tanpa mempertimbangkan waktu transaksi maupun sebaran wilayah pembeli. Penelitian ini mengisi celah tersebut dengan mengintegrasikan tiga dimensi sekaligus dalam satu kerangka analisis.

Penelitian oleh (Kamalludin et al., 2023) berjudul "Penerapan Algoritme Klasterisasi K-Means Kinerja Produk dengan Analisis Recency Frequency Monetary pada Cafe XYZ" menggunakan pendekatan RFM sebagai dasar pembentukan fitur sebelum proses *clustering* dijalankan. Penelitian ini memperlihatkan bahwa K-Means relevan digunakan untuk menganalisis kinerja produk pada usaha skala kecil. Namun, cakupannya masih terbatas pada segmentasi kinerja produk secara umum dan belum membahas pola penjualan berdasarkan waktu maupun sebaran wilayah transaksi, sehingga hasil yang diperoleh tidak mencerminkan pola temporal maupun geografis penjualan.

Selanjutnya, penelitian yang dilakukan (Fitri et al., 2023) berjudul "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means pada PT XYZ" secara eksplisit menjadikan wilayah sebagai variabel utama dalam *clustering* penjualan, dengan hasil klasterisasi yang dievaluasi menggunakan Silhouette Score. Penelitian ini menjadi referensi penting karena menunjukkan bahwa dimensi wilayah dapat dijadikan dasar pengelompokan yang bermakna. Namun, analisis hanya mencakup satu dimensi wilayah tanpa mengintegrasikan dimensi waktu transaksi maupun kategori produk, serta dilakukan pada perusahaan besar dan bukan UMKM berbasis platform marketplace digital seperti Shopee.

Penelitian oleh (Eriskiannisa Febrianty Luchia Awalina, 2023) menerapkan K-Means untuk mensegmentasi pelanggan pada data transaksi online retail dengan mempertimbangkan variabel waktu pembelian dan frekuensi transaksi yang berjudul "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means *Clustering* pada Transaksi Online Retail". Hasilnya menunjukkan bahwa K-Means efektif digunakan dalam lingkungan *e-commerce*. Meski demikian, penelitian ini hanya berfokus pada segmentasi pelanggan secara umum, tanpa menyertakan

analisis distribusi geografis pembeli per wilayah, klasifikasi produk secara spesifik, maupun kerangka kerja CRISP-DM dalam proses analisisnya.

Penelitian oleh (Dermawan et al., 2024) berjudul "Analisis Data Penjualan Menggunakan Algoritma K-Means *Clustering* Pada Toko Daun Indah di Shopee" membuktikan bahwa kombinasi K-Means, kerangka CRISP-DM, dan *Davies-Bouldin Index* efektif diterapkan pada data penjualan toko di Shopee. Penelitian ini paling relevan dengan konteks penelitian yang sedang dilakukan karena menggunakan objek yang sama, yaitu toko di *platform* Shopee, serta alat evaluasi yang sama, yaitu *Davies-Bouldin Index*. Namun, cakupannya masih terbatas pada pengelompokan berdasarkan atribut produk dan variasi nama, tanpa mengintegrasikan dimensi waktu transaksi dan wilayah pengiriman sebagai variabel analisis.

3.3 *Data Preprocessing*

Data penelitian ini diperoleh dari toko GWMM Shop pada platform Shopee dan mencakup seluruh transaksi penjualan produk parfum periode Januari 2024 hingga Desember 2025. Dataset terdiri dari 4.163 transaksi dengan total 42 atribut, yang memuat informasi nomor pesanan, nama produk, jenis produk, harga awal, harga setelah diskon, jumlah transaksi, metode pembayaran, opsi pengiriman, serta lokasi penerima berdasarkan kota dan provinsi. Adapun langkah *preprocessing* meliputi:

a. *Data Cleansing*

Pada tahap data *cleaning*, dilakukan beberapa proses untuk memastikan kualitas data. Pertama, dilakukan pengecekan nilai kosong pada seluruh atribut dataset, dan data dengan nilai kosong pada atribut nomor pesanan dihapus karena tidak dapat digunakan dalam proses *clustering*. Kedua, data duplikat yang merepresentasikan transaksi sama dihapus. Ketiga, dilakukan pemilihan fitur kolom yang digunakan, yaitu nama produk, tanggal, nama provinsi, jumlah transaksi, dan total pembayaran. Keempat, kolom tanggal diubah ke format bulan.

b. *Data Integration*

Data yang digunakan dalam penelitian ini bersumber dari satu file ekspor Shopee yang sudah mencakup seluruh informasi yang diperlukan, sehingga tidak diperlukan proses penggabungan data dari sumber yang berbeda.

c. *Data Transformation*

Kemudian pada tahap *data transformation*, dilakukan beberapa penyesuaian agar data dapat digunakan dalam pemodelan K-Means. Pertama, data kategorikal pada atribut nama produk dan provinsi diubah ke dalam bentuk numerik menggunakan teknik *encoding*. Terakhir, dilakukan proses normalisasi menggunakan metode *Min-Max* terhadap atribut jumlah transaksi, total pembayaran, bulan, nama produk dan provinsi yang sudah dikonversi menjadi numerik. Langkah ini dilakukan agar semua atribut berada dalam skala yang sama dan tidak mendominasi proses klusterisasi.

3.4 Penentuan Data Latih dan Data Uji

Penelitian ini menggunakan algoritma K-Means *Clustering* yang merupakan bagian dari *unsupervised learning*. Karena tidak terdapat variabel target atau label, seluruh data yang tersedia digunakan langsung dalam proses pemodelan tanpa pemisahan antara data latih dan data uji seperti pada *supervised learning*.

3.5 Pengembangan Model

Pengembangan model dilakukan dengan menerapkan algoritma K-Means *Clustering* menggunakan Google Colaboratory dan RapidMiner. Jumlah kluster optimal ditentukan melalui *Elbow Method* dengan menghitung nilai *Within Cluster Sum of Squares* (WCSS) pada berbagai nilai k, lalu memilih titik di mana penurunan WCSS mulai melandai. Setelah jumlah kluster ditentukan, proses klusterisasi dijalankan di RapidMiner menggunakan operator K-Means dengan konfigurasi parameter yang sesuai, termasuk jumlah kluster dan inisialisasi *centroid*. Seluruh data yang telah dinormalisasi digunakan langsung karena K-Means tidak memerlukan pemisahan data latih dan data uji. Hasil klusterisasi kemudian dianalisis untuk mengidentifikasi karakteristik dominan tiap kluster berdasarkan dimensi waktu, produk, dan wilayah, sehingga dapat memberikan wawasan yang dapat diinterpretasikan oleh GWMM Shop.

3.6 Teknik Pengujian

Teknik pengujian dalam penelitian ini menggunakan pendekatan gabungan antara *Elbow Method* dan *Davies-Bouldin Index* (DBI) untuk menentukan jumlah kluster yang paling optimal. *Elbow Method* digunakan untuk mendapatkan estimasi awal jumlah kluster dengan memvisualisasikan titik siku pada grafik hubungan antara nilai k dan WCSS, yang kemudian menjadi acuan awal jumlah kluster yang diuji lebih lanjut. Hasil estimasi tersebut divalidasi secara kuantitatif menggunakan DBI, yang mengevaluasi kualitas kluster dengan mengukur rasio antara kepadatan di dalam kluster dan pemisahannya dengan kluster lain, sehingga jumlah kluster dengan nilai DBI terendah dipilih sebagai solusi akhir. Selain pengujian kualitas kluster, dilakukan pula perhitungan manual algoritma K-Means menggunakan minimal 5 data sampel dari dataset yang telah dinormalisasi. Perhitungan manual ini mencakup penetapan centroid awal, perhitungan jarak Euclidean setiap data terhadap masing-masing centroid, pengelompokan data ke kluster terdekat, dan pembaruan nilai centroid hingga konvergensi tercapai. Hasil perhitungan manual kemudian dibandingkan dengan *output* RapidMiner untuk memastikan konsistensi proses klusterisasi.

BAB IV ANALISIS DAN PEMBAHASAN

4.1 Pengumpulan Data

Penelitian ini menggunakan data transaksi penjualan nyata dari toko GWMM SHOP, sebuah toko yang berjualan produk parfum di platform *e-commerce* Shopee. Data diperoleh langsung melalui fitur ekspor data pada *dashboard* penjual (*Seller Center*) Shopee, di mana pemilik toko dapat mengunduh riwayat pesanan dalam format Excel (.xlsx). Proses pengunduhan dilakukan secara manual oleh peneliti dengan masuk ke akun penjual GWMM SHOP dan mengekspor seluruh data transaksi berstatus Selesai dalam periode yang telah ditentukan.

Dataset yang diperoleh terdiri dari 4.163 baris data dengan 42 atribut dengan rentang waktu dari bulan Januari 2024 sampai Desember 2025. Atribut tersebut mencakup informasi tiap transaksi penjualan seperti nomor pesanan, nama produk, jumlah pembelian, metode pembayaran, dan lokasi pengiriman pelanggan. Data ini merupakan data primer yang bersumber langsung dari aktivitas bisnis nyata GWMM SHOP, sehingga mencerminkan pola transaksi dan perilaku pembelian pelanggan yang aktual. Adapun atribut yang terdapat dalam dataset GWMM SHOP sebagai berikut:

Tabel 4.1 Atribut Dataset GWMM SHOP

No.	Nama Atribut	Keterangan
1.	No. Pesanan	Nomor unik identifikasi setiap pesanan di Shopee
2.	Status Pesanan	Status akhir pesanan (Selesai, Dibatalkan, dll.)
3.	No. Resi	Nomor resi pengiriman dari ekspedisi
4.	Opsi Pengiriman	Jenis layanan pengiriman yang dipilih pembeli
5.	Antar Ke Counter/Pick up	Metode pengambilan paket oleh kurir
6.	Metode Pembayaran	Cara pembayaran yang digunakan (SPayLater, COD, Online Payment, dll.)
7.	SKU Induk	Kode stok produk (tidak terisi pada dataset ini)
8.	Nama Produk	Nama lengkap produk yang dibeli
9.	Jenis Produk	Kategori jenis produk
10.	Harga Awal	Harga produk sebelum diskon (Rp)
12.	Harga Setelah Diskon	Harga produk setelah diskon diterapkan (Rp)
13.	Jumlah	Kuantitas produk yang dipesan
14.	Returned Quantity	Jumlah produk yang dikembalikan
15.	Total Harga Produk	Total nilai harga produk dalam satu pesanan (Rp)
16.	Total Diskon	Total potongan diskon yang diperoleh pembeli (Rp)
17.	Diskon Dari Penjual	Potongan diskon yang ditanggung penjual (Rp)

18.	Diskon Dari Shopee	Potongan diskon yang ditanggung Shopee (Rp)
19.	Berat Produk	Berat produk dalam satuan gram
20.	Jumlah Produk di Pesan	Total unit produk dalam satu nota pesanan
21.	Total Berat	Total berat keseluruhan produk dalam pesanan
22.	Voucher Ditanggung Penjual	Nilai voucher yang menjadi beban penjual (Rp)
23.	Cashback Koin	Jumlah koin Shopee yang dikembalikan ke pembeli
24.	Voucher Ditanggung Shopee	Nilai voucher yang ditanggung Shopee (Rp)
25.	Paket Diskon	Keterangan paket diskon yang berlaku
26.	Paket Diskon (Diskon dari Shopee)	Nilai diskon Shopee dari paket bundling (Rp)
27.	Paket Diskon (Diskon dari Penjual)	Nilai diskon penjual dari paket bundling (Rp)
28.	Potongan Koin Shopee	Jumlah koin Shopee yang digunakan sebagai pembayaran (Rp)
29.	Diskon Kartu Kredit	Potongan khusus kartu kredit tertentu (Rp)
30.	Ongkos Kirim Dibayar Pembeli	Biaya pengiriman yang dibayar oleh pembeli (Rp)
31.	Estimasi Potongan Biaya Pengiriman	Estimasi subsidi ongkir yang diberikan platform (Rp)
32.	Ongkos Kirim Pengembalian Barang	Biaya ongkir untuk proses retur produk (Rp)
33.	Total Pembayaran	Total nilai yang dibayarkan oleh pembeli (Rp)
34.	Perkiraan Ongkos Kirim	Estimasi total biaya pengiriman (Rp)
35.	Catatan dari Pembeli	Pesan atau instruksi khusus dari pembeli
36.	Catatan	Catatan tambahan dari sistem Shopee
37.	Username (Pembeli)	Nama pengguna akun Shopee pembeli
38.	Nama Penerima	Nama lengkap penerima paket
39.	No. Telepon	Nomor telepon penerima
40.	Alamat Pengiriman	Alamat lengkap tujuan pengiriman
41.	Kota/Kabupaten	Kota atau kabupaten tujuan pengiriman
42.	Provinsi	Provinsi tujuan pengiriman

Berikut pada Gambar 4.1 adalah contoh data transaksi awal dari GWMM SHOP sebelum dilakukan proses pembersihan dan transformasi.

No. Baris	Detail Pengiriman	Alamat ke counter/pick up	Transaksi	Metode Pembayaran	Uraian Produk	Isi Produk	Range Awal
1	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-02	COD (Bayar di Tempat)	Vial Parfum 5ml
2	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-03	Saldo Shopeepay	Extrait De Parfum 50ml
3	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-03	Saldo Shopeepay	Eau De Parfum 100ml
4	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-04	Saldo Shopeepay	Omenta The Winner Perfume Pria Eau De Toilette 100 ML
5	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-04	Online Payment	Fogg Perfume Midnight Purple 100ml Eau De Parfum
6	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-04	COD (Bayar di Tempat)	Exhange Extrait De Parfum Superwoman Series 50ml - Terra Nova
7	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-05	Alfamart/Alfamidi/Dan+Dan	Fogg Perfume Midnight Purple 100ml Eau De Parfum
8	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-05	Online Payment	Omenta Pink Nal Woman Eau De Parfum 100 ML
9	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-08	Online Payment	Omenta The Winner Perfume Pria Eau De Toilette 100 ML
10	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-11	COD (Bayar di Tempat)	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempua
11	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-12	Seabank Bayar Instan	Fogg Perfume Exot Premium Many Flowers 200ml Eau De Parfum
12	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-12	Saldo Shopeepay	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempua
13	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-12	Saldo Shopeepay	Mykonos Baby Love Parfum EDP 50 ML
14	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-13	Kartu Kredit/Debit	Extrait De Parfum 250ml
15	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-15	Saldo Shopeepay	Onix Fragrance Eau de Parfum Sencoperty 30ml
16	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-15	Kartu Kredit/Debit	ICE FROST Extrait De Parfum 30ml
17	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-16	Seabank Bayar Instan	Mykonos Milk Drops 50ml
18	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-16	Seabank Bayar Instan	Mykonos Milk Drops 50ml
19	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-17	Saldo Shopeepay	BRICCO - Rinsu EDT Parfum 40ml
20	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-18	Saldo Shopeepay	Fogg Perfume Midnight Purple 100ml Eau De Parfum
21	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-18	Kartu Kredit/Debit	ICE FROST Extrait De Parfum 30ml
22	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-18	Online Payment	Exhange Extrait De Parfum Superwoman Series 50ml - Terra Nova
23	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-19	Online Payment	Kokulu Women Body Selter - Bundling 3pcs 50ML
24	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-20	Kartu Kredit/Debit	Mykonos Milk Drops 50ml
25	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-20	Seabank Bayar Instan	Parfum Parfums 100ml
26	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-22	COD (Bayar di Tempat)	Fragrance World Suits Unisex Eau De Parfum 100 ML
27	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-23	Saldo Shopeepay	Onix Fragrance Eau de Parfum Sencoperty 30ml
28	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-23	Online Payment	Refill Parfum 30ml
29	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-23	Kartu Kredit/Debit	Mykonos Milk Drops 50ml
30	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-25	Online Payment	Cosplay - Rinsu Day 1 Blooming Series Extrait De Parfum
31	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-25	Saldo Shopeepay	Russ Perfume Unimonts Blue Unisex 50ml
32	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-25	Online Payment	SAINTI Perfume - Naughty Friday - EDP 30ml
33	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-25	Seabank Bayar Instan	Dynamical Course Pour Homme Eau EDP 100 ML
34	Selesai	501138050940482	Same Day-Godend Same Day	Pickup	2024-01-25	Seabank Bayar Instan	Fragrance World Suits Unisex Eau De Parfum 100 ML

Gambar 4.1 Dataset Mentah GWMM SHOP

4.2 Data Preprocessing / Pra-Pemrosesan Data

Sebelum data diolah lebih lanjut, diperlukan tahap pra-pemrosesan data agar hasil analisis *data mining* lebih optimal. Pada penelitian ini, pra-pemrosesan mencakup tiga proses utama yaitu *data cleaning* terhadap data transaksi penjualan GWMM SHOP, *data integration*, dan transformasi data. Ketiga proses ini dilakukan untuk memastikan data siap digunakan pada tahap pemodelan klusterisasi menggunakan algoritma *K-Means*.

4.2.1 Data Cleaning

Tahap pertama dalam preprocessing adalah pembersihan data (*data cleaning*). Langkah ini mencakup dua sub proses utama, yaitu penghapusan data duplikat, penanganan nilai yang hilang (*missing values*).

a. Penghapusan Data Duplikat

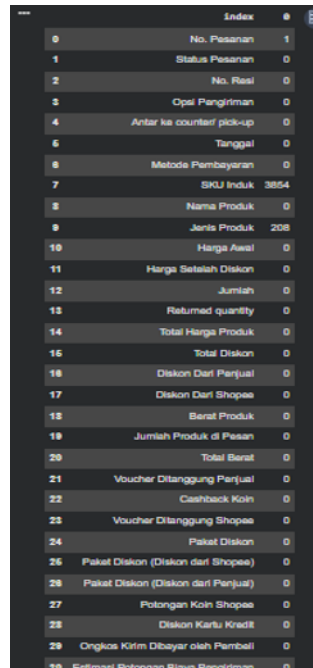
Pada tahap ini dilakukan dilakukan terhadap dataset transaksi penjualan GWMM SHOP yang diekspor dari platform Shopee. Dataset awal terdiri dari 4.163 baris dengan 42 atribut. Pengecekan duplikasi data dilakukan berdasarkan kolom No. Pesanan dan ditemukan sebanyak 309 baris duplikat. Setelah baris duplikat dihapus, jumlah data bersih yang tersisa menjadi 3.854 baris seperti pada Tabel 4.2.

Tabel 4.2 Hasil Penghapusan Data Duplikat

Kondisi	Jumlah Baris	Keterangan
Sebelum penghapusan data duplikat	4.163 baris	Data Mentah Awal
Baris duplikat yang dihapus	309 baris	Teridentifikasi duplikat berdasarkan No. Pesanan
Setelah penghapusan data duplikat	3.854 baris	Data siap diproses lebih lanjut

b. Penanganan *Missing Values* / Nilai Kosong

Pengecekan *missing values* dilakukan pada seluruh kolom dataset. Hasilnya, kolom SKU Induk ditemukan kosong seluruhnya dengan 3.854 missing values karena kolom ini memang tidak digunakan oleh GWMM SHOP dalam pengelolaan toko. Kolom Jenis Produk juga memiliki nilai kosong sebanyak 208 baris. Ringkasan kondisi missing values pada kolom utama disajikan pada Gambar 4.2 sebagai berikut.



Index		
0	No. Pesanan	1
1	Status Pesanan	0
2	No. Resi	0
3	Opsi Pengiriman	0
4	Antar ke counter/pick-up	0
5	Tanggal	0
6	Metode Pembayaran	0
7	SKU Induk	3854
8	Nama Produk	0
9	Jenis Produk	208
10	Harga Awal	0
11	Harga Setelah Diskon	0
12	Jumlah	0
13	Returned quantity	0
14	Total Harga Produk	0
15	Total Diskon	0
16	Diskon Dari Penjual	0
17	Diskon Dari Shopee	0
18	Berat Produk	0
19	Jumlah Produk di Pesan	0
20	Total Berat	0
21	Voucher Ditunggung Penjual	0
22	Cashback Koin	0
23	Voucher Ditunggung Shopee	0
24	Paket Diskon	0
25	Paket Diskon (Diskon dari Shopee)	0
26	Paket Diskon (Diskon dari Penjual)	0
27	Potongan Koin Shopee	0
28	Diskon Kartu Kredit	0
29	Ongkos Kirim Dibayar oleh Pembeli	0
30	Estimasi Potongan Biaya Pengiriman	0

Gambar 4.2 Hasil Pengecekan *Missing Values*

Berdasarkan hasil pengecekan tersebut Kolom SKU Induk dihapus karena tidak memiliki data sama sekali dan tidak relevan untuk analisis. Kolom lain yang tidak relevan untuk klasterisasi, seperti No. Resi, Opsi Pengiriman, Antar ke *counter/pick-up*, dan sejumlah atribut diskon yang tidak signifikan, juga ikut dihapus. Penghapusan atribut-atribut ini dilakukan untuk menyederhanakan dataset agar lebih fokus pada fitur yang merepresentasikan pola transaksi dan perilaku pelanggan GWMM SHOP.

c. Standarisasi Kolom Provinsi

Dalam penelitian ini, standarisasi nama provinsi dilakukan dengan membuat kamus acuan yang merujuk pada 34 nama provinsi resmi sesuai pembagian wilayah administratif Indonesia. Setiap variasi penulisan disesuaikan ke bentuk baku, di mana nama "Jakarta" dalam berbagai variasinya distandarisasi menjadi "DKI JAKARTA" sesuai nama resmi provinsi tersebut. Pendekatan ini sejalan dengan prinsip data cleaning untuk data kategorikal, yaitu nilai yang tidak konsisten perlu diperiksa lebih dulu melalui pengecekan nilai unik, kemudian

digantikan atau dikelompokkan sesuai kategori yang seharusnya. Hal serupa juga berlaku pada data yang digunakan dalam penelitian ini. Setelah dilakukan pemeriksaan terhadap 4.163 baris data, field Provinsi ditemukan memiliki 15 nilai unik yang seluruhnya sudah konsisten dengan format penulisan resmi, termasuk wilayah DKI Jakarta yang dalam dataset ini sudah tercatat secara baku sebagai "DKI JAKARTA".

d. Pemilihan Fitur Kolom

Dataset awal GWMM SHOP memiliki 42 atribut yang mencakup berbagai informasi transaksi, mulai dari data operasional pengiriman, informasi pembeli, rincian diskon dan voucher, hingga data produk dan lokasi. Tidak semua *atribut* tersebut relevan untuk tujuan klasterisasi pola penjualan yang menjadi fokus penelitian ini. Adapapun atribut yang terpilih untuk digunakan dalam analisis meliputi kolom tanggal, Nama Produk, Provinsi, Jumlah, Total Harga Produk seperti yang ditampilkan pada Gambar 4.3.

	Tanggal	Nama Produk	Provinsi	Jumlah	Total Harga Produk
0	2025-03-15	FYNE Extrait De Parfume	NUSA TENGGARA TIMUR (NTT)	1	89000
1	2025-03-06	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	NUSA TENGGARA TIMUR (NTT)	2	218000
2	2024-12-14	Fragrance World Cocktail Intense Unisex Eau De...	BANTEN	4	1064000
3	2024-12-02	Verbena Night Legend Man Eau De Parfum 100 ML	JAWA BARAT	5	610000
4	2025-07-22	BRODO - Rote EDT Parfume 40ml	BALI	2	198000

Gambar 4.3 Pemilihan Fitur Kolom Terpakai

4.2.2 Data Integration

Pada tahap ini, *Data integration* secara umum dilakukan dengan menggabungkan data dari berbagai sumber berbeda menjadi satu dataset yang utuh. Dalam penelitian ini, proses tersebut tidak perlu dilakukan karena data sudah tersedia dalam satu file yang diekspor langsung dari *Seller Center Shopee*. File tersebut sudah mencakup seluruh informasi transaksi yang dibutuhkan, mulai dari data produk, jumlah pembelian, total harga, hingga lokasi pengiriman. Karena semua informasi sudah tergabung dalam satu sumber, dataset langsung bisa diproses pada tahap transformasi berikutnya.

4.2.3 Transformasi Data

Transformasi data adalah tahap dalam *preprocessing* yang mengubah data mentah menjadi format yang relevan untuk dianalisis. Pada penelitian ini, transformasi mencakup beberapa langkah. Pertama, Kolom tanggal diubah menjadi representasi bulan dalam bentuk angka agar lebih mudah dianalisis. Kolom nama produk dan provinsi yang awalnya berbentuk kategorikal dikonversi ke format numerik menggunakan metode *encoding*, karena algoritma K-Means hanya dapat memproses data dalam bentuk angka. Setelah encoding, dilakukan normalisasi data menggunakan *Min-Max Normalization* untuk menyetarakan skala antar atribut, sehingga

atribut dengan nilai besar tidak mendominasi perhitungan jarak dalam proses klasterisasi.

a. Ekstraksi Fitur Waktu

Pada tahap awal transformasi data, mengubah tanggal menjadi Bulan yang diperoleh dari kolom Tanggal. Kolom ini menghasilkan angka bulan dalam format numerik, seperti 1 untuk Januari, 2 untuk Februari, dan seterusnya hingga 12 untuk Desember. Penambahan atribut ini bertujuan agar analisis pola penjualan bulanan toko GWMM SHOP dapat dilakukan. Informasi bulan perlu diikutsertakan karena volume transaksi bisa berbeda antar bulan.

Tabel 4.3 Contoh Ekstraksi Fitur Waktu

Tanggal	Bulan
01-01-2024 / 01-01-2025	1
01-02-2024 / 01-02-2025	2
01-03-2024 / 01-03-2025	3
01-04-2024 / 01-04-2025	4
01-05-2024 / 01-05-2025	5
01-06-2024 / 01-06-2025	6
01-07-2024 / 01-07-2025	7
01-08-2024 / 01-08-2025	8
01-09-2024 / 01-09-2025	9
01-10-2024 / 01-10-2025	10
01-11-2024 / 01-11-2025	11
01-12-2024 / 01-12-2025	12

b. Mengubah Data Kategorikal Menjadi Numerikal

Tahap ini mencakup pengubahan seluruh atribut yang digunakan dalam klasterisasi ke format *numerik*. Proses dimulai dari kolom Nama Produk yang sebelumnya berisi data teks, kemudian setiap nama produk parfum diganti dengan nilai numerik seperti yang ditampilkan pada Tabel 4.4.

Tabel 4.4 Transformasi Kategori Nama Produk

Nama Produk	Nilai Numerik
(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0
Aksesoris Gantungan Parfum	1
BRODO - Rote EDT Perfume 40ml	2
Crusita - Busy Day Blooming Series Extrait De Perfume	3
Dynamique Course Pour Homme Men EDP 100 ML	4
FYNE Extrait De Perfume	5
Fogg Perfume Midnight Purple 100ml Eau De Perfume	6

Fogg Perfume Scent Premium Many Flowers 100ml Eau De Perfume	7
Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	8
Fragrance World Suits Unisex Eau De Parfum 100 ML	9
GARVI Perfume - Naughty Friday - EDP 50ml	10
HINT IDENTITY Eau de Parfum 50ml	11
HINT Noble Eau de Parfum 50ml	12
ICE FROST Extrait De Perfume 35ml	13
Kokulu Women Best Seller - Bundling 2pcs 50ML	14
Mezuka Perfume Aqua Kiss Eau de Perfume 50ml	15
Mykonos Baby Love Parfum EDP 50 ML	16
Mykonos Milk Drips 50ml	17
Mykonos Verve Extrait De Parfum 50 ML	18
NARSIS Hypnotic Desire Eau de Perfume 50ml	19
Omerta Pink Ice Woman Eau de Parfum 100 ML	20
Omerta The Winner Perfume Pria Eau de Toilette 100 ML	21
Onix Fragrance Eau de Perfume Senoparty 30Ml	22
Pouch Parfum Mini	23
Refill Parfum 30ml	24
Revlon Charlie Fragrance - White EDT Parfum 100 ml	25
Russ Perfume Unomento Blue Unisex 50ml	26
SAFF & Co. Extrait de Parfum - CASCABEL 30ml	27
Sample Pack 3ml x5	28
Scentplus Dominant 35ml	29
Verbena Night Legend Man Eau De Parfum 100 ML	30
Vial Parfum 35ml	31
Xchange Extrait De Perfume Supernova Series 50ml - Terra Nova	32

Selain itu, pengubahan atribut provinsi juga dilakukan, yang kemudian diubah dari format teks menjadi angka. Konversi ini

bertujuan agar nilai provinsi dapat diproses dalam algoritma K-Means. Berikut hasil konversinya terdapat pada Tabel 4.5.

Tabel 4.5 Transformasi Kategori Provinsi

Provinsi	Nilai Numerik
Bali	0
Banten	1
Bengkulu	2
DKI Jakarta	3
Gorontalo	4
Jawa Barat	5
Jawa Tengah	6
Jawa Timur	7
Kalimantan Timur	8
Maluku	9
Nusa Tenggara Timur (NTT)	10
Papua	11
Riau	12
Sulawesi Selatan	13
Sumatra Utara	14

Pengubahan ke format *numerik* diperlukan karena algoritma K-Means hanya dapat memproses data dalam bentuk angka. Dengan langkah ini, seluruh informasi dari kolom Nama Produk dan Provinsi sudah direpresentasikan secara numerik dan siap untuk dianalisis lebih lanjut.

...	Bulan	Nama Produk_Encoded	Provinsi_Encoded	Jumlah	Total Harga Produk	...
0	3	5	10	1	89000	
1	3	15	10	2	218000	
2	12	8	1	4	1064000	
3	12	30	5	5	610000	
4	7	2	0	2	198000	
...	...	...	...	...	...	...
204	12	23	10	1	12000	
205	3	31	10	1	5000	
206	12	20	3	4	536400	
207	12	12	3	3	399000	
208	12	0	1	4	1396000	

Gambar 4.4 Dataset Setelah Dilakukan Transformasi Numerik

c. *Data Normalisasi*

Setelah proses mengubah nilai kategorikal menjadi numerikal, semua atribut numerik yang dipakai dalam pemodelan, yaitu Nama Produk_Encoded, Provinsi_Encoded, Jumlah, dan Total Harga Produk,

harus dinormalisasi terlebih dahulu sebelum masuk ke tahap berikutnya. Alasannya sederhana, skala nilai antar atribut terlampau jauh berbeda. Atribut Total Harga Produk dinyatakan dalam satuan rupiah yang nilainya bisa mencapai jutaan bahkan miliaran, sedangkan atribut Jumlah dan hasil pengubahan atribut kategorikal menjadi numerikal berada dalam rentang yang jauh lebih kecil. Kondisi ini bermasalah karena algoritma K-Means menghitung kedekatan antardata menggunakan jarak Euclidean. Apabila satu atribut memiliki nilai yang jauh lebih besar, atribut tersebut akan otomatis mendominasi perhitungan, sementara atribut lain nyaris tidak memberi pengaruh apa pun pada hasil klusterisasi. Akibatnya, klaster yang terbentuk tidak mencerminkan keseluruhan karakteristik data. Untuk mengatasi masalah ini, digunakan metode Min-Max Normalization yang mengubah semua nilai atribut ke dalam rentang 0 hingga 1 dengan rumus:

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Keterangan:

x' = Nilai data baru setelah dinormalisasi.

x_i = Nilai data asli yang sedang diproses.

$\min(x)$ = Nilai minimum dari suatu atribut.

$\max(x)$ = Nilai maksimal dari suatu atribut.

Maka dapat diketahui dengan rumus Min-Max pada data nama produk GWMM SHOP sebagai berikut:

$$x' = \frac{5 - 0}{32 - 0} = 0,15625$$

Normalisasi dilakukan seterusnya sampai 3.854 *record* dari setiap atribut Nama Produk, Provinsi, Jumlah dan Total Harga Produk.

a. Bulan Normalisasi

Pada Tabel 4.6 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Bulan.

Tabel 4.6 Bulan Normalisasi

Bulan	Nilai Normalisasi
1	0
2	0.091
3	0.182
4	0.273
5	0.364
6	0.455
7	0.545

8	0.636
9	0.727
10	0.818
11	0.909
12	1

b. Data Nama Produk

Pada Tabel 4.7 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Nama Produk.

Tabel 4.7 Hasil Normalisasi Nama Produk

Nama Produk	Nilai Numerik
(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0
Aksesoris Gantungan Parfum	0,03125
BRODO - Rote EDT Perfume 40ml	0,0625
Crusita - Busy Day Blooming Series Extrait De Perfume	0,09375
Dynamique Course Pour Homme Men EDP 100 ML	0,125
FYNE Extrait De Perfume	0,15625
Fogg Perfume Midnight Purple 100ml Eau De Perfume	0,1875
Fogg Perfume Scent Premium Many Flowers 100ml Eau De Perfume	0,21875
Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	0,25
Fragrance World Suits Unisex Eau De Parfum 100 ML	0,28125
GARVI Perfume - Naughty Friday - EDP 50ml	0,3125
HINT IDENTITY Eau de Parfum 50ml	0,34375
HINT Noble Eau de Parfum 50ml	0,375
ICE FROST Extrait De Perfume 35ml	0,40625
Kokulu Women Best Seller - Bundling 2pcs 50ML	0,4375
Mezuca Perfume Aqua Kiss Eau de Perfume 50ml	0,46875
Mykonos Baby Love Parfum EDP 50 ML	0,5
Mykonos Milk Drips 50ml	0,53125
Mykonos Verve Extrait De Parfum 50 ML	0,5625
NARSIS Hypnotic Desire Eau de Perfume 50ml	0,59375

Omerta Pink Ice Woman Eau de Parfum 100 ML	0,625
Omerta The Winner Parfume Pria Eau de Toilette 100 ML	0,65625
Onix Fragrance Eau de Perfume Senoparty 30Ml	0,6875
Pouch Parfum Mini	0,71875
Refill Parfum 30ml	0,75
Revlon Charlie Fragrance - White EDT Parfum 100 ml	0,78125
Russ Parfume Unomento Blue Unisex 50ml	0,8125
SAFF & Co. Extrait de Parfum - CASCAVEL 30ml	0,84375
Sample Pack 3ml x5	0,875
Scentplus Dominant 35ml	0,90625
Verbena Night Legend Man Eau De Parfum 100 ML	0,9375
Vial Parfum 35ml	0,96875
Xchange Extrait De Parfume Supernova Series 50ml - Terra Nova	1

c. Data Provinsi

Pada Tabel 4.8 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Provinsi.

Tabel 4.8 Hasil Normalisasi Provinsi

Provinsi	Nilai Numerik
Bali	0
Banten	0,071428571
Bengkulu	0,142857143
DKI Jakarta	0,214285714
Gorontalo	0,285714286
Jawa Barat	0,357142857
Jawa Tengah	0,428571429
Jawa Timur	0,5
Kalimantan Timur	0,571428571
Maluku	0,642857143
Nusa Tenggara Timur (NTT)	0,714285714
Papua	0,785714286
Riau	0,857142857
Sulawesi Selatan	0,928571429
Sumatra Utara	1

d. Data Jumlah

Pada Tabel 4.9 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Jumlah.

Tabel 4.9 Hasil Normalisasi Jumlah

Bulan	Nama Produk	Jumlah
3	FYNE Extrait De Parfume	0.000000
3	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	0.166667
12	Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	0.500000
12	Verbena Night Legend Man Eau De Parfum 100 ML	0.666667
7	BRODO - Rote EDT Parfume 40ml	0.166667
.....		
12	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0.500000
7	GARVI Parfume - Naughty Friday - EDP 50ml	0.000000
7	HINT IDENTITY Eau de Parfum 50ml	1.000000
12	Russ Parfume Unomento Blue Unisex 50ml	0.071429
12	SAFF & Co. Extrait de Parfum - CASCAVEL 30ml	0.071429

e. Data Total Harga Produk

Pada Tabel 4.10 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Total Harga Produk.

Tabel 4. 10 Hasil Normalisasi Total Harga Produk

Bulan	Nama Produk	Total Harga Produk
3	FYNE Extrait De Parfume	0.034454
3	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	0.087367
12	Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	0.434372
12	Verbena Night Legend Man Eau De Parfum 100 ML	0.248154
7	BRODO - Rote EDT Parfume 40ml	0.079163
.....		
12	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0.713700
7	GARVI Parfume - Naughty Friday - EDP 50ml	0.087367

7	HINT IDENTITY Eau de Parfum 50ml	0.078425
12	Russ Parfume Unomento Blue Unisex 50ml	0.132075
12	SAFF & Co. Extrait de Parfum - CASCABEL 30ml	0.364643

Maka dapat diketahui nilai data pada dataset GWMM SHOP setelah dilakukan normalisasi pada gambar di bawah ini.

	Bulan	Nama Produk_Encoded	Provinsi_Encoded	Jumlah	Total Harga Produk
0	0.181818	0.15625	0.714286	0.000000	0.034454
1	0.181818	0.46875	0.714286	0.166667	0.087367
2	1.000000	0.25000	0.071429	0.500000	0.434372
3	1.000000	0.93750	0.357143	0.666667	0.248154
4	0.545455	0.06250	0.000000	0.166667	0.079163
...	...	...	...	...	...
204	1.000000	0.71875	0.714286	0.000000	0.002871
205	0.181818	0.96875	0.714286	0.000000	0.000000
206	1.000000	0.62500	0.214286	0.500000	0.217966
207	1.000000	0.37500	0.214286	0.333333	0.161608
208	1.000000	0.00000	0.071429	0.500000	0.570550

Gambar 4.5 Dataset Setelah Dilakukan Normalisasi

4.3 Pemodelan

4.3.1 Hasil Komparasi Model

Untuk menentukan jumlah kluster yang paling optimal, dilakukan komparasi model berdasarkan nilai *Davies-Bouldin Index* (DBI). Nilai DBI yang rendah menandakan kualitas klusterisasi yang baik karena jarak antarkluster cukup jauh dan data di dalam setiap kluster terbilang kompak. Pada penelitian ini, dua algoritma dibandingkan, yaitu K-Means dan K-Medoids, masing-masing diuji dengan berbagai variasi jumlah kluster. Seluruh proses pengujian dijalankan menggunakan RapidMiner, dan hasilnya ditampilkan pada tabel berikut sebagai dasar pemilihan model terbaik.

a. Pemodelan Dengan Algoritma *K-Means*

Tahap pemodelan pertama menggunakan algoritma K-Means dengan menguji 4 variasi jumlah kluster, yaitu 2, 3, 4, dan 5 kluster. Nilai DBI dari setiap variasi tersebut ditampilkan pada Tabel 4.11 berikut.

Tabel 4.11 Hasil DBI pada Algoritma *K-Means*

Jumlah Cluster	Nilai DBI
<i>Cluster 2</i>	1.097
<i>Cluster 3</i>	1.146
<i>Cluster 4</i>	1.002
<i>Cluster 5</i>	1.077

Berdasarkan hasil pengujian, nilai DBI terendah diperoleh pada jumlah klaster 4, yaitu sebesar 1,002. Nilai DBI yang paling kecil mencerminkan kualitas klasterisasi terbaik, sehingga model K-Means bekerja paling optimal dengan 4 klaster.

b. Pemodelan Dengan Algoritma *K-Medoids*

Pemodelan berikutnya dilakukan menggunakan algoritma K-Medoids dengan variasi jumlah klaster yang sama, yaitu 2, 3, 4, dan 5 klaster. Nilai DBI dari setiap variasi tersebut ditampilkan pada Tabel 4.12 berikut.

Tabel 4.12 Hasil DBI pada Algoritma *K-medoids*

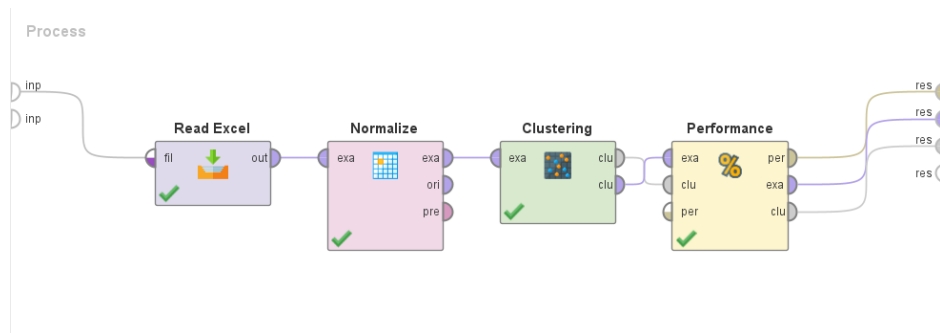
Jumlah Cluster	Nilai DBI
<i>Cluster 2</i>	1.993
<i>Cluster 3</i>	1.113
<i>Cluster 4</i>	1.079
<i>Cluster 5</i>	1.114

Berdasarkan Tabel 4.12, nilai DBI terendah pada algoritma K-Medoids diperoleh pada jumlah klaster 4, yaitu sebesar 1,079. Namun angka ini masih lebih tinggi dibandingkan seluruh nilai DBI yang dihasilkan oleh algoritma K-Means.

Setelah membandingkan kedua algoritma, K-Means dengan 4 klaster menghasilkan nilai DBI paling rendah, yaitu 1,002. Oleh karena itu, algoritma K-Means dipilih sebagai metode klasterisasi terbaik dalam penelitian ini dengan jumlah klaster optimal sebanyak 4 klaster.

4.3.2 Penyajian Model Terbaik

Berdasarkan hasil evaluasi, algoritma K-Means dengan 4 klaster terpilih sebagai model terbaik dibandingkan K-Medoids. Pemilihan ini didasarkan pada nilai Davies-Bouldin Index (DBI) terendah sebesar 1,002, yang menunjukkan bahwa pemisahan antarklaster cukup baik dan data di dalam setiap klaster cukup kompak. Proses klasterisasi sendiri dilakukan pada data yang sudah melewati tahap normalisasi menggunakan metode Min-Max Normalization. Atribut yang digunakan dalam pemodelan ini ada lima, yaitu Bulan, Jumlah, Total Harga Produk, Nama Produk yang sudah dikonversi ke bentuk numerik, dan Provinsi yang sudah dikodekan ke bentuk numerik. Alur proses pemodelan K-Means *Clustering* pada RapidMiner ditampilkan pada Gambar 4.6 berikut.



Gambar 4.6 Penyajian Model dengan RapidMiner

Setelah proses klasterisasi dilakukan, diperoleh hasil distribusi data ke dalam empat *cluster*, dengan rincian Gambar 4.7 sebagai berikut:

Cluster Model

```

Cluster 0: 1203 items
Cluster 1: 1449 items
Cluster 2: 652 items
Cluster 3: 550 items
Total number of items: 3854
  
```

Gambar 4.7 Hasil Jumlah Cluster

Untuk memahami karakteristik tiap klaster yang terbentuk, dilakukan analisis deskriptif berdasarkan nilai minimum, rata-rata, dan maksimum dari setiap atribut yang digunakan dalam proses klasterisasi. Hasilnya disajikan pada Tabel 4.13 berikut.

Tabel 4.13 Data Jumlah Minimal, Maksimal, dan Rata-rata Pada Setiap Atribut Data

		Bulan	Nama Produk_Encoded	Provinsi_Encoded	Jumlah	Total Harga Produk
Cluster 0	Min	1	0	7	1	10.000
	Rata-Rata	6,089	9,389	11,522	1,513	162.596,010
	Max	12	25	14	4	1.396.000
Cluster 1	Min	2	0	0	1	12.000
	Rata-Rata	11,619	18,144	4,346	4,511	725.445,687
	Max	12	32	8	7	2.443.000
	Min	1	0	0	1	10.000

Cluster 2	Rata-Rata	5,678	11,553	1,045	1,789	234.784,909
	Max	12	29	7	6	1.745.000
Cluster 3	Min	1	14	2	1	5.000
	Rata-Rata	3,997	27,066	8,862	1,417	93.116,564
	Max	12	32	14	7	894.000

Berdasarkan hasil pemodelan K-Means dengan $k=4$, terbentuk empat klaster dari total 3.854 data transaksi. Setiap klaster dibentuk berdasarkan atribut Bulan, Nama Produk_Encoded, Provinsi_Encoded, Jumlah, dan Total Harga Produk yang sudah melalui proses transformasi dan normalisasi. Analisis terhadap keempat klaster ini dilakukan untuk menggambarkan pola penjualan GWMM SHOP, sebagaimana diuraikan berikut ini.

a. Cluster 0

Klaster 0 berisi 1.203 data transaksi dengan rata-rata bulan sebesar 6,089, yang menunjukkan transaksi pada klaster ini terkonsentrasi di bulan Juni hingga September. Rata-rata jumlah produk per transaksi tercatat sebesar 1,513 unit dengan rata-rata total harga transaksi sebesar Rp162.596. Produk yang paling banyak muncul pada klaster ini adalah Mykonos Baby Love Parfum EDP 50 ML dan Fyne Extrait De Parfume, yaitu parfum segmen mid-range yang menjadi pilihan utama konsumen di luar musim belanja besar. Provinsi dengan transaksi terbanyak pada klaster ini adalah Sumatera Utara, yang mengindikasikan bahwa pasar di luar Pulau Jawa tetap aktif sepanjang pertengahan tahun.

b. Cluster 1

Klaster 1 memiliki jumlah data terbanyak, yaitu 1.449 transaksi, dengan rata-rata total harga produk tertinggi sebesar Rp725.445 per transaksi. Rata-rata bulan sebesar 11,619 menunjukkan klaster ini didominasi transaksi pada bulan Oktober hingga Desember, yang bertepatan dengan momen Harbolnas, Natal, dan Tahun Baru. Rata-rata jumlah produk per transaksi juga tercatat paling tinggi, yaitu 4,511 unit, yang menandakan konsumen cenderung membeli lebih banyak produk sekaligus pada periode tersebut. Provinsi dengan transaksi paling aktif adalah DKI Jakarta, Jawa Barat, dan Jawa Timur, selaras dengan konsentrasi populasi pengguna *e-commerce* terbesar di Indonesia.

c. Cluster 2

Klaster 2 berisi 652 data transaksi dengan rata-rata bulan sebesar 5,678, yang menunjukkan transaksi pada klaster ini terkonsentrasi di bulan Mei hingga Juli. Klaster ini mencatat nilai rata-rata total harga produk sedang, yaitu Rp234.784, dengan rata-rata jumlah produk per transaksi paling kecil sebesar 1,789 unit. Produk

yang paling banyak muncul adalah parfum lokal berharga terjangkau seperti Xchange Extrait De Parfume Supernova Series 50ml - Terra Nova dan Sample Pack 3ml x5. Yang menarik, provinsi dengan transaksi terbanyak justru Papua, NTT, dan Kalimantan Timur, yang mengindikasikan adanya segmen pasar aktif di wilayah Indonesia timur meskipun volume transaksinya masih relatif Rendah.

d. Cluster 3

Klaster 3 berisi 550 data transaksi dengan rata-rata bulan sebesar 3,997, yang menunjukkan transaksi pada klaster ini terkonsentrasi di bulan Januari hingga Mei. Klaster ini mencatat nilai rata-rata total harga produk rendah, yaitu Rp93.116, dengan rata-rata jumlah produk per transaksi paling kecil sebesar 1,417 unit. Produk yang paling banyak muncul adalah parfum lokal berharga terjangkau seperti HINT IDENTITY dan ICE FROST Extrait De Parfume 35ml. Yang menarik, provinsi dengan transaksi terbanyak justru Bali, Gorontalo, dan Bengkulu, yang mengindikasikan adanya segmen pasar aktif di wilayah Indonesia timur meskipun volume transaksinya masih sangat rendah.

Tabel 4.14 Karakteristik Setiap Cluster

<i>Cluster</i>	Jumlah	Nama Produk	Provinsi	Total Harga Produk	Bulan
0	Tinggi	Mykonos Baby Love	Sumatra Utara	Rp162.596	7-9
1	Sangat Tinggi	Hint Noble	DKI Jakarta	Rp725.445	10-12
2	Sedang	Xchange Extrait De Parfume Supernova Series 50ml	Papua	Rp234.784	5-7
3	Rendah	HINT IDENTITY Eau De Parfume 50ml	Bali	Rp93.116	1-5

Secara keseluruhan, keempat klaster yang terbentuk menggambarkan empat pola penjualan yang berbeda berdasarkan dimensi waktu, nama produk, jumlah, total harga produk, dan wilayah. Klaster 1 merepresentasikan periode penjualan puncak yang perlu dijadikan prioritas dalam persiapan stok dan strategi promosi. Klaster 0 dan klaster 2 menggambarkan periode penjualan stabil di pertengahan tahun yang dapat dimanfaatkan untuk memperluas kampanye

pemasaran ke pasar luar Jawa. Sementara itu, Klaster 3 mencerminkan periode paling sepi yang justru bisa dijadikan peluang untuk menjalankan program promosi awal tahun guna mendorong peningkatan transaksi.

4.3.3 Perhitungan Algoritma Terpilih

Pada tahapan ini dilakukan proses perhitungan manual menggunakan algoritma yang telah terpilih, yaitu K-Means *Clustering*. Pengolahan data dilakukan setelah data melalui proses transformasi dan normalisasi, sehingga data telah siap untuk dianalisis menggunakan K-Means. Perhitungan dilakukan dengan menggunakan sampel dataset sebanyak 5 data untuk mempermudah penjabaran proses secara manual. Data yang digunakan mencakup atribut Nama Produk (normalisasi), Bulan (normalisasi), Jumlah (normalisasi), Total Harga Produk (normalisasi), dan Provinsi (normalisasi), yang akan diproses lebih lanjut dalam tahap klasterisasi.

Tabel 4.15 Dataset pada Algoritma Terpilih

Nama Produk	Bulan	Nama Produk Norm	Provinsi_Norm	Jumlah_Norm	Total Harga Norm
Mykonos Baby Love Parfum EDP 50ML	0.455	0,167	1,000	0,000	0,056
HINT Noble Eau De Parfume 50ml	1,000	0,375	0,214	0,667	0,271
Xchange Extrait De Parfume Supernova Series 50ml	0.182	1,000	0,786	0,000	0,051
HINT IDENTITY Eau De Parfume 50ml	0,000	0,344	0,000	0,333	0,132
FYNE Extrait De Parfume	0,182	0,156	0,714	0,000	0,034

Setelah data dinormalisasi, proses klasterisasi menggunakan algoritma K-Means dapat dimulai. Tahapan awal yang dilakukan adalah sebagai berikut.

1. Jumlah klaster yang digunakan dalam penelitian ini adalah 4 klaster ($k=4$), sesuai dengan hasil evaluasi DBI yang telah dilakukan pada tahap sebelumnya.
2. Centroid awal setiap klaster ditentukan dengan cara mengambil data secara acak dari sampel yang telah disiapkan sebelumnya. Data centroid tersebut adalah sebagai berikut.

Tabel 4. 16 Titik Pusat Awal Tiap Cluster

Centroid Awal	Bulan	Nama Produk_Norm	Provinsi_Norm	Jumlah_Norm	Total Harga_Norm
C0	0.455	0,167	1,000	0,000	0,056
C1	1,000	0,375	0,214	0,667	0,271
C2	0.182	1,000	0,786	0,000	0,051
C3	0,000	0,344	0,000	0,333	0,132

3. Setelah centroid awal ditentukan, langkah berikutnya adalah menghitung jarak setiap data terhadap tiap centroid menggunakan rumus Euclidean Distance. Perhitungan ini bertujuan untuk mengukur seberapa dekat setiap data terhadap pusat klaster yang telah ditetapkan.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Berikut jarak data 1 dengan setiap centroid :

$$d(1, C0)$$

$$= \sqrt{(0,455 - 0,455)^2 + (0,167 - 0,167)^2 + (1,000 - 1,000)^2 + (0,000 - 0,000)^2 + (0,056 - 0,056)^2}$$

$$= 0,000$$

$$d(1, C1)$$

$$= \sqrt{(0,455 - 1,000)^2 + (0,167 - 0,375)^2 + (1,000 - 0,214)^2 + (0,000 - 0,667)^2 + (0,056 - 0,271)^2}$$

$$= 1,204$$

$$d(1, C2)$$

$$= \sqrt{(0,455 - 0,182)^2 + (0,167 - 1,000)^2 + (1,000 - 0,786)^2 + (0,000 - 0,000)^2 + (0,056 - 0,051)^2}$$

$$= 0,902$$

$$\begin{aligned}
& d(1, C3) \\
&= \sqrt{(0,455 - 0,000)^2 + (0,167 - 0,344)^2 + (1,000 - 0,000)^2 + (0,000 - 0,333)^2 + (0,056 - 0,132)^2} \\
&= 1,164
\end{aligned}$$

Berikut jarak data 2 dengan setiap centroid :

$$\begin{aligned}
& d(2, C0) \\
&= \sqrt{(1,000 - 0,455)^2 + (0,375 - 0,167)^2 + (0,214 - 1,000)^2 + (0,667 - 0,000)^2 + (0,271 - 0,056)^2} \\
&= 1,204
\end{aligned}$$

$$\begin{aligned}
& d(2, C1) \\
&= \sqrt{(1,000 - 1,000)^2 + (0,375 - 0,375)^2 + (0,214 - 0,214)^2 + (0,667 - 0,667)^2 + (0,271 - 0,271)^2} \\
&= 0,000
\end{aligned}$$

$$\begin{aligned}
& d(2, C2) \\
&= \sqrt{(1,000 - 0,182)^2 + (0,375 - 1,000)^2 + (0,214 - 0,786)^2 + (0,667 - 0,000)^2 + (0,271 - 0,051)^2} \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
& d(2, C3) \\
&= \sqrt{(1,000 - 0,000)^2 + (0,375 - 0,344)^2 + (0,214 - 0,000)^2 + (0,667 - 0,333)^2 + (0,271 - 0,132)^2} \\
&= 1,085
\end{aligned}$$

Berikut jarak data 3 dengan setiap centroid :

$$\begin{aligned}
& d(3, C0) \\
&= \sqrt{(0,182 - 0,455)^2 + (1,000 - 0,167)^2 + (0,786 - 1,000)^2 + (0,000 - 0,000)^2 + (0,051 - 0,056)^2} \\
&= 0,902
\end{aligned}$$

$$\begin{aligned}
& d(3, C1) \\
&= \sqrt{(0,182 - 1,000)^2 + (1,000 - 0,375)^2 + (0,786 - 0,214)^2 + (0,000 - 0,667)^2 + (0,051 - 0,271)^2} \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
& d(3, C2) \\
&= \sqrt{(0,182 - 0,182)^2 + (1,000 - 1,000)^2 + (0,786 - 0,786)^2 + (0,000 - 0,000)^2 + (0,051 - 0,051)^2} \\
&= 0,000
\end{aligned}$$

$$\begin{aligned}
& d(3, C3) \\
&= \sqrt{(0,182 - 0,000)^2 + (1,000 - 0,344)^2 + (0,786 - 0,000)^2 + (0,000 - 0,333)^2 + (0,051 - 0,132)^2} \\
&= 1,095
\end{aligned}$$

Berikut jarak data 4 dengan setiap centroid :

$$\begin{aligned}
& d(4, C0) \\
&= \sqrt{(0,000 - 0,455)^2 + (0,344 - 0,167)^2 + (0,000 - 1,000)^2 + (0,333 - 0,000)^2 + (0,132 - 0,056)^2} \\
&= 1,164
\end{aligned}$$

$$\begin{aligned}
& d(4, C1) \\
&= \sqrt{(0,000 - 1,000)^2 + (0,344 - 0,375)^2 + (0,000 - 0,214)^2 + (0,333 - 0,667)^2 + (0,132 - 0,271)^2} \\
&= 1,085
\end{aligned}$$

$$\begin{aligned}
& d(4, C2) \\
&= \sqrt{(0,000 - 0,182)^2 + (0,344 - 1,000)^2 + (0,000 - 0,786)^2 + (0,333 - 0,000)^2 + (0,132 - 0,051)^2} \\
&= 1,095
\end{aligned}$$

$$\begin{aligned}
& d(4, C3) \\
&= \sqrt{(0,000 - 0,000)^2 + (0,344 - 0,344)^2 + (0,000 - 0,000)^2 + (0,333 - 0,333)^2 + (0,132 - 0,132)^2} \\
&= 0,000
\end{aligned}$$

Berikut jarak data 5 dengan setiap centroid :

$$\begin{aligned}
& d(5, C0) \\
&= \sqrt{(0,182 - 0,455)^2 + (0,156 - 0,167)^2 + (0,714 - 1,000)^2 + (0,000 - 0,000)^2 + (0,034 - 0,056)^2} \\
&= 0,396
\end{aligned}$$

$$\begin{aligned}
& d(5, C1) \\
&= \sqrt{(0,182 - 1,000)^2 + (0,156 - 0,375)^2 + (0,714 - 0,214)^2 + (0,000 - 0,667)^2 + (0,034 - 0,271)^2} \\
&= 1,211
\end{aligned}$$

$$\begin{aligned}
& d(5, C2) \\
&= \sqrt{(0,182 - 0,182)^2 + (0,156 - 1,000)^2 + (0,714 - 0,786)^2 + (0,000 - 0,000)^2 + (0,034 - 0,051)^2} \\
&= 0,847
\end{aligned}$$

$$\begin{aligned}
& d(5, C3) \\
&= \sqrt{(0,182 - 0,000)^2 + (0,156 - 0,344)^2 + (0,714 - 0,000)^2 + (0,000 - 0,333)^2 + (0,034 - 0,132)^2} \\
&= 0,836
\end{aligned}$$

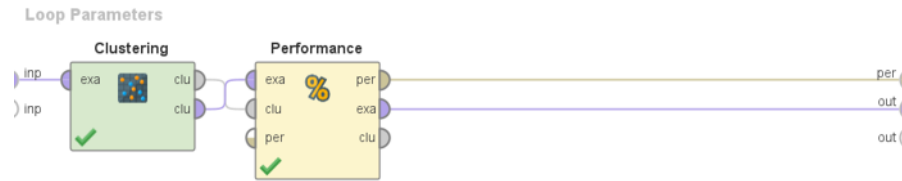
Berdasarkan hasil perhitungan jarak setiap data terhadap seluruh centroid, setiap data dikelompokkan ke dalam klaster dengan nilai jarak terkecil. Hasil pengelompokan data ke dalam klaster yang sesuai ditampilkan pada Tabel 4.16 berikut.

Tabel 4.17 Hasil Perhitungan Setiap Data ke Setiap Pusat *Cluster*

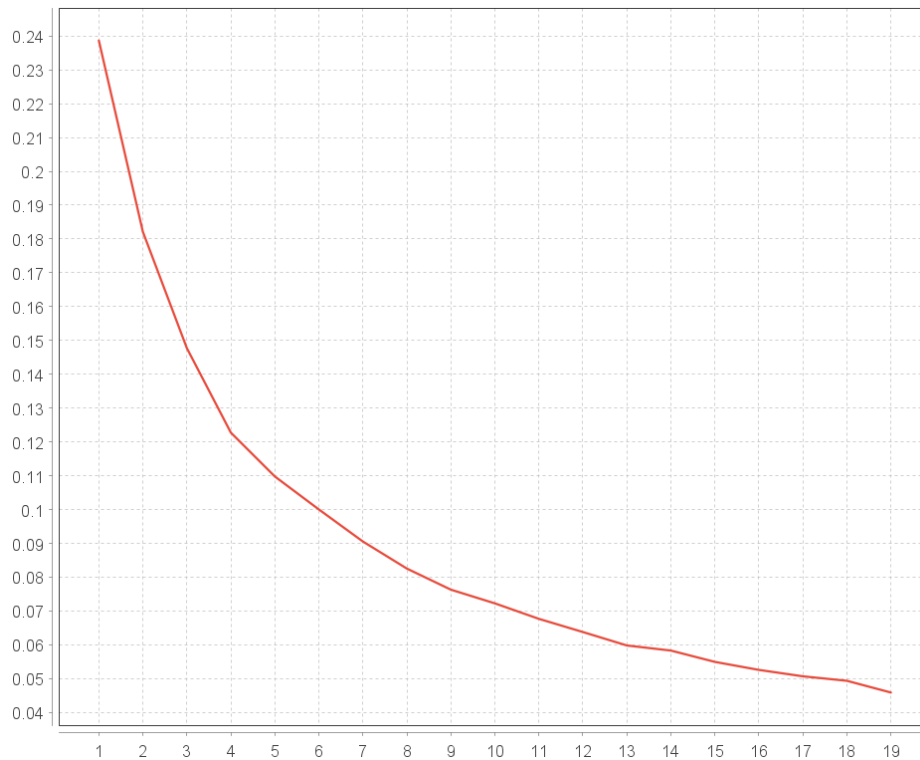
Nama Produk	Jarak Ke C0	Jarak Ke C1	Jarak Ke C2	Jarak Ke C3	Cluster
Mykonos Baby Love Parfum EDP 50 ML	0,000	1,204	0,902	1,164	C0
HINT Noble Eau de Parfum 50ml	1,204	0,000	1,371	1,085	C1
Xchange Extrait De Parfume Supernova Series 50ml	0,902	1,371	0,000	1,095	C2
HINT IDENTITY Eau De Parfume 50ml	1,164	1,085	1,095	0,000	C3
FYNE Extrait De Parfume	0,396	1,211	0,847	0,836	C0

4.4 Pengujian

Untuk menentukan jumlah kluster yang paling sesuai, penelitian ini menggunakan dua metode evaluasi, yaitu *Elbow Method* dan *Davies-Bouldin Index* (DBI), yang keduanya dijalankan melalui RapidMiner. Tahapan pengujian Elbow Method dapat dilihat pada Gambar 4.8, sedangkan hasil visualisasinya ditampilkan pada Gambar 4.9.



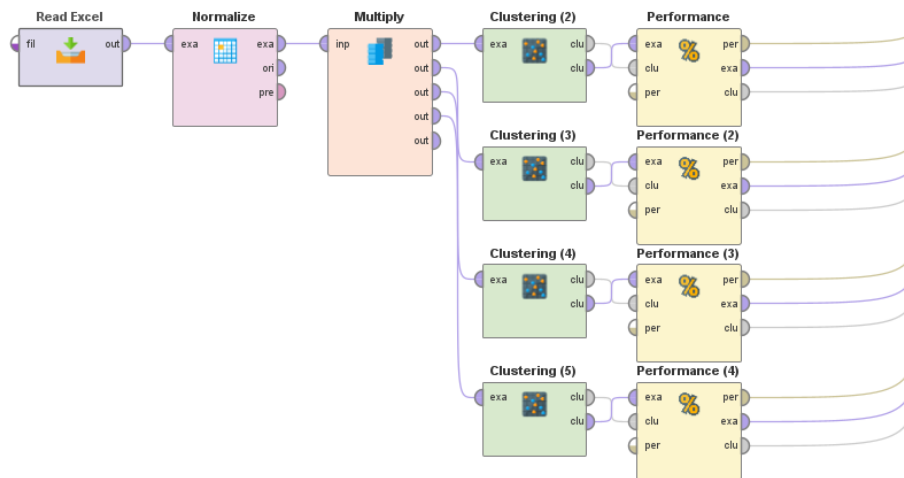
Gambar 4.8 Pengujian *Elbow Method* Menggunakan RapidMiner



Gambar 4.9 Grafik *Elbow Method*

Gambar 4.9 memperlihatkan hasil *Elbow Method* yang menggambarkan hubungan antara jumlah kluster (k) dan nilai *average within centroid distance*. Grafik menunjukkan penurunan nilai yang tajam dari kluster 2 hingga kluster 4. Setelah k=4, penurunan nilai mulai melambat dan kurva bergerak lebih landai, yang berarti penambahan kluster berikutnya tidak lagi memberikan pengurangan jarak yang berarti terhadap pusat kluster.

Dari pola tersebut, titik "siku" atau *elbow* terlihat paling jelas pada $k=4$, sehingga jumlah klaster optimal berdasarkan metode ini adalah 4 klaster.



Gambar 4.10 Pengujian Dengan *Davies Bouldien Index*

Tabel 4.18 Tabel Hasil Pengujian *Davies Bouldien Index*

Jumlah Cluster	Nilai DBI
Cluster 2	1.097
Cluster 3	1.146
Cluster 4	1.002
Cluster 5	1.077

Pengujian lanjutan dilakukan menggunakan metode *Davies-Bouldin Index* (DBI) untuk memperkuat hasil yang diperoleh. Hasil evaluasi menunjukkan nilai DBI terendah sebesar 1,002 juga ditemukan pada jumlah klaster 4, yang mencerminkan bahwa pemisahan antarklaster pada model ini paling optimal dibanding variasi jumlah klaster lainnya. Dengan demikian, baik dari sisi visualisasi Elbow maupun evaluasi kuantitatif DBI, jumlah 4 klaster ditetapkan sebagai jumlah klaster optimal dalam penelitian ini.

BAB V PENUTUP

5.1 Kesimpulan

- a. Berdasarkan hasil penelitian menggunakan RapidMiner dengan algoritma K-Means *Clustering*, model terbaik diperoleh pada jumlah 4 klaster dengan nilai *Davies-Bouldin Index* (DBI) sebesar 1,002. Nilai ini menunjukkan bahwa model yang terbentuk sudah cukup baik karena menghasilkan kelompok data yang jelas dan terpisah, sehingga dapat digunakan untuk memahami pola penjualan parfum pada GWMM SHOP berdasarkan dimensi waktu, produk, dan wilayah.
- b. Hasil pengelompokan membagi data transaksi menjadi empat klaster berdasarkan dimensi bulan, jumlah produk per transaksi, total harga transaksi, dan provinsi tujuan pengiriman. Klaster 0 berisi 1.203 transaksi dengan rata-rata bulan sebesar 6,089 yang terkonsentrasi pada bulan Juni hingga September. Rata-rata jumlah produk pada klaster ini sebesar 1,513 unit dengan rata-rata total harga transaksi sebesar Rp162.596, didominasi oleh produk Mykonos Baby Love Parfum EDP 50ML dan Fyne Extrait De Parfume dengan wilayah tujuan terbanyak Sumatera Utara. Klaster 1 merupakan klaster dengan jumlah transaksi terbanyak yaitu 1.449 transaksi dan rata-rata total harga tertinggi sebesar Rp725.445. Rata-rata bulan pada klaster ini sebesar 11,619 yang bertepatan dengan momen Harbolnas, Natal, dan Tahun Baru, dengan rata-rata jumlah produk per transaksi juga tertinggi yaitu 4,511 unit, serta wilayah tujuan terbanyak berada di DKI Jakarta, Jawa Barat, dan Jawa Timur. Klaster 2 berisi 652 transaksi dengan rata-rata bulan sebesar 5,678 yang terkonsentrasi pada bulan Mei hingga Juli. Rata-rata total harga transaksi pada klaster ini tergolong sedang yaitu Rp234.784 dengan rata-rata jumlah produk per transaksi sebesar 1,789 unit, didominasi produk Xchange Extrait De Parfume Supernova Series 50ml Terra Nova dan Sample Pack 3ml x5, dengan wilayah tujuan terbanyak berada di Papua, Nusa Tenggara Timur, dan Kalimantan Timur. Klaster 3 berisi 550 transaksi dengan rata-rata bulan sebesar 3,997 yang terkonsentrasi pada bulan Januari hingga Mei. Klaster ini memiliki rata-rata total harga transaksi paling rendah yaitu Rp93.116 dengan rata-rata jumlah produk per transaksi sebesar 1,417 unit, didominasi produk HINT *Identity* dan Ice Frost Extrait De Parfume 35ml, dengan wilayah tujuan terbanyak berada di Bali, Gorontalo, dan Bengkulu.
- c. Dari keempat klaster yang terbentuk, diperoleh beberapa rekomendasi strategi bisnis bagi GWMM SHOP. Klaster 1 menunjukkan penjualan tertinggi dengan nilai transaksi yang jauh melampaui klaster lain, sehingga GWMM SHOP perlu menyiapkan stok produk best seller dalam jumlah lebih banyak menjelang bulan Oktober hingga Desember serta memperkuat promosi pada wilayah Jawa, khususnya DKI Jakarta, Jawa Barat, dan Jawa Timur, mengingat periode ini bertepatan dengan Harbolnas, Natal, dan Tahun Baru yang menjadi puncak belanja konsumen. Klaster 0 mencerminkan penjualan yang stabil pada

pertengahan tahun dengan konsumen yang cenderung memilih produk mid-range, sehingga momentum ini dapat dimanfaatkan untuk mempertahankan loyalitas pelanggan di wilayah luar Jawa seperti Sumatera Utara melalui program promosi berkala. Klaster 2 dan Klaster 3 sama-sama mencerminkan segmen pasar dengan nilai transaksi yang lebih kecil dan didominasi produk berharga terjangkau, namun keduanya menunjukkan bahwa wilayah Indonesia timur dan tengah seperti Papua, Nusa Tenggara Timur, Kalimantan Timur, Bali, Gorontalo, dan Bengkulu tetap memiliki aktivitas transaksi yang perlu diperhatikan. GWMM SHOP dapat menjadikan wilayah-wilayah tersebut sebagai target ekspansi pasar melalui strategi diskon atau bundling produk terjangkau, sehingga potensi pasar di luar Pulau Jawa dapat dioptimalkan secara bertahap.

5.2 Saran

- a. Penelitian ini masih memiliki ruang untuk dikembangkan lebih lanjut, baik dengan menambahkan atribut lain seperti metode pembayaran atau jenis produk secara lebih rinci, menggunakan algoritma *clustering* yang berbeda, maupun menerapkan pendekatan ini pada UMKM dengan kategori produk lain guna memperoleh hasil analisis yang lebih mendalam dan sesuai konteks bisnis setiap pelaku usaha.
- b. Hasil klasterisasi ini dapat dijadikan dasar pertimbangan bagi pemilik GWMM SHOP dalam menentukan waktu, produk, dan wilayah yang perlu mendapat perhatian lebih, terutama dalam perencanaan stok, strategi promosi, dan target pemasaran ke depan.
- c. Penelitian berikutnya dapat mencoba membandingkan hasil K-Means dengan algoritma *clustering* lain selain K-Medoids, seperti DBSCAN atau Hierarchical *Clustering*, untuk melihat apakah ada pendekatan yang mampu menghasilkan nilai DBI lebih rendah pada data UMKM berbasis Shopee.
- d. penelitian berikutnya juga dapat diarahkan pada pembangunan model prediktif untuk mengestimasi lonjakan permintaan pada periode Oktober hingga Desember tahun berjalan. Pendekatan ini akan memberi nilai tambah praktis, karena hasil penelitian tidak hanya bersifat deskriptif terhadap pola yang telah terjadi, tetapi juga dapat difungsikan sebagai alat bantu perencanaan operasional bagi GWMM SHOP.

DAFTAR PUSTAKA

- Ananda, P. S. ... Sembiring, I. (2023). KMeans Clustering Menggunakan RapidMiner dalam Segmentasi Pelanggan dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal. *Jurnal BATIRSI*, 6(2).
- Dasopang, N. (2024). E-Commerce Bisnis Dan Internet. *JIEMAS: Jurnal Ilmiah Ekonomi, Manajemen Dan Syariah*, 3(1). <https://doi.org/10.55883/jiemas.v3i1>
- Dermawan, H. D. ... Wijaya, Y. A. (2024). CESS Analisis Data Penjualan Menggunakan Algoritma K-Means Clustering Pada Toko Daun Indah Di Shopee Analysis of Sales Data Using K-Means Clustering Algorithm at Daun Indah Shop in Shopee. *CESS (Journal of Computing Engineering, System and Science)*, 9(1), 175–191. <https://jurnal.unimed.ac.id/2012/index.php/cess/en/article/view/55093/pdf>
- Dona, D. ... Rahayu, D. (2025). Penerapan Data Mining Terhadap Penjualan Produk Kosmetik Menggunakan Algoritma K-Means. *Riau Jurnal Teknik Informatika*, 4(3), 502–513. <https://doi.org/10.30606/rjti.v4i3.4275>
- Eriskiannisa Febrianty Luchia Awalina, W. I. R. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions. *Jurnal Teknologi Dan Informasi*, 13(September), 122–137. <https://doi.org/10.34010/jati.v13i2>
- Fatkhudin, A. ... Zade, N. A. P. (2023). Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika). *Jurnal Minfo Polgan*, 12(2), 777–783. <https://doi.org/10.33395/jmp.v12i2.12494>
- Fitri, E. M. ... Wantoro, A. (2023). Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada PT XYZ. *Jurnal Komputasi*, 11(2), 157–168.
- Gori, T. ... Al Fatta, H. (2024). Preprocessing Data Dan Klasifikasi Untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Ilmu, G. J. M. (2024). Gudang Jurnal Multidisiplin Ilmu Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas. *Gudang Jurnal Multidisiplin Ilmu*, 2(November), 86–94. <http://gudangjurnal.com/index.php/gjmi/article/view/1040/933>
- Jollyta, D. ... Gustientiedina. (2023). Pengoptimalan Pengukuran Bregman Divergences Menggunakan Davies Bouldin Index. *JOISIE Journal Of Information System And Informatics Engineering*, 7(1), 174–180.
- Kamalludin, A., & Budianto, et al. (2023). Penerapan Algoritme Klasterisasi K-Means Kinerja Produk Dengan Analisis Recency Frequency Monetary Pada Cafe Xyz. *Prosiding Seminar ...*, 2(April), 258–266. <http://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/631%0Ahttps://senafiti.budiluhur.ac.id/index.php/senafiti/article/download/631/271>

- Khalif, A. ... Sari, B. N. (2024). Klasterisasi Tingkat Kemiskinan di Indonesia Menggunakan Algoritma K-Means. *Generation Journal*, 8(1), 54–62. <https://ojs.unpkediri.ac.id/index.php/gj/article/view/21470/3747>
- Pratiwi, D. S. ... Siregar, A. C. (2023). Peningkatan Keterampilan Pengolahan Data dengan Microsoft Excel terhadap Peserta Didik Madrasah Aliyah Al-Uswah Samarinda. *Jurnal Pengabdian Kepada Masyarakat Nusantara (JPkMN)*, 4(4), 4542–4549. <https://doi.org/10.55338/jpkmn.v4i4.2166>
- Ramadhan, M. I. ... Syafria, F. (2026). Analisis Clustering Menggunakan Metode K-Means untuk Mengidentifikasi Pola Kepuasan Alumni. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 6(1), 31–41. <https://doi.org/10.57152/malcom.v6i1.2401>
- Riadi, R., & Mesran. (2023). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume. *Journal of Informatics, Electrical and Electronics Engineering*, 2(4), 138–145. <https://doi.org/10.47065/jieeee.v2i4.1181>
- Vania, P., & Sari, B. N. (2023). Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means. *Jurnal Ilmiah Wahana Pendidikan*, 9(21), 547–558. <https://doi.org/10.5281/zenodo.10081332>
- Wahyudi, S. I., & Wibowo, A. (2022). Implementasi Metode K-Means Clustering Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI), September*, 470–478. <https://senafti.budiluhur.ac.id/index.php>
- Wilyani, F. ... Aslimar, F. (2024). Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory. *Jurnal Pengabdian Pada Masyarakat Indonesia (JPPMI)*, 3(1), 8–14. <https://doi.org/10.55606/jppmi.v3i1.1087>

**Penerapan Algoritma K-Means Clustering untuk Analisis Pola
Penjualan Berdasarkan Waktu, Produk, dan Wilayah pada
GWMM SHOP**

**Diajukan untuk memenuhi salah satu persyaratan memperoleh
gelar Sarjana Komputer (S. Kom)**

TUGAS AKHIR



**Oleh:
Febri Ardana
2212500199**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2026**

**Penerapan Algoritma K-Means Clustering untuk Analisis Pola
Penjualan Berdasarkan Waktu, Produk, dan Wilayah pada
GWMM SHOP**

**Diajukan untuk memenuhi salah satu persyaratan memperoleh
gelar Sarjana Komputer (S. Kom)**

TUGAS AKHIR



**Oleh:
Febri Ardana
2212500199**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2026**



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

LEMBAR PENGESAHAN

Nama	: Febri Ardana
Nomor Induk Mahasiswa	: 2212500199
Program Studi	: Sistem Informasi
Peminatan	: Data Science
Jenjang Studi	: Strata 1
Judul	: PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS POLA PENJUALAN BERDASARKAN WAKTU, PRODUK, DAN WILAYAH PADA GWMM SHOP



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 13 Juli 2026

Tim Penguji:

Ketua	: Lauw Li Hin, S.Kom., M.Kom.
Anggota	: Agnes Aryasanti, S.Kom., M.Kom.
Pembimbing	: Muhammad Ainur Rony, S.Kom., M.T.I.
Ketua Program Studi	: Bima Cahya Putra, S.Kom., M.Kom.

ABSTRAK
PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK ANALISIS
POLA PENJUALAN BERDASARKAN WAKTU, PRODUK, DAN
WILAYAH PADA GWMM SHOP

Oleh : Febri Ardana (NIM : 2212500199)

Pesatnya pertumbuhan *e-commerce* di Indonesia mendorong semakin banyak pelaku UMKM memanfaatkan platform *marketplace* seperti Shopee sebagai saluran penjualan utama. Namun sebagian besar UMKM masih mengelola bisnisnya secara konvensional tanpa memanfaatkan data transaksi yang sebenarnya sudah mereka miliki. GWMM SHOP, sebuah toko parfum yang berjualan di Shopee, mengalami permasalahan serupa karena belum memiliki sistem analisis yang mampu mengidentifikasi pola penjualan secara spesifik berdasarkan waktu, jenis produk, dan wilayah pembeli. Akibatnya, keputusan terkait manajemen stok, waktu promosi, dan target pemasaran masih dilakukan secara spekulatif dan belum berbasis data. Penelitian ini bertujuan menerapkan metode *data mining* dengan algoritma K-Means *Clustering* untuk mengelompokkan data transaksi penjualan GWMM SHOP berdasarkan tiga dimensi utama, yaitu waktu (bulan), produk, dan wilayah (provinsi). Data yang digunakan merupakan data *primer* berupa riwayat transaksi penjualan yang diekspor langsung dari *Seller Center Shopee* milik GWMM SHOP pada periode Januari 2024 hingga Desember 2025, dengan jumlah awal 4.163 baris data yang menyusut menjadi 3.854 baris setelah melalui tahap data cleaning. Proses analisis mengikuti metodologi CRISP-DM, mulai dari pemahaman data, persiapan data seperti label *encoding* dan normalisasi *Min-Max*, pemodelan, hingga evaluasi menggunakan *Elbow Method* dan *Davies-Bouldin Index* (DBI). Hasil komparasi antara algoritma K-Means dan K-Medoids menunjukkan bahwa K-Means dengan 4 klaster memberikan nilai DBI paling rendah, yaitu sebesar 1,002, sehingga dipilih sebagai model terbaik dalam penelitian ini. Keempat klaster yang terbentuk memiliki karakteristik yang berbeda. Klaster 0 berisi 1.203 transaksi yang terkonsentrasi pada bulan Juni hingga September dengan rata-rata total harga Rp162.596 dan wilayah dominan Sumatera Utara. Klaster 1 merupakan klaster dengan jumlah transaksi terbanyak, yaitu 1.449 transaksi, dengan rata-rata total harga tertinggi sebesar Rp725.445, terjadi pada bulan Oktober hingga Desember bertepatan dengan momen Harbolnas, Natal, dan Tahun Baru, dengan wilayah dominan DKI Jakarta, Jawa Barat, dan Jawa Timur. Klaster 2 berisi 652 transaksi pada bulan Mei hingga Juli dengan rata-rata total harga Rp234.784 dan wilayah dominan Papua, Nusa Tenggara Timur, serta Kalimantan Timur. Klaster 3 berisi 550 transaksi dengan nilai transaksi paling rendah, yaitu rata-rata Rp93.116, terkonsentrasi pada bulan Januari hingga Mei dengan wilayah dominan Bali, Gorontalo, dan Bengkulu. Hasil penelitian ini menghasilkan rekomendasi bagi GWMM SHOP dalam pengambilan keputusan operasional berbasis pola penjualan, mencakup strategi manajemen stok menjelang akhir tahun, penentuan waktu promosi, dan perluasan target pemasaran ke wilayah potensial di luar Pulau Jawa.

Kata Kunci: Data Mining, K-Means Clustering, Pola Penjualan, UMKM, E-Commerce.

PERNYATAAN TIDAK PLAGIAT

SURAT PERNYATAAN TIDAK PLAGIAT DAN PERSETUJUAN PUBLIKASI

Saya yang bertanda tangan dibawah ini :

Nama : FEBRI ARDANA
NIM : 2212500199
Program Studi : SISTEM INFORMASI
Bidang Peminatan : DATA SCIENCE
Jenjang Studi : STRATA 1
Fakultas : TEKNOLOGI INFORMASI

Menyatakan bahwa TUGAS AKHIR yang berjudul:

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS POLA

PENJUALAN BERDASARKAN WAKTU, PRODUK, DAN WILAYAH PADA GWMM SHOP

Merupakan:

1. Karya tulis saya sebagai laporan tugas akhir yang asli dan belum pernah diajukan untuk mendapatkan gelar akademik apapun, baik di Universitas Budi Luhur maupun di perguruan tinggi lainnya.
2. Karya tulis ini bukan saduran/terjemahan, dan murni gagasan, rumusan dan pelaksanaan penelitian/implementasi saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing akademik dan pembimbing organisasi tempat riset.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Saya menyerahkan hak milik atas karya tulis ini kepada Universitas Budi Luhur, dan oleh karenanya Universitas Budi Luhur berhak melakukan pengelolaan atas karya tulis ini sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh berdasarkan karya tulis ini, serta sanksi lainnya sesuai dengan norma di Universitas Budi Luhur dan Undang-Undang yang berlaku.

Jakarta, 26 Juni 2026

SOA0007372647
Febri ardana

PERNYATAAN PENGGUNAAN AI GENERATIF

SURAT PERNYATAAN PENGGUNAAN AI GENERATIF DALAM PENULISAN KARYA ILMIAH

Saya yang bertandatangan di bawah ini:

Nama Lengkap : Febri Ardana
NIM : 2212500199
Program Studi : Sistem Informasi
Fakultas : Teknologi Informasi
Judul Karya Ilmiah : Penerapan Algoritma K-Means Clustering Untuk Analisis Pola
Penjualan Berdasarkan Waktu, Produk, dan Wilayah Pada
GWMM SHOP
Jenis Karya Ilmiah : Tugas Akhir

Dengan ini saya menyatakan bahwa dalam penyusunan karya ilmiah sebagaimana judul di atas:

1. **TIDAK** menggunakan alat AI Generatif dalam bentuk apapun.

Atau

MENGUNAKAN alat AI Generatif dengan rincian sebagaimana tercantum pada bagian
Rincian Penggunaan AI Generatif di bawah ini:

No	Nama Alat AI	Tujuan Penggunaan	Bagian Karya yang Dibantu
1	Claude (Anthropic)	Mempertajam pemikiran mengenai fenomena dan gap penelitian.	BAB I Pendahuluan (Latar Belakang)
2	Claude (Anthropic)	Membantu menyusun alur kerangka berpikir.	BAB II Landasan Teori
3	Claude (Anthropic)	Membantu merumuskan poin-poin Kesimpulan yang perlu dihasilkan.	BAB V Kesimpulan dan Saran
4	Claude (Anthropic)	Membantu merumuskan poin-poin yang terdapat pada slide presentasi.	Slide Presentasi (PPT)
5	QuillBot AI	Memparafrase kalimat yang kurang ilmiah agar sesuai gaya bahasa akademik.	Seluruh Naskah Tugas Akhir

- Penggunaan AI generatif sebagaimana dirinci di atas semata-mata bersifat sebagai alat bantu dan tidak menggantikan kontribusi intelektual orisinal saya.
 - Seluruh luaran AI yang digunakan telah diverifikasi kebenarannya melalui sumber-sumber primer yang dapat diandalkan.
 - Saya tidak memasukkan data pribadi partisipan penelitian atau data sensitif lainnya ke dalam platform AI generatif.
- Seluruh ide pokok, argumen utama, analisis, dan kesimpulan dalam artikel ilmiah ini merupakan hasil karya intelektual saya sendiri.
 - Saya bertanggung jawab penuh atas keseluruhan isi, kebenaran, dan orisinalitas karya ilmiah ini.

- Saya bersedia menerima sanksi akademik sesuai ketentuan yang berlaku apabila di kemudian hari terbukti bahwa terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini.

Jakarta, 11 Juli 2026


(Febri Ardana)

KATA PENGANTAR

Segala puji syukur ke hadirat Allah SWT atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan Skripsi yang berjudul **"PENERAPAN ALGORITMA K-MEANS *CLUSTERING* UNTUK ANALISIS POLA PENJUALAN BERDASARKAN WAKTU, PRODUK, DAN WILAYAH PADA GWMM SHOP"** ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) di Universitas Budi Luhur, Fakultas Teknologi Informasi, Program Studi Sistem Informasi.

Penelitian ini bertujuan untuk menerapkan metode Data Mining dengan menggunakan algoritma K-Means dalam menganalisis pola penjualan produk pada data transaksi penjualan. Melalui penelitian ini, diharapkan dapat memberikan wawasan yang bermanfaat bagi dunia bisnis dalam menyusun strategi pemasaran berbasis data, serta memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang Data Mining.

Dalam penyusunan skripsi ini, penulis menyadari bahwa keberhasilan penelitian ini tidak terlepas dari bimbingan, dukungan, dan bantuan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan ucapan terima kasih, penulis menyampaikan apresiasi sebesar-besarnya kepada:

1. Allah SWT, atas segala nikmat, rahmat, dan kemudahan yang telah diberikan, sehingga skripsi ini dapat diselesaikan dengan baik.
2. Kedua orang tua dan keluarga, yang senantiasa memberikan doa, dukungan, dan motivasi tanpa henti selama penulis menempuh pendidikan.
3. Ibu Anisa Rahmat Purnama Putri, selaku kekasih penulis atas dukungan dan perhatiannya dalam pengerjaan proses Tugas Akhir dari awal hingga akhir.
4. Bapak Prof. Dr. Agus Setyo Budi, M.Sc., selaku Rektor Universitas Budi Luhur, atas dukungan dan perhatiannya dalam penyelenggaraan pendidikan yang berkualitas.
5. Bapak Dr. Ir. Deni Mahdiana, S.Kom., M.M., M.Kom., selaku Deputy Rektor Bidang Akademik, atas segala dukungan yang telah diberikan.
6. Bapak Bima Cahya Putra, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi, yang telah memberikan arahan serta persetujuan dalam proses penyusunan skripsi ini.
7. Bapak Dr. Ir. Achmad Solichin, S.Kom., M.T.I., selaku Dekan Fakultas Teknologi Informasi, atas segala dukungan yang telah diberikan.
8. Bapak M. Ainur Rony, S.Kom., M.TI., selaku Dosen Pembimbing Skripsi, atas bimbingan, saran, dan motivasi yang telah diberikan selama proses penelitian hingga penyusunan skripsi ini selesai.
9. Ibu Yesi Puspita Dewi, S.Kom., M.Kom., selaku dosen penasehat akademik atas dukungan dan komitmennya dalam mendukung kelancaran kegiatan akademik.
10. Seluruh dosen Fakultas Teknologi Informasi Universitas Budi Luhur, atas ilmu, wawasan, dan inspirasi yang telah diberikan selama masa perkuliahan.
11. Rekan-rekan mahasiswa Universitas Budi Luhur, khususnya Program Studi Sistem Informasi, yang telah memberikan dukungan dan semangat selama proses pengerjaan skripsi ini.

12. Pemilik GWMM SHOP, yang telah memberikan izin dan kepercayaan kepada penulis untuk menggunakan data transaksi penjualan sebagai bahan penelitian ini.
13. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah turut berkontribusi dalam penyelesaian skripsi ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan guna penyempurnaan penelitian ini. Semoga hasil penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang Data Mining dan Analisis Pola Penjualan.

Jakarta, Juni 2026
Penulis,



Febri Ardana
(2212500199)

DAFTAR TABEL

Tabel 4.1 Atribut Dataset GWMM SHOP	17
Tabel 4.2 Hasil Penghapusan Data Duplikat.....	19
Tabel 4.3 Contoh Ekstraksi Fitur Waktu	22
Tabel 4.4 Transformasi Kategori Nama Produk	22
Tabel 4.5 Transformasi Kategori Provinsi	24
Tabel 4.6 Bulan Normalisasi.....	25
Tabel 4.7 Hasil Normalisasi Nama Produk.....	26
Tabel 4.8 Hasil Normalisasi Provinsi	27
Tabel 4.9 Hasil Normalisasi Jumlah	28
Tabel 4. 10 Hasil Normalisasi Total Harga Produk	28
Tabel 4.11 Hasil DBI pada Algoritma K-Means	30
Tabel 4.12 Hasil DBI pada Algoritma K-medoids.....	30
Tabel 4.13 Data Jumlah Minimal, Maksimal, dan Rata-rata Pada Setiap Atribut Data	31
Tabel 4.14 Karakteristik Setiap <i>Cluster</i>	33
Tabel 4.15 Dataset pada Algoritma Terpilih	34
Tabel 4. 16 Titik Pusat Awal Tiap <i>Cluster</i>	35
Tabel 4.17 Hasil Perhitungan Setiap Data ke Setiap Pusat <i>Cluster</i>	38
Tabel 4.18 Tabel Hasil Pengujian Davies Bouldien Index.....	40

DAFTAR GAMBAR

Gambar 2.1 CRISP-DM.....	5
Gambar 3.1 Tahapan Penelitian	12
Gambar 4.1 Dataset Mentah GWMM SHOP	19
Gambar 4.2 Hasil Pengecekan Missing Values.....	20
Gambar 4.3 Pemilihan Fitur Kolom Terpakai.....	21
Gambar 4.4 Dataset Setelah Dilakukan Transformasi Numerik.....	24
Gambar 4.5 Dataset Setelah Dilakukan Normalisasi.....	29
Gambar 4.6 Penyajian Model dengan RapidMiner	31
Gambar 4.7 Hasil Jumlah <i>Cluster</i>	31
Gambar 4.8 Pengujian Elbow Method Menggunakan RapidMiner	39
Gambar 4.9 Grafik Elbow Method	39
Gambar 4.10 Pengujian Dengan Davies Bouldien Index	40

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK.....	ii
PERNYATAAN TIDAK PLAGIAT	iii
PERNYATAAN PENGGUNAAN AI GENERATIF	iv
KATA PENGANTAR	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
DAFTAR ISI.....	ix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Masalah.....	2
1.3 Tujuan dan Manfaat Penelitian	3
1.4 Sistematika Penulisan	3
BAB II STUDI PUSTAKA.....	5
2.1 Teori – Teori Umum.....	5
2.1.1 Pengertian <i>Data Mining</i>	5
2.1.2 CRISP – DM	5
2.1.3 <i>K-Means</i>	6
2.1.4 Metode Elbow	7
2.1.5 <i>Davies Bouldin Index</i> (DBI)	7
2.2 Teori Khusus	8
2.2.1 <i>E-Commerce</i>	8
2.2.2 Klusterisasi.....	8
2.2.3 Transformasi Data.....	9
2.2.4 Normalisasi Data.....	9
2.3 <i>Tools</i> Yang Digunakan	9
2.3.1 Google Collab	9
2.3.2 RapidMiner	10
2.3.3 Microsoft Excel.....	10
2.4 Studi Literatur	10
BAB III METODOLOGI PENELITIAN	12
3.1 Tahapan Penelitian.....	12
3.1.1 Studi Literatur	12
3.1.2 Identifikasi Masalah.....	12
3.1.3 Analisa Masalah.....	13
3.1.4 Kesimpulan dan Saran	14
3.2 Perbedaan dari Penelitian Sebelumnya.....	14
3.3 <i>Data Preprocessing</i>	15
3.4 Penentuan Data Latih dan Data Uji	16
3.5 Pengembangan Model.....	16
3.6 Teknik Pengujian	16
BAB IV ANALISIS DAN PEMBAHASAN.....	17
4.1 Pengumpulan Data.....	17

4.2	<i>Data Preprocessing / Pra-Pemrosesan Data</i>	19
4.2.1	<i>Data Cleaning</i>	19
4.2.2	<i>Data Integration</i>	21
4.2.3	Transformasi Data	21
4.3	Pemodelan	29
4.3.1	Hasil Komparasi Model	29
4.3.2	Penyajian Model Terbaik	30
4.3.3	Perhitungan Algoritma Terpilih	34
4.4	Pengujian	39
BAB V PENUTUP		41
5.1	Kesimpulan	41
5.2	Saran	42
DAFTAR PUSTAKA		43
LAMPIRAN		45

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, pertumbuhan *e-commerce* di Indonesia berlangsung cukup pesat. Berdasarkan Survei *E-Commerce* yang dirilis Badan Pusat Statistik (BPS) pada tahun 2025, jumlah usaha *e-commerce* di Indonesia pada akhir tahun 2023 mencapai 3,82 juta usaha yang tersebar di 38 provinsi, dengan nilai transaksi sebesar Rp1.100,87 triliun atau naik 27,40 persen dibandingkan tahun sebelumnya. Kondisi ini mendorong semakin banyak pelaku UMKM untuk mengandalkan platform *marketplace* seperti Shopee sebagai saluran penjualan utama mereka. Salah satu sektor yang ikut terdampak adalah industri parfum, yang permintaannya dipengaruhi oleh faktor musiman, momen tertentu seperti Lebaran dan Natal, serta preferensi konsumen yang berbeda di setiap wilayah. Keragaman variasi produk seperti merek, volume, jenis aroma, dan segmen harga membuat analisis pola penjualan semakin kompleks dan tidak mungkin dikerjakan secara manual.

GWMM SHOP adalah salah satu UMKM yang bergerak di bidang penjualan parfum melalui platform Shopee. Selama beroperasi, toko ini telah menghasilkan data transaksi dalam jumlah yang cukup besar, mulai dari waktu pembelian, jenis produk yang terjual, hingga wilayah pengiriman pembeli. Namun, data tersebut belum pernah dimanfaatkan secara serius untuk mendukung pengambilan keputusan bisnis. Pengelolaan toko masih dilakukan secara konvensional, dan keputusan terkait stok, waktu promosi, serta target pasar masih bergantung pada perkiraan dan intuisi pemilik usaha. Akibatnya, potensi yang tersimpan dalam data transaksi historis belum pernah diolah secara sistematis.

Dari kondisi tersebut muncul tiga masalah yang konkret. Pertama, GWMM SHOP belum memahami produk mana yang paling banyak terjual sehingga pengadaan stok masih bersifat spekulatif. Kedua, toko ini belum memahami kapan lonjakan transaksi terjadi sehingga momen puncak seperti *harbolnas* atau akhir tahun sering belum terantisipasi dengan baik. Ketiga, GWMM SHOP belum memahami provinsi mana yang paling aktif bertransaksi sehingga strategi pemasaran belum dapat diarahkan ke wilayah yang paling potensial. Ketiga masalah ini bermuara pada satu persoalan yang sama, yaitu data transaksi yang sudah ada belum pernah diolah secara terstruktur. Dalam persaingan *marketplace* yang semakin ketat, kemampuan membaca data bukan lagi nilai tambah, melainkan kebutuhan dasar.

Pendekatan yang relevan untuk menjawab permasalahan ini adalah data mining, khususnya teknik *clustering*. *Clustering* memungkinkan pengelompokan data transaksi ke dalam kelompok-kelompok yang memiliki karakteristik serupa tanpa memerlukan label kategori sebelumnya, sehingga cocok digunakan pada data penjualan yang belum terstruktur. Algoritma K-Means dipilih karena efisien secara komputasi, mudah diimplementasikan, dan menghasilkan keluaran yang mudah diinterpretasikan untuk keperluan analisis bisnis. Penentuan jumlah kluster yang optimal dilakukan

menggunakan Elbow Method, sedangkan kualitas hasil clustering diukur menggunakan metrik Davies-Bouldin Index (DBI). Dengan pendekatan ini, pola penjualan berdasarkan dimensi waktu, produk, dan wilayah dapat diidentifikasi secara sistematis dari data transaksi GWMM SHOP.

Beberapa penelitian terdahulu telah membuktikan efektivitas K-Means dalam konteks analisis penjualan. (Dermawan et al., 2024) pada Toko Daun Indah di Shopee berhasil menemukan pola pengelompokan produk berdasarkan tingkat penjualan menggunakan K-Means dengan kerangka CRISP-DM yang divalidasi menggunakan Davies-Bouldin Index. (Riadi et al., 2023) juga berhasil mengelompokkan produk parfum ke dalam beberapa kluster berdasarkan karakteristik penjualan sehingga manajemen toko dapat menyusun strategi pengadaan yang lebih tepat. (Dona et al., 2025) menunjukkan bahwa pengelompokan produk kosmetik menggunakan K-Means dapat mencegah kehabisan stok sekaligus menjadi acuan perencanaan persediaan yang lebih akurat. Meski demikian, penelitian tersebut umumnya hanya berfokus pada satu dimensi analisis dan belum mengintegrasikan tiga dimensi sekaligus, yaitu waktu, produk, dan wilayah, dalam satu kerangka clustering yang komprehensif pada objek UMKM parfum berbasis Shopee.

1.2 Masalah

a. Identifikasi Masalah

Berdasarkan latar belakang dari penelitian ini maka dapat diidentifikasi permasalahan sebagai berikut:

- 1) Data transaksi penjualan pada GWMM SHOP belum dianalisis secara mendalam.
- 2) Belum tersedia pengelompokan pola penjualan berdasarkan waktu, produk, dan wilayah yang menyebabkan momen penting sering terlewat, pengelolaan stok berdasarkan produk yang paling laris, serta tidak mengetahui provinsi yang sering melakukan transaksi.
- 3) Pengambilan keputusan masih dilakukan secara spekulatif yang dapat menyebabkan kerugian bagi perusahaan.

b. Batasan Masalah

Batasan atau ruang lingkup masalah dalam penelitian ini yaitu:

- 1) Data yang digunakan dalam penelitian adalah data transaksi penjualan shopee pada GWMM SHOP selama 2 tahun terakhir dari tahun 2024 sampai 2025.
- 2) Analisis yang digunakan hanya berfokus pada 3 atribut utama yaitu produk, waktu (bulan), dan wilayah (provinsi).
- 3) Metode data mining yang digunakan pada penelitian ini menggunakan algoritma K-Means.
- 4) *Clustering* yang digunakan adalah sangat tinggi, tinggi, sedang dan rendah.
- 5) Penelitian bertujuan untuk mengelompokkan dan menganalisis pola penjualan, bukan melakukan prediksi (*forecasting*).

- 6) Hasil analisis yang diperoleh digunakan sebagai bahan rekomendasi dalam pengambilan keputusan, tanpa membahas implemementasi sistem secara *real-time*.

c. Rumusan Masalah

- 1) Bagaimana penerapan algoritma K-Means dalam mengelompokkan data penjualan pada GWMM SHOP?
- 2) Bagaimana pola penjualan yang terbentuk berdasarkan waktu, produk, dan wilayah?
- 3) Bagaimana hasil clustering dapat digunakan untuk mendukung pengambilan keputusan bisnis?

1.3 Tujuan dan Manfaat Penelitian

a. Tujuan

- 1) Mengolah data transaksi penjualan agar dapat menghasilkan informasi yang berguna.
- 2) Menerapkan algoritma K-Means *Clustering* pada data penjualan.
- 3) Mengelompokkan data penjualan berdasarkan pola tertentu.
- 4) Mengidentifikasi pola penjualan berdasarkan waktu, produk, dan wilayah.

b. Manfaat

Penelitian ini diharapkan bermanfaat bagi tiga pihak. Bagi GWMM SHOP, hasil penelitian memberikan gambaran berbasis data mengenai produk terlaris, waktu terjadinya lonjakan transaksi, dan provinsi dengan aktivitas pembelian tertinggi. Data tersebut dapat digunakan perusahaan sebagai dasar dalam mengambil keputusan operasional dan pemasaran yang lebih tepat sasaran. Bagi dunia akademik, penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang membahas penerapan algoritma *clustering* pada analisis penjualan UMKM berbasis *platform digital*. Sementara itu, bagi peneliti, penelitian ini menjadi kesempatan untuk mempraktikkan ilmu *data mining* dan algoritma K-Means yang selama ini hanya dipelajari secara teori di perkuliahan, kali ini diterapkan langsung pada permasalahan bisnis yang nyata dan terukur.

1.4 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II STUDI PUSTAKA

Bab ini menjelaskan tentang teori-teori yang mendasari penelitian, mulai dari konsep data mining, *tools* yang digunakan, algoritma K-Means *Clustering*, hingga tinjauan penelitian terdahulu yang relevan.

BAB III METODOLOGI PENELITIAN

Menjelaskan metodologi penelitian, termasuk kerangka kerja CRISP-DM yang digunakan, metode pengumpulan data, variabel penelitian, dan alur analisis.

BAB IV ANALISIS DAN PEMBAHASAN

Menyajikan hasil analisis dan pembahasan secara menyeluruh, mulai dari pengumpulan data, preprocessing, pemodelan K-Means, pengujian dengan Davies-Bouldin Index, hingga interpretasi setiap klaster yang terbentuk.

BAB V KESIMPULAN

Berisi kesimpulan yang menjawab seluruh rumusan masalah secara spesifik, serta saran untuk penelitian dan pengembangan lebih lanjut.

BAB II

STUDI PUSTAKA

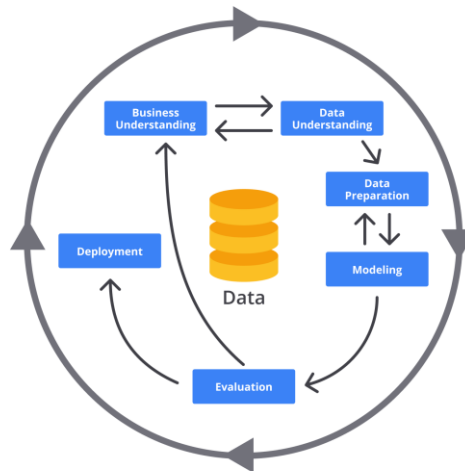
2.1 Teori – Teori Umum

2.1.1 Pengertian *Data Mining*

Menurut (Ilmu, 2024) *Data mining* secara sederhana dapat dipahami sebagai proses menggali informasi yang berguna dari kumpulan data berukuran besar yang belum terstruktur, dengan tujuan menemukan pola atau hubungan yang sebelumnya tidak terlihat. Dalam konteks bisnis, pendekatan ini berguna karena mampu mengolah volume data transaksi yang besar menjadi informasi yang bisa langsung dipakai untuk mengambil keputusan. Secara umum, proses *data mining* mencakup beberapa tahapan mulai dari seleksi data, *preprocessing*, transformasi, penerapan algoritma, sampai interpretasi hasil.

2.1.2 CRISP – DM

Menurut (Dermawan et al., 2024) CRISP-DM (*Cross Industry Standard Process for Data Mining*) adalah kerangka kerja analisis data yang sudah banyak dipakai di berbagai industri karena alurnya yang runtut dan tidak terlalu kaku untuk disesuaikan. Kerangka ini membawa peneliti melewati proses dari identifikasi masalah bisnis hingga penerapan hasil analisis ke dalam solusi yang bisa langsung digunakan, termasuk dalam konteks UMKM. Seperti terlihat pada Gambar 2.1, CRISP-DM mencakup 6 tahapan proses yang saling terhubung satu sama lain.



Gambar 2.1 CRISP-DM

a. *Business Understanding*

Tahap awal ini berfokus pada pemahaman tujuan serta kebutuhan proyek dari sudut pandang bisnis secara menyeluruh. Peneliti harus mampu menerjemahkan masalah nyata pemilik usaha menjadi tujuan teknis data mining yang tepat.

- b. *Data Understanding*
Tahap ini melibatkan pengumpulan data awal serta pemeriksaan kualitas data untuk mengenali karakteristik dasar dari informasi tersedia. Peneliti melakukan eksplorasi guna menemukan wawasan pertama atau mendeteksi adanya masalah seperti data tidak valid. Untuk menentukan variabel yang paling relevan untuk diolah pada tahap selanjutnya.
- c. *Data Preparation*
Persiapan data mencakup aktivitas pembersihan serta transformasi data mentah menjadi format yang siap untuk digunakan dalam proses modeling.
- d. *Modeling*
Tahap ini dilakukan dengan menerapkan algoritma *clustering* pada dataset yang telah disiapkan sebelumnya. Peneliti menentukan parameter optimal untuk mendapatkan hasil pengelompokan data yang paling akurat sesuai dengan pengujian model.
- e. *Evaluation*
Setelah model berhasil dibuat maka tahap evaluasi dilakukan untuk mengukur kualitas hasil pengelompokan menggunakan metrik pengujian. Peneliti meninjau kembali apakah hasil analisis sudah sesuai dengan tujuan bisnis yang telah ditetapkan pada tahap pertama.
- f. *Deployment*
Tahap akhir ini gunanya menginterpretasikan hasil data mining dalam bentuk laporan atau rekomendasi strategi untuk digunakan dalam pengambilan keputusan.

2.1.3 *K-Means*

Menurut (Fatkhudin et al., 2023) *K-Means* adalah metode *clustering* dalam kategori *unsupervised machine learning* yang bekerja dengan cara mengelompokkan data ke dalam sejumlah k kelompok berdasarkan kemiripan karakteristik antar datanya. Metode ini banyak dipilih karena cara kerjanya yang tidak terlalu rumit dan tidak membutuhkan komputasi yang berat. Dalam ranah data mining, *K-Means* masuk ke dalam kelompok metode *partitioning*, yaitu pendekatan yang memecah dataset menjadi beberapa bagian tanpa ada irisan di antara kelompok-kelompok tersebut sehingga setiap titik data hanya akan berada di satu *cluster*. Berikut metode dalam algoritma *K-Means*:

- a. Menentukan nilai k optimal.
- b. Hitung jarak antara setiap data dengan masing-masing centroid menggunakan rumus Euclidean Distance.
- c. Menentukan data mana yang masuk ke *cluster* tertentu, menggunakan pengukuran jarak antara setiap titik data dengan centroid menggunakan rumus *Euclidean Distance*:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Keterangan:

x_i = data kriteria.

μ_j = *centroid* pada *cluster* ke-j.

- d. Masukkan setiap data ke dalam *cluster* yang *centroid*-nya paling dekat.
- e. Setelah semua data ditetapkan ke *cluster* masing-masing, *centroid* setiap *cluster* dihitung ulang menggunakan rata-rata dari seluruh data yang ada di dalamnya:

$$C_k = \frac{1}{n_k} \sum d_i$$

Keterangan:

C_k = Titik pusat (*centroid*) baru untuk *cluster* ke-k.

n_k = Jumlah total titik data yang tergabung ke dalam *cluster* ke-k.

d_i = Titik data ke-i yang menjadi anggota dari *cluster* ke-k.

- f. Ulangi langkah 3 sampai 5 sampai posisi *centroid* tidak lagi berubah.

2.1.4 Metode Elbow

Metode Elbow adalah salah satu pendekatan yang umum digunakan untuk menentukan jumlah kluster optimal dalam algoritma K-Means *Clustering*. Cara kerjanya cukup langsung: algoritma K-Means dijalankan berulang kali dengan nilai k yang berbeda, mulai dari k=1 hingga batas nilai k yang telah ditentukan, lalu pada setiap percobaan dihitung nilai *Within Cluster Sum of Squares* (WCSS) atau yang sering disebut inerti. WCSS sendiri mengukur total kuadrat jarak antara setiap titik data dengan *centroid* klasternya masing-masing (Vania et al., 2023), yang secara matematis dituliskan sebagai:

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

WCSS = total jumlah kuadrat jarak seluruh data terhadap *centroid*nya.

K = jumlah *cluster* yang sedang diuji.

C_k = himpunan data yang termasuk dalam *cluster* ke-k.

x_i = titik data ke-i dalam *cluster* C_k

μ_k = nilai *centroid* dari *cluster* ke-k

$\|x_i - \mu_k\|^2$ = kuadrat jarak Euclidean antara data x_i dengan *centroid* μ_k

2.1.5 Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan metode evaluasi *cluster* yang digunakan untuk menentukan jumlah *cluster* yang optimal dengan didukung pengukuran jarak yang tepat. Hasil akhir DBI berupa nilai

numerik yang mencerminkan kualitas keseluruhan dari proses *clustering* yang telah dilakukan. Semakin kecil nilai DBI yang dihasilkan, semakin baik kualitas *cluster* tersebut karena menandakan bahwa data dalam satu *cluster* sangat rapat satu sama lain sekaligus terpisah jauh dari *cluster* lainnya. Prinsip inilah yang menjadikan DBI sebagai salah satu metode evaluasi internal yang objektif karena tidak memerlukan data label eksternal untuk menilai kualitas *clustering*, sehingga sangat cocok dipadukan dengan algoritma K-Means yang bersifat *unsupervised*. (Jollyta et al., 2023) Adapun rumus untuk mengetahui nilai DBI sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{d(c_i, c_j)} \right)$$

Keterangan:

DBI = nilai akhir *Davies Bouldin Index*.

K = jumlah *cluster* yang terbentuk.

S_i = nilai *scatter* dari *cluster* ke- i

S_j = nilai *scatter* dari *cluster* ke- j

$d(\mu_i, \mu_j)$ = jarak antara centroid *cluster* ke- i dan ke- j

$\max_{j \neq i}$ = nilai rasio tertinggi terhadap semua *cluster* lain selain *cluster* ke- i

2.2 Teori Khusus

2.2.1 E-Commerce

E-commerce atau perdagangan elektronik merupakan bentuk transaksi jual beli barang maupun jasa yang dilakukan melalui media elektronik, khususnya dengan memanfaatkan jaringan internet sebagai sarana utamanya. Dalam perkembangannya, *e-commerce* tidak hanya mencakup aktivitas pembelian dan penjualan produk secara online, tetapi juga melibatkan berbagai proses bisnis seperti pemasaran digital, manajemen logistik, hingga layanan purna jual yang semuanya terintegrasi dalam satu ekosistem digital (Dasopang, 2024). Kehadiran *e-commerce* telah mengubah cara pelaku usaha berinteraksi dengan konsumen secara signifikan, karena transaksi dapat dilakukan kapan saja dan di mana saja tanpa terbatas oleh jarak maupun waktu.

2.2.2 Klasterisasi

Klasterisasi atau *clustering* merupakan salah satu teknik dalam analisis data yang bertujuan untuk mengelompokkan sekumpulan objek atau data ke dalam beberapa kelompok (*cluster*) berdasarkan tingkat kemiripan atau kedekatan karakteristiknya. Data yang berada dalam satu *cluster* memiliki kemiripan yang tinggi satu sama lain, sedangkan data yang berada di *cluster* berbeda memiliki karakteristik yang jauh berbeda. Proses pengelompokan ini dilakukan secara otomatis tanpa menggunakan label kelas yang telah ditentukan sebelumnya, sehingga *clustering* termasuk ke dalam kategori *unsupervised learning* dalam bidang *machine*

learning dan *data mining* (Khalif et al., 2024). Dalam konteks *e-commerce*, teknik *clustering* sangat berguna untuk mengelompokkan pelanggan berdasarkan pola perilaku transaksinya, sehingga hasil pengelompokan tersebut dapat dimanfaatkan oleh pelaku usaha untuk menyusun strategi pemasaran yang lebih tepat sasaran.

2.2.3 Transformasi Data

Menurut (Gori et al., 2024) Transformasi data adalah salah satu tahap dalam *preprocessing* yang mengubah format data agar kompatibel dengan algoritma yang akan dipakai. Secara umum, proses *preprocessing* mencakup beberapa langkah seperti integrasi data, pelabelan, transformasi, dan pembersihan data. Semua langkah ini dikerjakan sebelum algoritma *machine learning* mulai berjalan. Ketika dataset memiliki kolom bertipe kategorikal, salah satu teknik yang sering dipilih adalah *Label Encoding*. Teknik ini bekerja dengan mengonversi setiap nilai kategori pada suatu kolom menjadi angka, sehingga algoritma *machine learning* bisa membaca dan memproses kolom tersebut.

2.2.4 Normalisasi Data

Normalisasi data adalah proses menyamakan skala nilai pada setiap atribut numerik agar tidak ada satu atribut yang nilainya jauh lebih besar dari atribut lain dan secara tidak sengaja mendominasi hasil perhitungan. Ini menjadi langkah yang perlu dilakukan sebelum menjalankan algoritma berbasis jarak seperti K-Means, karena algoritma tersebut mengandalkan perhitungan jarak Euclidean yang sangat terpengaruh oleh perbedaan skala antara atribut. (Ramadhan et al., 2026) Dalam konteks penelitian berbasis K-Means, normalisasi dilakukan pada atribut numerik menggunakan *Min-Max Scaling* agar setiap atribut berada dalam rentang yang setara, sehingga bias dalam perhitungan jarak dapat ditekan. Metode yang digunakan adalah *Min-Max Scaler*, yaitu teknik yang mengubah setiap nilai data ke dalam rentang 0 hingga 1 menggunakan rumus berikut:

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Keterangan:

x' = Nilai data baru setelah dinormalisasi.

x_i = Nilai data asli yang sedang diproses.

$\min(x)$ = Nilai minimum dari suatu atribut.

$\max(x)$ = Nilai maksimum dari suatu atribut.

2.3 Tools Yang Digunakan

2.3.1 Google Collab

Google Colaboratory, atau Google Colab, adalah platform berbasis *cloud* yang dikembangkan oleh Google dan dapat diakses gratis melalui *browser* tanpa perlu menginstal perangkat lunak apapun. Pengguna bisa langsung mulai menulis kode Python tanpa setup awal, berbagi *notebook* dengan orang lain, dan memanfaatkan sumber daya komputasi untuk

kebutuhan pemrosesan data yang cukup berat. Platform ini berjalan di atas lingkungan Jupyter *Notebook*, sehingga pengguna bisa menulis kode, menjalankannya langsung, dan menyimpan hasilnya dalam satu dokumen yang terhubung dengan Google Drive (Wilyani et al., 2024).

2.3.2 RapidMiner

RapidMiner adalah perangkat lunak *open source* untuk analisis data yang dirilis di bawah lisensi AGPL versi 3. Platform ini sudah dipakai di lebih dari 40 negara dan mencakup berbagai kebutuhan dalam proses data mining, mulai dari *preprocessing* hingga evaluasi model. Yang membedakan RapidMiner dari tools lain adalah pendekatannya yang berbasis antarmuka visual *drag-and-drop*, sehingga alur kerja analisis bisa dibangun tanpa harus menulis kode dari awal. Kemudahan ini yang menjadi salah satu pertimbangan penggunaannya dalam penelitian ini, terutama untuk mempercepat proses eksplorasi dan pengujian model (Ananda et al., 2023).

2.3.3 Microsoft Excel

Microsoft Excel adalah perangkat lunak *spreadsheet* yang dikembangkan oleh Microsoft dan menjadi bagian dari paket Microsoft Office. Excel digunakan untuk mengolah data secara otomatis, mulai dari perhitungan sederhana, penggunaan rumus dan fungsi, pembuatan tabel dan grafik, hingga manajemen data untuk mendukung proses pengambilan keputusan (Pratiwi et al., 2023).

2.4 Studi Literatur

Penelitian mengenai pengelompokan data dalam konteks pengelolaan stok produk telah banyak dilakukan oleh berbagai peneliti sebelumnya, khususnya dengan menggunakan algoritma K-Means *Clustering*. Penelitian oleh (Wahyudi et al., 2022) yang berjudul "Implementasi Metode K-Means *Clustering* Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos" menunjukkan bahwa algoritma K-Means dapat dimanfaatkan untuk mengelompokkan data stok produk pada toko online berskala UMKM yang bergerak di bidang perdagangan kaos. Penelitian ini menggunakan data transaksi penjualan internal toko dengan atribut kode produk, nama produk, dan jumlah stok yang bergerak, kemudian dievaluasi menggunakan Davies-Bouldin Index. Hasil penelitian menunjukkan bahwa algoritma K-Means berhasil membentuk pengelompokan produk berdasarkan tingkat perputaran stok, sehingga dapat digunakan sebagai dasar pengambilan keputusan operasional dalam manajemen inventaris toko.

Penelitian yang dilakukan oleh (Kamalludin et al., 2023) berjudul "Penerapan Algoritme Klasterisasi K-Means Kinerja Produk dengan Analisis Recency Frequency Monetary pada Cafe XYZ" menerapkan pendekatan RFM (*Recency, Frequency, Monetary*) sebagai dasar pembentukan fitur sebelum dilakukan proses *clustering* terhadap data transaksi pada Cafe XYZ. Dataset yang digunakan memiliki atribut tanggal transaksi, id pelanggan,

nama produk, frekuensi pembelian, dan nilai transaksi. Hasil penelitian ini menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan kinerja produk pada usaha skala kecil, sehingga pengelola Cafe XYZ dapat menyusun strategi pemasaran yang lebih terarah berdasarkan karakteristik produk pada masing-masing klaster yang terbentuk.

Penelitian lain oleh (Fitri et al., 2023) berjudul "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means pada PT XYZ" menjadikan wilayah sebagai variabel utama dalam proses *clustering* data penjualan pada sebuah perusahaan besar. Dataset yang digunakan mencakup atribut wilayah penjualan, jumlah transaksi per wilayah, dan total nilai penjualan, dengan evaluasi klasterisasi menggunakan Silhouette Score. Hasil penelitian ini menunjukkan bahwa klasterisasi berbasis wilayah mampu memetakan distribusi penjualan secara geografis, sehingga informasinya dapat digunakan oleh PT XYZ sebagai dasar pengambilan keputusan strategis dalam menentukan target pasar di setiap wilayah.

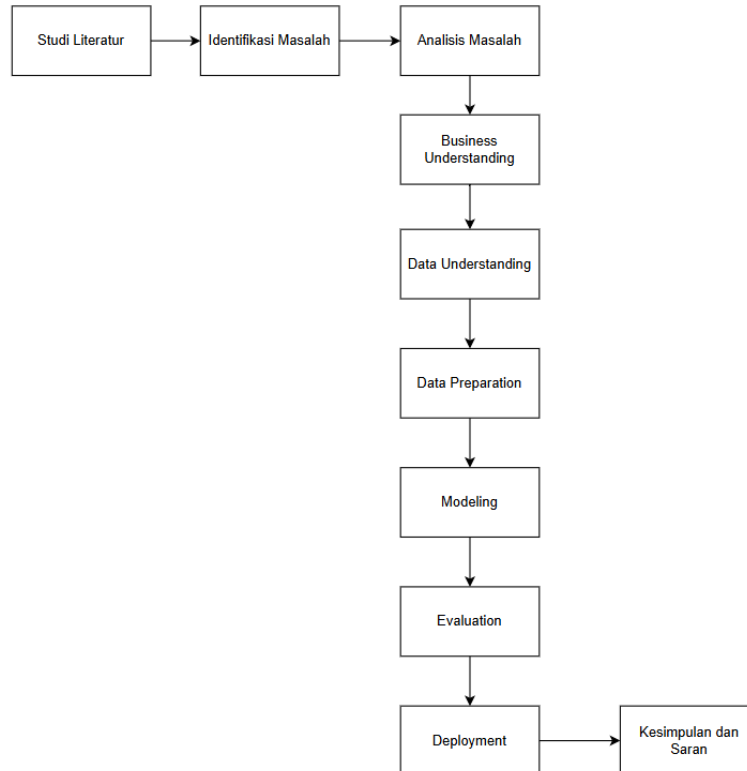
(Eriskiannisa Febrianty Luchia Awalina, 2023), dalam penelitiannya yang berjudul "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means *Clustering* pada Transaksi Online Retail", membahas penerapan algoritma K-Means dalam mensegmentasi pelanggan berdasarkan data transaksi pada platform online retail dengan atribut id pelanggan, tanggal pembelian, frekuensi transaksi, dan total nilai belanja. Hasil penelitian ini menunjukkan bahwa K-Means *Clustering* efektif diterapkan pada lingkungan *e-commerce*, di mana terbentuknya beberapa segmen pelanggan dengan karakteristik belanja yang berbeda dapat dimanfaatkan sebagai dasar penyusunan strategi pemasaran yang lebih terarah dan personal bagi setiap segmen.

Sebagai perbandingan terakhir, (Dermawan et al., 2024) melalui penelitiannya yang berjudul "Analisis Data Penjualan Menggunakan Algoritma K-Means *Clustering* Pada Toko Daun Indah di Shopee" menerapkan algoritma K-Means dengan kerangka metodologi CRISP-DM untuk menganalisis data penjualan sebuah toko yang beroperasi di platform Shopee. Dataset yang digunakan bersumber dari data ekspor Shopee Seller Center toko tersebut, dengan atribut nama produk, kategori produk, dan jumlah terjual, yang dievaluasi menggunakan Davies-Bouldin Index (DBI). Hasil penelitian ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan produk berdasarkan tingkat penjualannya, sehingga hasil klasterisasi tersebut dapat digunakan oleh pemilik Toko Daun Indah untuk menentukan strategi promosi dan pengelolaan stok produk yang lebih efektif di platform Shopee.

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini menggunakan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai kerangka kerja analisis data. CRISP-DM membagi proses penelitian ke dalam enam tahap yang berurutan, sebagaimana ditunjukkan pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

Berikut adalah penjelasan dari setiap tahapan yang digambarkan pada Gambar 3.1:

3.1.1 Studi Literatur

Studi literatur dilakukan untuk mengumpulkan referensi dari jurnal ilmiah, buku teks, dan laporan penelitian sebelumnya. Tahap ini digunakan untuk memahami teori *clustering*, cara kerja metode K-Means, dan konteks bisnis penjualan UMKM di *marketplace* digital. Selain itu, hasil studi literatur menjadi dasar dalam menyusun kerangka pemikiran dan mengidentifikasi celah penelitian yang membedakan penelitian ini dari penelitian sebelumnya.

3.1.2 Identifikasi Masalah

Berdasarkan studi literatur dan kondisi di lapangan, peneliti menemukan bahwa GWMM SHOP belum mengolah data transaksi penjualannya secara sistematis. Data transaksi yang tersedia di platform

Shopee belum pernah dianalisis untuk menemukan pola penjualan berdasarkan waktu, jenis produk, dan wilayah pengiriman. Dari kondisi ini, masalah penelitian dirumuskan sebagai kebutuhan untuk mengetahui pola penjualan yang bisa mendukung pengambilan keputusan bisnis secara berbasis data.

3.1.3 Analisa Masalah

Analisa masalah dilakukan melalui penerapan enam sub-tahap CRISP-DM:

- a. *Business Understanding*
Proses ini bertujuan memahami kebutuhan bisnis GWMM SHOP sebagai UMKM parfum yang berjualan di platform Shopee. Tujuan bisnis kemudian diterjemahkan ke dalam tujuan data mining, yaitu mengelompokkan pola penjualan berdasarkan waktu (bulan), produk, dan wilayah (provinsi) guna mendukung pengambilan keputusan operasional dan strategi pemasaran.
- b. *Data Understanding*
Mengeksplorasi data awal transaksi penjualan GWMM SHOP di Shopee selama periode 2024 hingga 2025. Eksplorasi dilakukan dengan memeriksa kualitas data, mengidentifikasi nilai yang hilang, dan memetakan atribut yang tersedia. Dari proses ini, ditetapkan tiga variabel utama yang akan digunakan dalam analisis, yaitu waktu transaksi (bulan), nama produk, dan wilayah pembeli (provinsi).
- c. *Data Preparation*
Melakukan seleksi variabel penting (bulan, produk, provinsi, dan jumlah transaksi), pembersihan data (menangani data hilang, duplikat, dan inkonsistensi), serta transformasi data berupa encoding kategorikal dan normalisasi numerik menggunakan metode Min-Max agar kompatibel dengan algoritma K-Means.
- d. *Modeling*
Menerapkan algoritma K-Means pada dataset yang telah disiapkan dan menentukan jumlah kluster optimal menggunakan Elbow Method. Setiap kluster yang terbentuk diidentifikasi karakteristiknya berdasarkan dimensi waktu, produk, dan wilayah.
- e. *Evaluation*
Mengevaluasi hasil klasterisasi menggunakan Davies-Bouldin Index (DBI) untuk mengukur kualitas pengelompokan secara kuantitatif. Nilai DBI yang lebih kecil menunjukkan hasil klasterisasi yang lebih baik. Selain itu dilakukan validasi visual untuk memastikan kluster yang terbentuk sesuai dengan karakteristik bisnis yang diharapkan.
- f. *Deployment*
Menyusun laporan hasil analisis dan memberikan rekomendasi strategi bisnis kepada GWMM SHOP berdasarkan karakteristik setiap kluster yang terbentuk, mencakup rekomendasi terkait manajemen stok, waktu promosi, dan target pemasaran berdasarkan wilayah.

3.1.4 Kesimpulan dan Saran

Bagian ini menyajikan temuan dari seluruh tahapan CRISP-DM, mencakup pola klaster yang terbentuk berdasarkan dimensi waktu, produk, dan wilayah. Dari temuan tersebut, dihasilkan rekomendasi praktis bagi GWMM SHOP terkait pengadaan stok, penentuan waktu promosi, dan penentuan wilayah target pemasaran yang paling potensial.

3.2 Perbedaan dari Penelitian Sebelumnya

Penelitian oleh (Wahyudi et al., 2022) menerapkan K-Means untuk mengelompokkan produk berdasarkan tingkat perputaran stok pada toko online berskala UMKM berjudul "Implementasi Metode K-Means *Clustering* Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos". Hasilnya membuktikan bahwa K-Means relevan digunakan pada konteks produk toko online skala kecil. Namun, analisisnya hanya mencakup satu dimensi, yaitu pergerakan stok, tanpa mempertimbangkan waktu transaksi maupun sebaran wilayah pembeli. Penelitian ini mengisi celah tersebut dengan mengintegrasikan tiga dimensi sekaligus dalam satu kerangka analisis.

Penelitian oleh (Kamalludin et al., 2023) berjudul "Penerapan Algoritme Klasterisasi K-Means Kinerja Produk dengan Analisis Recency Frequency Monetary pada Cafe XYZ" menggunakan pendekatan RFM sebagai dasar pembentukan fitur sebelum proses *clustering* dijalankan. Penelitian ini memperlihatkan bahwa K-Means relevan digunakan untuk menganalisis kinerja produk pada usaha skala kecil. Namun, cakupannya masih terbatas pada segmentasi kinerja produk secara umum dan belum membahas pola penjualan berdasarkan waktu maupun sebaran wilayah transaksi, sehingga hasil yang diperoleh tidak mencerminkan pola temporal maupun geografis penjualan.

Selanjutnya, penelitian yang dilakukan (Fitri et al., 2023) berjudul "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means pada PT XYZ" secara eksplisit menjadikan wilayah sebagai variabel utama dalam *clustering* penjualan, dengan hasil klasterisasi yang dievaluasi menggunakan Silhouette Score. Penelitian ini menjadi referensi penting karena menunjukkan bahwa dimensi wilayah dapat dijadikan dasar pengelompokan yang bermakna. Namun, analisis hanya mencakup satu dimensi wilayah tanpa mengintegrasikan dimensi waktu transaksi maupun kategori produk, serta dilakukan pada perusahaan besar dan bukan UMKM berbasis platform marketplace digital seperti Shopee.

Penelitian oleh (Eriskiannisa Febrianty Luchia Awalina, 2023) menerapkan K-Means untuk mensegmentasi pelanggan pada data transaksi online retail dengan mempertimbangkan variabel waktu pembelian dan frekuensi transaksi yang berjudul "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means *Clustering* pada Transaksi Online Retail". Hasilnya menunjukkan bahwa K-Means efektif digunakan dalam lingkungan *e-commerce*. Meski demikian, penelitian ini hanya berfokus pada segmentasi pelanggan secara umum, tanpa menyertakan

analisis distribusi geografis pembeli per wilayah, klasifikasi produk secara spesifik, maupun kerangka kerja CRISP-DM dalam proses analisisnya.

Penelitian oleh (Dermawan et al., 2024) berjudul "Analisis Data Penjualan Menggunakan Algoritma K-Means *Clustering* Pada Toko Daun Indah di Shopee" membuktikan bahwa kombinasi K-Means, kerangka CRISP-DM, dan *Davies-Bouldin Index* efektif diterapkan pada data penjualan toko di Shopee. Penelitian ini paling relevan dengan konteks penelitian yang sedang dilakukan karena menggunakan objek yang sama, yaitu toko di *platform* Shopee, serta alat evaluasi yang sama, yaitu *Davies-Bouldin Index*. Namun, cakupannya masih terbatas pada pengelompokan berdasarkan atribut produk dan variasi nama, tanpa mengintegrasikan dimensi waktu transaksi dan wilayah pengiriman sebagai variabel analisis.

3.3 *Data Preprocessing*

Data penelitian ini diperoleh dari toko GWMM Shop pada platform Shopee dan mencakup seluruh transaksi penjualan produk parfum periode Januari 2024 hingga Desember 2025. Dataset terdiri dari 4.163 transaksi dengan total 42 atribut, yang memuat informasi nomor pesanan, nama produk, jenis produk, harga awal, harga setelah diskon, jumlah transaksi, metode pembayaran, opsi pengiriman, serta lokasi penerima berdasarkan kota dan provinsi. Adapun langkah *preprocessing* meliputi:

a. *Data Cleansing*

Pada tahap data *cleaning*, dilakukan beberapa proses untuk memastikan kualitas data. Pertama, dilakukan pengecekan nilai kosong pada seluruh atribut dataset, dan data dengan nilai kosong pada atribut nomor pesanan dihapus karena tidak dapat digunakan dalam proses *clustering*. Kedua, data duplikat yang merepresentasikan transaksi sama dihapus. Ketiga, dilakukan pemilihan fitur kolom yang digunakan, yaitu nama produk, tanggal, nama provinsi, jumlah transaksi, dan total pembayaran. Keempat, kolom tanggal diubah ke format bulan.

b. *Data Integration*

Data yang digunakan dalam penelitian ini bersumber dari satu file ekspor Shopee yang sudah mencakup seluruh informasi yang diperlukan, sehingga tidak diperlukan proses penggabungan data dari sumber yang berbeda.

c. *Data Transformation*

Kemudian pada tahap *data transformation*, dilakukan beberapa penyesuaian agar data dapat digunakan dalam pemodelan K-Means. Pertama, data kategorikal pada atribut nama produk dan provinsi diubah kedalam bentuk numerik menggunakan teknik *encoding*. Terakhir, dilakukan proses normalisasi menggunakan metode *Min-Max* terhadap atribut jumlah transaksi, total pembayaran, bulan, nama produk dan provinsi yang sudah dikonversi menjadi numerik. Langkah ini dilakukan agar semua atribut berada dalam skala yang sama dan tidak mendominasi proses klusterisasi.

3.4 Penentuan Data Latih dan Data Uji

Penelitian ini menggunakan algoritma K-Means *Clustering* yang merupakan bagian dari *unsupervised learning*. Karena tidak terdapat variabel target atau label, seluruh data yang tersedia digunakan langsung dalam proses pemodelan tanpa pemisahan antara data latih dan data uji seperti pada *supervised learning*.

3.5 Pengembangan Model

Pengembangan model dilakukan dengan menerapkan algoritma K-Means *Clustering* menggunakan Google Colaboratory dan RapidMiner. Jumlah kluster optimal ditentukan melalui *Elbow Method* dengan menghitung nilai *Within Cluster Sum of Squares* (WCSS) pada berbagai nilai k, lalu memilih titik di mana penurunan WCSS mulai melandai. Setelah jumlah kluster ditentukan, proses klusterisasi dijalankan di RapidMiner menggunakan operator K-Means dengan konfigurasi parameter yang sesuai, termasuk jumlah kluster dan inisialisasi *centroid*. Seluruh data yang telah dinormalisasi digunakan langsung karena K-Means tidak memerlukan pemisahan data latih dan data uji. Hasil klusterisasi kemudian dianalisis untuk mengidentifikasi karakteristik dominan tiap kluster berdasarkan dimensi waktu, produk, dan wilayah, sehingga dapat memberikan wawasan yang dapat diinterpretasikan oleh GWMM Shop.

3.6 Teknik Pengujian

Teknik pengujian dalam penelitian ini menggunakan pendekatan gabungan antara *Elbow Method* dan *Davies-Bouldin Index* (DBI) untuk menentukan jumlah kluster yang paling optimal. *Elbow Method* digunakan untuk mendapatkan estimasi awal jumlah kluster dengan memvisualisasikan titik siku pada grafik hubungan antara nilai k dan WCSS, yang kemudian menjadi acuan awal jumlah kluster yang diuji lebih lanjut. Hasil estimasi tersebut divalidasi secara kuantitatif menggunakan DBI, yang mengevaluasi kualitas kluster dengan mengukur rasio antara kepadatan di dalam kluster dan pemisahannya dengan kluster lain, sehingga jumlah kluster dengan nilai DBI terendah dipilih sebagai solusi akhir. Selain pengujian kualitas kluster, dilakukan pula perhitungan manual algoritma K-Means menggunakan minimal 5 data sampel dari dataset yang telah dinormalisasi. Perhitungan manual ini mencakup penetapan centroid awal, perhitungan jarak Euclidean setiap data terhadap masing-masing centroid, pengelompokan data ke kluster terdekat, dan pembaruan nilai centroid hingga konvergensi tercapai. Hasil perhitungan manual kemudian dibandingkan dengan *output* RapidMiner untuk memastikan konsistensi proses klusterisasi.

BAB IV ANALISIS DAN PEMBAHASAN

4.1 Pengumpulan Data

Penelitian ini menggunakan data transaksi penjualan nyata dari toko GWMM SHOP, sebuah toko yang berjualan produk parfum di platform *e-commerce* Shopee. Data diperoleh langsung melalui fitur ekspor data pada *dashboard* penjual (*Seller Center*) Shopee, di mana pemilik toko dapat mengunduh riwayat pesanan dalam format Excel (.xlsx). Proses pengunduhan dilakukan secara manual oleh peneliti dengan masuk ke akun penjual GWMM SHOP dan mengekspor seluruh data transaksi berstatus Selesai dalam periode yang telah ditentukan.

Dataset yang diperoleh terdiri dari 4.163 baris data dengan 42 atribut dengan rentang waktu dari bulan Januari 2024 sampai Desember 2025. Atribut tersebut mencakup informasi tiap transaksi penjualan seperti nomor pesanan, nama produk, jumlah pembelian, metode pembayaran, dan lokasi pengiriman pelanggan. Data ini merupakan data primer yang bersumber langsung dari aktivitas bisnis nyata GWMM SHOP, sehingga mencerminkan pola transaksi dan perilaku pembelian pelanggan yang aktual. Adapun atribut yang terdapat dalam dataset GWMM SHOP sebagai berikut:

Tabel 4.1 Atribut Dataset GWMM SHOP

No.	Nama Atribut	Keterangan
1.	No. Pesanan	Nomor unik identifikasi setiap pesanan di Shopee
2.	Status Pesanan	Status akhir pesanan (Selesai, Dibatalkan, dll.)
3.	No. Resi	Nomor resi pengiriman dari ekspedisi
4.	Opsi Pengiriman	Jenis layanan pengiriman yang dipilih pembeli
5.	Antar Ke Counter/Pick up	Metode pengambilan paket oleh kurir
6.	Metode Pembayaran	Cara pembayaran yang digunakan (SPayLater, COD, Online Payment, dll.)
7.	SKU Induk	Kode stok produk (tidak terisi pada dataset ini)
8.	Nama Produk	Nama lengkap produk yang dibeli
9.	Jenis Produk	Kategori jenis produk
10.	Harga Awal	Harga produk sebelum diskon (Rp)
12.	Harga Setelah Diskon	Harga produk setelah diskon diterapkan (Rp)
13.	Jumlah	Kuantitas produk yang dipesan
14.	Returned Quantity	Jumlah produk yang dikembalikan
15.	Total Harga Produk	Total nilai harga produk dalam satu pesanan (Rp)
16.	Total Diskon	Total potongan diskon yang diperoleh pembeli (Rp)
17.	Diskon Dari Penjual	Potongan diskon yang ditanggung penjual (Rp)

18.	Diskon Dari Shopee	Potongan diskon yang ditanggung Shopee (Rp)
19.	Berat Produk	Berat produk dalam satuan gram
20.	Jumlah Produk di Pesan	Total unit produk dalam satu nota pesanan
21.	Total Berat	Total berat keseluruhan produk dalam pesanan
22.	Voucher Ditanggung Penjual	Nilai voucher yang menjadi beban penjual (Rp)
23.	Cashback Koin	Jumlah koin Shopee yang dikembalikan ke pembeli
24.	Voucher Ditanggung Shopee	Nilai voucher yang ditanggung Shopee (Rp)
25.	Paket Diskon	Keterangan paket diskon yang berlaku
26.	Paket Diskon (Diskon dari Shopee)	Nilai diskon Shopee dari paket bundling (Rp)
27.	Paket Diskon (Diskon dari Penjual)	Nilai diskon penjual dari paket bundling (Rp)
28.	Potongan Koin Shopee	Jumlah koin Shopee yang digunakan sebagai pembayaran (Rp)
29.	Diskon Kartu Kredit	Potongan khusus kartu kredit tertentu (Rp)
30.	Ongkos Kirim Dibayar Pembeli	Biaya pengiriman yang dibayar oleh pembeli (Rp)
31.	Estimasi Potongan Biaya Pengiriman	Estimasi subsidi ongkir yang diberikan platform (Rp)
32.	Ongkos Kirim Pengembalian Barang	Biaya ongkir untuk proses retur produk (Rp)
33.	Total Pembayaran	Total nilai yang dibayarkan oleh pembeli (Rp)
34.	Perkiraan Ongkos Kirim	Estimasi total biaya pengiriman (Rp)
35.	Catatan dari Pembeli	Pesan atau instruksi khusus dari pembeli
36.	Catatan	Catatan tambahan dari sistem Shopee
37.	Username (Pembeli)	Nama pengguna akun Shopee pembeli
38.	Nama Penerima	Nama lengkap penerima paket
39.	No. Telepon	Nomor telepon penerima
40.	Alamat Pengiriman	Alamat lengkap tujuan pengiriman
41.	Kota/Kabupaten	Kota atau kabupaten tujuan pengiriman
42.	Provinsi	Provinsi tujuan pengiriman

Berikut pada Gambar 4.1 adalah contoh data transaksi awal dari GWMM SHOP sebelum dilakukan proses pembersihan dan transformasi.

No. Baris	Detail Pengiriman	Detail Pesanan	Detail Produk	Detail Harga
1	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
2	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
3	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
4	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
5	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
6	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
7	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
8	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
9	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
10	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
11	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
12	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
13	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
14	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
15	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
16	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
17	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
18	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
19	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
20	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
21	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
22	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
23	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
24	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
25	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
26	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
27	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
28	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
29	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
30	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
31	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
32	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
33	Selesai	501138050940482	Same Day-Godend Same Day	Pickup
34	Selesai	501138050940482	Same Day-Godend Same Day	Pickup

Gambar 4.1 Dataset Mentah GWMM SHOP

4.2 Data Preprocessing / Pra-Pemrosesan Data

Sebelum data diolah lebih lanjut, diperlukan tahap pra-pemrosesan data agar hasil analisis *data mining* lebih optimal. Pada penelitian ini, pra-pemrosesan mencakup tiga proses utama yaitu *data cleaning* terhadap data transaksi penjualan GWMM SHOP, *data integration*, dan transformasi data. Ketiga proses ini dilakukan untuk memastikan data siap digunakan pada tahap pemodelan klusterisasi menggunakan algoritma *K-Means*.

4.2.1 Data Cleaning

Tahap pertama dalam preprocessing adalah pembersihan data (*data cleaning*). Langkah ini mencakup dua sub proses utama, yaitu penghapusan data duplikat, penanganan nilai yang hilang (*missing values*).

a. Penghapusan Data Duplikat

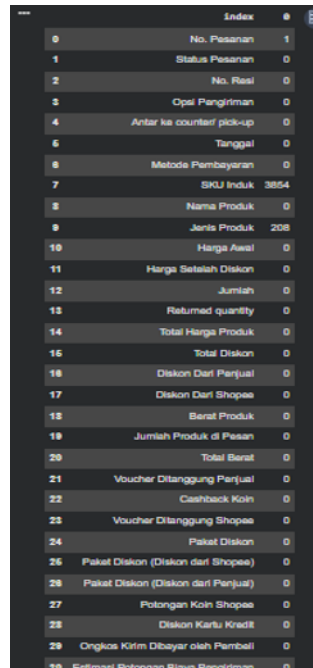
Pada tahap ini dilakukan dilakukan terhadap dataset transaksi penjualan GWMM SHOP yang diekspor dari platform Shopee. Dataset awal terdiri dari 4.163 baris dengan 42 atribut. Pengecekan duplikasi data dilakukan berdasarkan kolom No. Pesanan dan ditemukan sebanyak 309 baris duplikat. Setelah baris duplikat dihapus, jumlah data bersih yang tersisa menjadi 3.854 baris seperti pada Tabel 4.2.

Tabel 4.2 Hasil Penghapusan Data Duplikat

Kondisi	Jumlah Baris	Keterangan
Sebelum penghapusan data duplikat	4.163 baris	Data Mentah Awal
Baris duplikat yang dihapus	309 baris	Teridentifikasi duplikat berdasarkan No. Pesanan
Setelah penghapusan data duplikat	3.854 baris	Data siap diproses lebih lanjut

b. Penanganan *Missing Values* / Nilai Kosong

Pengecekan *missing values* dilakukan pada seluruh kolom dataset. Hasilnya, kolom SKU Induk ditemukan kosong seluruhnya dengan 3.854 missing values karena kolom ini memang tidak digunakan oleh GWMM SHOP dalam pengelolaan toko. Kolom Jenis Produk juga memiliki nilai kosong sebanyak 208 baris. Ringkasan kondisi missing values pada kolom utama disajikan pada Gambar 4.2 sebagai berikut.



Index		
0	No. Pesanan	1
1	Status Pesanan	0
2	No. Resi	0
3	Opsi Pengiriman	0
4	Antar ke counter/pick-up	0
5	Tanggal	0
6	Metode Pembayaran	0
7	SKU Induk	3854
8	Nama Produk	0
9	Jenis Produk	208
10	Harga Awal	0
11	Harga Setelah Diskon	0
12	Jumlah	0
13	Returned quantity	0
14	Total Harga Produk	0
15	Total Diskon	0
16	Diskon Dari Penjual	0
17	Diskon Dari Shopee	0
18	Berat Produk	0
19	Jumlah Produk di Pesan	0
20	Total Berat	0
21	Voucher Ditunggung Penjual	0
22	Cashback Koin	0
23	Voucher Ditunggung Shopee	0
24	Paket Diskon	0
25	Paket Diskon (Diskon dari Shopee)	0
26	Paket Diskon (Diskon dari Penjual)	0
27	Potongan Koin Shopee	0
28	Diskon Kartu Kredit	0
29	Ongkos Kirim Dibayar oleh Pembeli	0
30	Estimasi Potongan Biaya Pengiriman	0

Gambar 4.2 Hasil Pengecekan *Missing Values*

Berdasarkan hasil pengecekan tersebut Kolom SKU Induk dihapus karena tidak memiliki data sama sekali dan tidak relevan untuk analisis. Kolom lain yang tidak relevan untuk klasterisasi, seperti No. Resi, Opsi Pengiriman, Antar ke *counter/pick-up*, dan sejumlah atribut diskon yang tidak signifikan, juga ikut dihapus. Penghapusan atribut-atribut ini dilakukan untuk menyederhanakan dataset agar lebih fokus pada fitur yang merepresentasikan pola transaksi dan perilaku pelanggan GWMM SHOP.

c. Standarisasi Kolom Provinsi

Dalam penelitian ini, standarisasi nama provinsi dilakukan dengan membuat kamus acuan yang merujuk pada 34 nama provinsi resmi sesuai pembagian wilayah administratif Indonesia. Setiap variasi penulisan disesuaikan ke bentuk baku, di mana nama "Jakarta" dalam berbagai variasinya distandarisasi menjadi "DKI JAKARTA" sesuai nama resmi provinsi tersebut. Pendekatan ini sejalan dengan prinsip data cleaning untuk data kategorikal, yaitu nilai yang tidak konsisten perlu diperiksa lebih dulu melalui pengecekan nilai unik, kemudian

digantikan atau dikelompokkan sesuai kategori yang seharusnya. Hal serupa juga berlaku pada data yang digunakan dalam penelitian ini. Setelah dilakukan pemeriksaan terhadap 4.163 baris data, field Provinsi ditemukan memiliki 15 nilai unik yang seluruhnya sudah konsisten dengan format penulisan resmi, termasuk wilayah DKI Jakarta yang dalam dataset ini sudah tercatat secara baku sebagai "DKI JAKARTA".

d. Pemilihan Fitur Kolom

Dataset awal GWMM SHOP memiliki 42 atribut yang mencakup berbagai informasi transaksi, mulai dari data operasional pengiriman, informasi pembeli, rincian diskon dan voucher, hingga data produk dan lokasi. Tidak semua *atribut* tersebut relevan untuk tujuan klasterisasi pola penjualan yang menjadi fokus penelitian ini. Adapapun atribut yang terpilih untuk digunakan dalam analisis meliputi kolom tanggal, Nama Produk, Provinsi, Jumlah, Total Harga Produk seperti yang ditampilkan pada Gambar 4.3.

	Tanggal	Nama Produk	Provinsi	Jumlah	Total Harga Produk
0	2025-03-15	FYNE Extrait De Parfume	NUSA TENGGARA TIMUR (NTT)	1	89000
1	2025-03-06	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	NUSA TENGGARA TIMUR (NTT)	2	218000
2	2024-12-14	Fragrance World Cocktail Intense Unisex Eau De...	BANTEN	4	1064000
3	2024-12-02	Verbena Night Legend Man Eau De Parfum 100 ML	JAWA BARAT	5	610000
4	2025-07-22	BRODO - Rote EDT Parfume 40ml	BALI	2	198000

Gambar 4.3 Pemilihan Fitur Kolom Terpakai

4.2.2 Data Integration

Pada tahap ini, *Data integration* secara umum dilakukan dengan menggabungkan data dari berbagai sumber berbeda menjadi satu dataset yang utuh. Dalam penelitian ini, proses tersebut tidak perlu dilakukan karena data sudah tersedia dalam satu file yang diekspor langsung dari *Seller Center Shopee*. File tersebut sudah mencakup seluruh informasi transaksi yang dibutuhkan, mulai dari data produk, jumlah pembelian, total harga, hingga lokasi pengiriman. Karena semua informasi sudah tergabung dalam satu sumber, dataset langsung bisa diproses pada tahap transformasi berikutnya.

4.2.3 Transformasi Data

Transformasi data adalah tahap dalam *preprocessing* yang mengubah data mentah menjadi format yang relevan untuk dianalisis. Pada penelitian ini, transformasi mencakup beberapa langkah. Pertama, Kolom tanggal diubah menjadi representasi bulan dalam bentuk angka agar lebih mudah dianalisis. Kolom nama produk dan provinsi yang awalnya berbentuk kategorikal dikonversi ke format numerik menggunakan metode *encoding*, karena algoritma K-Means hanya dapat memproses data dalam bentuk angka. Setelah encoding, dilakukan normalisasi data menggunakan *Min-Max Normalization* untuk menyetarakan skala antar atribut, sehingga

atribut dengan nilai besar tidak mendominasi perhitungan jarak dalam proses klusterisasi.

a. Ekstraksi Fitur Waktu

Pada tahap awal transformasi data, mengubah tanggal menjadi Bulan yang diperoleh dari kolom Tanggal. Kolom ini menghasilkan angka bulan dalam format numerik, seperti 1 untuk Januari, 2 untuk Februari, dan seterusnya hingga 12 untuk Desember. Penambahan atribut ini bertujuan agar analisis pola penjualan bulanan toko GWMM SHOP dapat dilakukan. Informasi bulan perlu diikutsertakan karena volume transaksi bisa berbeda antar bulan.

Tabel 4.3 Contoh Ekstraksi Fitur Waktu

Tanggal	Bulan
01-01-2024 / 01-01-2025	1
01-02-2024 / 01-02-2025	2
01-03-2024 / 01-03-2025	3
01-04-2024 / 01-04-2025	4
01-05-2024 / 01-05-2025	5
01-06-2024 / 01-06-2025	6
01-07-2024 / 01-07-2025	7
01-08-2024 / 01-08-2025	8
01-09-2024 / 01-09-2025	9
01-10-2024 / 01-10-2025	10
01-11-2024 / 01-11-2025	11
01-12-2024 / 01-12-2025	12

b. Mengubah Data Kategorikal Menjadi Numerikal

Tahap ini mencakup pengubahan seluruh atribut yang digunakan dalam klusterisasi ke format *numerik*. Proses dimulai dari kolom Nama Produk yang sebelumnya berisi data teks, kemudian setiap nama produk parfum diganti dengan nilai numerik seperti yang ditampilkan pada Tabel 4.4.

Tabel 4.4 Transformasi Kategori Nama Produk

Nama Produk	Nilai Numerik
(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0
Aksesoris Gantungan Parfum	1
BRODO - Rote EDT Perfume 40ml	2
Crusita - Busy Day Blooming Series Extrait De Perfume	3
Dynamique Course Pour Homme Men EDP 100 ML	4
FYNE Extrait De Perfume	5
Fogg Perfume Midnight Purple 100ml Eau De Perfume	6

Fogg Perfume Scent Premium Many Flowers 100ml Eau De Perfume	7
Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	8
Fragrance World Suits Unisex Eau De Parfum 100 ML	9
GARVI Perfume - Naughty Friday - EDP 50ml	10
HINT IDENTITY Eau de Parfum 50ml	11
HINT Noble Eau de Parfum 50ml	12
ICE FROST Extrait De Perfume 35ml	13
Kokulu Women Best Seller - Bundling 2pcs 50ML	14
Mezuka Perfume Aqua Kiss Eau de Perfume 50ml	15
Mykonos Baby Love Parfum EDP 50 ML	16
Mykonos Milk Drips 50ml	17
Mykonos Verve Extrait De Parfum 50 ML	18
NARSIS Hypnotic Desire Eau de Perfume 50ml	19
Omerta Pink Ice Woman Eau de Parfum 100 ML	20
Omerta The Winner Perfume Pria Eau de Toilette 100 ML	21
Onix Fragrance Eau de Perfume Senoparty 30ML	22
Pouch Parfum Mini	23
Refill Parfum 30ml	24
Revlon Charlie Fragrance - White EDT Parfum 100 ml	25
Russ Perfume Unomento Blue Unisex 50ml	26
SAFF & Co. Extrait de Parfum - CASCABEL 30ml	27
Sample Pack 3ml x5	28
Scentplus Dominant 35ml	29
Verbena Night Legend Man Eau De Parfum 100 ML	30
Vial Parfum 35ml	31
Xchange Extrait De Perfume Supernova Series 50ml - Terra Nova	32

Selain itu, pengubahan atribut provinsi juga dilakukan, yang kemudian diubah dari format teks menjadi angka. Konversi ini

bertujuan agar nilai provinsi dapat diproses dalam algoritma K-Means. Berikut hasil konversinya terdapat pada Tabel 4.5.

Tabel 4.5 Transformasi Kategori Provinsi

Provinsi	Nilai Numerik
Bali	0
Banten	1
Bengkulu	2
DKI Jakarta	3
Gorontalo	4
Jawa Barat	5
Jawa Tengah	6
Jawa Timur	7
Kalimantan Timur	8
Maluku	9
Nusa Tenggara Timur (NTT)	10
Papua	11
Riau	12
Sulawesi Selatan	13
Sumatra Utara	14

Pengubahan ke format *numerik* diperlukan karena algoritma K-Means hanya dapat memproses data dalam bentuk angka. Dengan langkah ini, seluruh informasi dari kolom Nama Produk dan Provinsi sudah direpresentasikan secara numerik dan siap untuk dianalisis lebih lanjut.

...	Bulan	Nama Produk_Encoded	Provinsi_Encoded	Jumlah	Total Harga Produk	...
0	3	5	10	1	89000	
1	3	15	10	2	218000	
2	12	8	1	4	1064000	
3	12	30	5	5	610000	
4	7	2	0	2	198000	
...	...	...	...	...	...	...
204	12	23	10	1	12000	
205	3	31	10	1	5000	
206	12	20	3	4	536400	
207	12	12	3	3	399000	
208	12	0	1	4	1396000	

Gambar 4.4 Dataset Setelah Dilakukan Transformasi Numerik

c. *Data Normalisasi*

Setelah proses mengubah nilai kategorikal menjadi numerikal, semua atribut numerik yang dipakai dalam pemodelan, yaitu Nama Produk_Encoded, Provinsi_Encoded, Jumlah, dan Total Harga Produk,

harus dinormalisasi terlebih dahulu sebelum masuk ke tahap berikutnya. Alasannya sederhana, skala nilai antar atribut terlampau jauh berbeda. Atribut Total Harga Produk dinyatakan dalam satuan rupiah yang nilainya bisa mencapai jutaan bahkan miliaran, sedangkan atribut Jumlah dan hasil pengubahan atribut kategorikal menjadi numerikal berada dalam rentang yang jauh lebih kecil. Kondisi ini bermasalah karena algoritma K-Means menghitung kedekatan antardata menggunakan jarak Euclidean. Apabila satu atribut memiliki nilai yang jauh lebih besar, atribut tersebut akan otomatis mendominasi perhitungan, sementara atribut lain nyaris tidak memberi pengaruh apa pun pada hasil klusterisasi. Akibatnya, klaster yang terbentuk tidak mencerminkan keseluruhan karakteristik data. Untuk mengatasi masalah ini, digunakan metode Min-Max Normalization yang mengubah semua nilai atribut ke dalam rentang 0 hingga 1 dengan rumus:

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Keterangan:

x' = Nilai data baru setelah dinormalisasi.

x_i = Nilai data asli yang sedang diproses.

$\min(x)$ = Nilai minimum dari suatu atribut.

$\max(x)$ = Nilai maksimal dari suatu atribut.

Maka dapat diketahui dengan rumus Min-Max pada data nama produk GWMM SHOP sebagai berikut:

$$x' = \frac{5 - 0}{32 - 0} = 0,15625$$

Normalisasi dilakukan seterusnya sampai 3.854 *record* dari setiap atribut Nama Produk, Provinsi, Jumlah dan Total Harga Produk.

a. Bulan Normalisasi

Pada Tabel 4.6 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Bulan.

Tabel 4.6 Bulan Normalisasi

Bulan	Nilai Normalisasi
1	0
2	0.091
3	0.182
4	0.273
5	0.364
6	0.455
7	0.545

8	0.636
9	0.727
10	0.818
11	0.909
12	1

b. Data Nama Produk

Pada Tabel 4.7 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Nama Produk.

Tabel 4.7 Hasil Normalisasi Nama Produk

Nama Produk	Nilai Numerik
(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0
Aksesoris Gantungan Parfum	0,03125
BRODO - Rote EDT Perfume 40ml	0,0625
Crusita - Busy Day Blooming Series Extrait De Perfume	0,09375
Dynamique Course Pour Homme Men EDP 100 ML	0,125
FYNE Extrait De Perfume	0,15625
Fogg Perfume Midnight Purple 100ml Eau De Perfume	0,1875
Fogg Perfume Scent Premium Many Flowers 100ml Eau De Perfume	0,21875
Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	0,25
Fragrance World Suits Unisex Eau De Parfum 100 ML	0,28125
GARVI Perfume - Naughty Friday - EDP 50ml	0,3125
HINT IDENTITY Eau de Parfum 50ml	0,34375
HINT Noble Eau de Parfum 50ml	0,375
ICE FROST Extrait De Perfume 35ml	0,40625
Kokulu Women Best Seller - Bundling 2pcs 50ML	0,4375
Mezuca Perfume Aqua Kiss Eau de Perfume 50ml	0,46875
Mykonos Baby Love Parfum EDP 50 ML	0,5
Mykonos Milk Drips 50ml	0,53125
Mykonos Verve Extrait De Parfum 50 ML	0,5625
NARSIS Hypnotic Desire Eau de Perfume 50ml	0,59375

Omerta Pink Ice Woman Eau de Parfum 100 ML	0,625
Omerta The Winner Parfume Pria Eau de Toilette 100 ML	0,65625
Onix Fragrance Eau de Perfume Senoparty 30Ml	0,6875
Pouch Parfum Mini	0,71875
Refill Parfum 30ml	0,75
Revlon Charlie Fragrance - White EDT Parfum 100 ml	0,78125
Russ Parfume Unomento Blue Unisex 50ml	0,8125
SAFF & Co. Extrait de Parfum - CASCAVEL 30ml	0,84375
Sample Pack 3ml x5	0,875
Scentplus Dominant 35ml	0,90625
Verbena Night Legend Man Eau De Parfum 100 ML	0,9375
Vial Parfum 35ml	0,96875
Xchange Extrait De Parfume Supernova Series 50ml - Terra Nova	1

c. Data Provinsi

Pada Tabel 4.8 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Provinsi.

Tabel 4.8 Hasil Normalisasi Provinsi

Provinsi	Nilai Numerik
Bali	0
Banten	0,071428571
Bengkulu	0,142857143
DKI Jakarta	0,214285714
Gorontalo	0,285714286
Jawa Barat	0,357142857
Jawa Tengah	0,428571429
Jawa Timur	0,5
Kalimantan Timur	0,571428571
Maluku	0,642857143
Nusa Tenggara Timur (NTT)	0,714285714
Papua	0,785714286
Riau	0,857142857
Sulawesi Selatan	0,928571429
Sumatra Utara	1

d. Data Jumlah

Pada Tabel 4.9 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Jumlah.

Tabel 4.9 Hasil Normalisasi Jumlah

Bulan	Nama Produk	Jumlah
3	FYNE Extrait De Parfume	0.000000
3	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	0.166667
12	Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	0.500000
12	Verbena Night Legend Man Eau De Parfum 100 ML	0.666667
7	BRODO - Rote EDT Parfume 40ml	0.166667
.....		
12	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0.500000
7	GARVI Parfume - Naughty Friday - EDP 50ml	0.000000
7	HINT IDENTITY Eau de Parfum 50ml	1.000000
12	Russ Parfume Unomento Blue Unisex 50ml	0.071429
12	SAFF & Co. Extrait de Parfum - CASCAVEL 30ml	0.071429

e. Data Total Harga Produk

Pada Tabel 4.10 terlihat hasil normalisasi data menggunakan metode Min-Max pada atribut Total Harga Produk.

Tabel 4. 10 Hasil Normalisasi Total Harga Produk

Bulan	Nama Produk	Total Harga Produk
3	FYNE Extrait De Parfume	0.034454
3	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	0.087367
12	Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	0.434372
12	Verbena Night Legend Man Eau De Parfum 100 ML	0.248154
7	BRODO - Rote EDT Parfume 40ml	0.079163
.....		
12	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	0.713700
7	GARVI Parfume - Naughty Friday - EDP 50ml	0.087367

7	HINT IDENTITY Eau de Parfum 50ml	0.078425
12	Russ Parfume Unomento Blue Unisex 50ml	0.132075
12	SAFF & Co. Extrait de Parfum - CASCABEL 30ml	0.364643

Maka dapat diketahui nilai data pada dataset GWMM SHOP setelah dilakukan normalisasi pada gambar di bawah ini.

	Bulan	Nama Produk_Encoded	Provinsi_Encoded	Jumlah	Total Harga Produk
0	0.181818	0.15625	0.714286	0.000000	0.034454
1	0.181818	0.46875	0.714286	0.166667	0.087367
2	1.000000	0.25000	0.071429	0.500000	0.434372
3	1.000000	0.93750	0.357143	0.666667	0.248154
4	0.545455	0.06250	0.000000	0.166667	0.079163
...	...	...	...	...	...
204	1.000000	0.71875	0.714286	0.000000	0.002871
205	0.181818	0.96875	0.714286	0.000000	0.000000
206	1.000000	0.62500	0.214286	0.500000	0.217966
207	1.000000	0.37500	0.214286	0.333333	0.161608
208	1.000000	0.00000	0.071429	0.500000	0.570550

Gambar 4.5 Dataset Setelah Dilakukan Normalisasi

4.3 Pemodelan

4.3.1 Hasil Komparasi Model

Untuk menentukan jumlah kluster yang paling optimal, dilakukan komparasi model berdasarkan nilai *Davies-Bouldin Index* (DBI). Nilai DBI yang rendah menandakan kualitas klusterisasi yang baik karena jarak antarkluster cukup jauh dan data di dalam setiap kluster terbilang kompak. Pada penelitian ini, dua algoritma dibandingkan, yaitu K-Means dan K-Medoids, masing-masing diuji dengan berbagai variasi jumlah kluster. Seluruh proses pengujian dijalankan menggunakan RapidMiner, dan hasilnya ditampilkan pada tabel berikut sebagai dasar pemilihan model terbaik.

a. Pemodelan Dengan Algoritma *K-Means*

Tahap pemodelan pertama menggunakan algoritma K-Means dengan menguji 4 variasi jumlah kluster, yaitu 2, 3, 4, dan 5 kluster. Nilai DBI dari setiap variasi tersebut ditampilkan pada Tabel 4.11 berikut.

Tabel 4.11 Hasil DBI pada Algoritma *K-Means*

Jumlah Cluster	Nilai DBI
<i>Cluster 2</i>	1.097
<i>Cluster 3</i>	1.146
<i>Cluster 4</i>	1.002
<i>Cluster 5</i>	1.077

Berdasarkan hasil pengujian, nilai DBI terendah diperoleh pada jumlah klaster 4, yaitu sebesar 1,002. Nilai DBI yang paling kecil mencerminkan kualitas klasterisasi terbaik, sehingga model K-Means bekerja paling optimal dengan 4 klaster.

b. Pemodelan Dengan Algoritma *K-Medoids*

Pemodelan berikutnya dilakukan menggunakan algoritma K-Medoids dengan variasi jumlah klaster yang sama, yaitu 2, 3, 4, dan 5 klaster. Nilai DBI dari setiap variasi tersebut ditampilkan pada Tabel 4.12 berikut.

Tabel 4.12 Hasil DBI pada Algoritma *K-medoids*

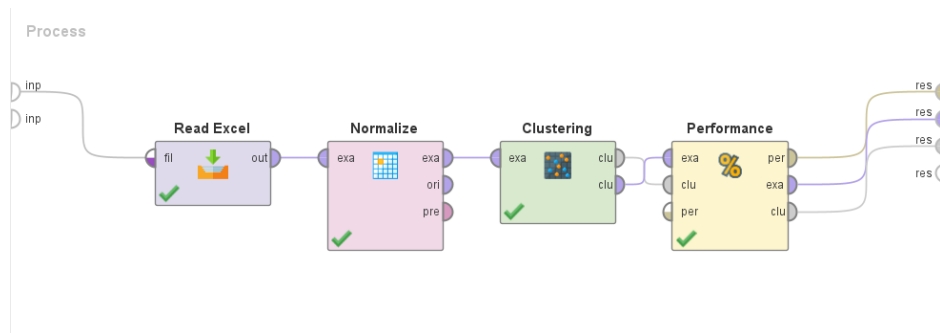
Jumlah Cluster	Nilai DBI
<i>Cluster 2</i>	1.993
<i>Cluster 3</i>	1.113
<i>Cluster 4</i>	1.079
<i>Cluster 5</i>	1.114

Berdasarkan Tabel 4.12, nilai DBI terendah pada algoritma K-Medoids diperoleh pada jumlah klaster 4, yaitu sebesar 1,079. Namun angka ini masih lebih tinggi dibandingkan seluruh nilai DBI yang dihasilkan oleh algoritma K-Means.

Setelah membandingkan kedua algoritma, K-Means dengan 4 klaster menghasilkan nilai DBI paling rendah, yaitu 1,002. Oleh karena itu, algoritma K-Means dipilih sebagai metode klasterisasi terbaik dalam penelitian ini dengan jumlah klaster optimal sebanyak 4 klaster.

4.3.2 Penyajian Model Terbaik

Berdasarkan hasil evaluasi, algoritma K-Means dengan 4 klaster terpilih sebagai model terbaik dibandingkan K-Medoids. Pemilihan ini didasarkan pada nilai Davies-Bouldin Index (DBI) terendah sebesar 1,002, yang menunjukkan bahwa pemisahan antarklaster cukup baik dan data di dalam setiap klaster cukup kompak. Proses klasterisasi sendiri dilakukan pada data yang sudah melewati tahap normalisasi menggunakan metode Min-Max Normalization. Atribut yang digunakan dalam pemodelan ini ada lima, yaitu Bulan, Jumlah, Total Harga Produk, Nama Produk yang sudah dikonversi ke bentuk numerik, dan Provinsi yang sudah dikodekan ke bentuk numerik. Alur proses pemodelan K-Means *Clustering* pada RapidMiner ditampilkan pada Gambar 4.6 berikut.



Gambar 4.6 Penyajian Model dengan RapidMiner

Setelah proses klasterisasi dilakukan, diperoleh hasil distribusi data ke dalam empat *cluster*, dengan rincian Gambar 4.7 sebagai berikut:

Cluster Model

```
Cluster 0: 1203 items
Cluster 1: 1449 items
Cluster 2: 652 items
Cluster 3: 550 items
Total number of items: 3854
```

Gambar 4.7 Hasil Jumlah Cluster

Untuk memahami karakteristik tiap klaster yang terbentuk, dilakukan analisis deskriptif berdasarkan nilai minimum, rata-rata, dan maksimum dari setiap atribut yang digunakan dalam proses klasterisasi. Hasilnya disajikan pada Tabel 4.13 berikut.

Tabel 4.13 Data Jumlah Minimal, Maksimal, dan Rata-rata Pada Setiap Atribut Data

		Bulan	Nama Produk_Encoded	Provinsi_Encoded	Jumlah	Total Harga Produk
Cluster 0	Min	1	0	7	1	10.000
	Rata-Rata	6,089	9,389	11,522	1,513	162.596,010
	Max	12	25	14	4	1.396.000
Cluster 1	Min	2	0	0	1	12.000
	Rata-Rata	11,619	18,144	4,346	4,511	725.445,687
	Max	12	32	8	7	2.443.000
	Min	1	0	0	1	10.000

<i>Cluster</i> 2	Rata-Rata	5,678	11,553	1,045	1,789	234.784,909
	Max	12	29	7	6	1.745.000
<i>Cluster</i> 3	Min	1	14	2	1	5.000
	Rata-Rata	3,997	27,066	8,862	1,417	93.116,564
	Max	12	32	14	7	894.000

Berdasarkan hasil pemodelan K-Means dengan $k=4$, terbentuk empat klaster dari total 3.854 data transaksi. Setiap klaster dibentuk berdasarkan atribut Bulan, Nama Produk_Encoded, Provinsi_Encoded, Jumlah, dan Total Harga Produk yang sudah melalui proses transformasi dan normalisasi. Analisis terhadap keempat klaster ini dilakukan untuk menggambarkan pola penjualan GWMM SHOP, sebagaimana diuraikan berikut ini.

a. *Cluster* 0

Klaster 0 berisi 1.203 data transaksi dengan rata-rata bulan sebesar 6,089, yang menunjukkan transaksi pada klaster ini terkonsentrasi di bulan Juni hingga September. Rata-rata jumlah produk per transaksi tercatat sebesar 1,513 unit dengan rata-rata total harga transaksi sebesar Rp162.596. Produk yang paling banyak muncul pada klaster ini adalah Mykonos Baby Love Parfum EDP 50 ML dan Fyne Extrait De Parfume, yaitu parfum segmen mid-range yang menjadi pilihan utama konsumen di luar musim belanja besar. Provinsi dengan transaksi terbanyak pada klaster ini adalah Sumatera Utara, yang mengindikasikan bahwa pasar di luar Pulau Jawa tetap aktif sepanjang pertengahan tahun.

b. *Cluster* 1

Klaster 1 memiliki jumlah data terbanyak, yaitu 1.449 transaksi, dengan rata-rata total harga produk tertinggi sebesar Rp725.445 per transaksi. Rata-rata bulan sebesar 11,619 menunjukkan klaster ini didominasi transaksi pada bulan Oktober hingga Desember, yang bertepatan dengan momen Harbolnas, Natal, dan Tahun Baru. Rata-rata jumlah produk per transaksi juga tercatat paling tinggi, yaitu 4,511 unit, yang menandakan konsumen cenderung membeli lebih banyak produk sekaligus pada periode tersebut. Provinsi dengan transaksi paling aktif adalah DKI Jakarta, Jawa Barat, dan Jawa Timur, selaras dengan konsentrasi populasi pengguna *e-commerce* terbesar di Indonesia.

c. *Cluster* 2

Klaster 2 berisi 652 data transaksi dengan rata-rata bulan sebesar 5,678, yang menunjukkan transaksi pada klaster ini terkonsentrasi di bulan Mei hingga Juli. Klaster ini mencatat nilai rata-rata total harga produk sedang, yaitu Rp234.784, dengan rata-rata jumlah produk per transaksi paling kecil sebesar 1,789 unit. Produk

yang paling banyak muncul adalah parfum lokal berharga terjangkau seperti Xchange Extrait De Parfume Supernova Series 50ml - Terra Nova dan Sample Pack 3ml x5. Yang menarik, provinsi dengan transaksi terbanyak justru Papua, NTT, dan Kalimantan Timur, yang mengindikasikan adanya segmen pasar aktif di wilayah Indonesia timur meskipun volume transaksinya masih relatif Rendah.

d. Cluster 3

Klaster 3 berisi 550 data transaksi dengan rata-rata bulan sebesar 3,997, yang menunjukkan transaksi pada klaster ini terkonsentrasi di bulan Januari hingga Mei. Klaster ini mencatat nilai rata-rata total harga produk rendah, yaitu Rp93.116, dengan rata-rata jumlah produk per transaksi paling kecil sebesar 1,417 unit. Produk yang paling banyak muncul adalah parfum lokal berharga terjangkau seperti HINT IDENTITY dan ICE FROST Extrait De Parfume 35ml. Yang menarik, provinsi dengan transaksi terbanyak justru Bali, Gorontalo, dan Bengkulu, yang mengindikasikan adanya segmen pasar aktif di wilayah Indonesia timur meskipun volume transaksinya masih sangat rendah.

Tabel 4.14 Karakteristik Setiap Cluster

<i>Cluster</i>	Jumlah	Nama Produk	Provinsi	Total Harga Produk	Bulan
0	Tinggi	Mykonos Baby Love	Sumatra Utara	Rp162.596	7-9
1	Sangat Tinggi	Hint Noble	DKI Jakarta	Rp725.445	10-12
2	Sedang	Xchange Extrait De Parfume Supernova Series 50ml	Papua	Rp234.784	5-7
3	Rendah	HINT IDENTITY Eau De Parfume 50ml	Bali	Rp93.116	1-5

Secara keseluruhan, keempat klaster yang terbentuk menggambarkan empat pola penjualan yang berbeda berdasarkan dimensi waktu, nama produk, jumlah, total harga produk, dan wilayah. Klaster 1 merepresentasikan periode penjualan puncak yang perlu dijadikan prioritas dalam persiapan stok dan strategi promosi. Klaster 0 dan klaster 2 menggambarkan periode penjualan stabil di pertengahan tahun yang dapat dimanfaatkan untuk memperluas kampanye

pemasaran ke pasar luar Jawa. Sementara itu, Klaster 3 mencerminkan periode paling sepi yang justru bisa dijadikan peluang untuk menjalankan program promosi awal tahun guna mendorong peningkatan transaksi.

4.3.3 Perhitungan Algoritma Terpilih

Pada tahapan ini dilakukan proses perhitungan manual menggunakan algoritma yang telah terpilih, yaitu K-Means *Clustering*. Pengolahan data dilakukan setelah data melalui proses transformasi dan normalisasi, sehingga data telah siap untuk dianalisis menggunakan K-Means. Perhitungan dilakukan dengan menggunakan sampel dataset sebanyak 5 data untuk mempermudah penjabaran proses secara manual. Data yang digunakan mencakup atribut Nama Produk (normalisasi), Bulan (normalisasi), Jumlah (normalisasi), Total Harga Produk (normalisasi), dan Provinsi (normalisasi), yang akan diproses lebih lanjut dalam tahap klasterisasi.

Tabel 4.15 Dataset pada Algoritma Terpilih

Nama Produk	Bulan	Nama Produk Norm	Provinsi_Norm	Jumlah_Norm	Total Harga Norm
Mykonos Baby Love Parfum EDP 50ML	0.455	0,167	1,000	0,000	0,056
HINT Noble Eau De Parfume 50ml	1,000	0,375	0,214	0,667	0,271
Xchange Extrait De Parfume Supernova Series 50ml	0.182	1,000	0,786	0,000	0,051
HINT IDENTITY Eau De Parfume 50ml	0,000	0,344	0,000	0,333	0,132
FYNE Extrait De Parfume	0,182	0,156	0,714	0,000	0,034

Setelah data dinormalisasi, proses klasterisasi menggunakan algoritma K-Means dapat dimulai. Tahapan awal yang dilakukan adalah sebagai berikut.

1. Jumlah klaster yang digunakan dalam penelitian ini adalah 4 klaster ($k=4$), sesuai dengan hasil evaluasi DBI yang telah dilakukan pada tahap sebelumnya.
2. Centroid awal setiap klaster ditentukan dengan cara mengambil data secara acak dari sampel yang telah disiapkan sebelumnya. Data centroid tersebut adalah sebagai berikut.

Tabel 4. 16 Titik Pusat Awal Tiap Cluster

Centroid Awal	Bulan	Nama Produk_Norm	Provinsi_Norm	Jumlah_Norm	Total Harga_Norm
C0	0.455	0,167	1,000	0,000	0,056
C1	1,000	0,375	0,214	0,667	0,271
C2	0.182	1,000	0,786	0,000	0,051
C3	0,000	0,344	0,000	0,333	0,132

3. Setelah centroid awal ditentukan, langkah berikutnya adalah menghitung jarak setiap data terhadap tiap centroid menggunakan rumus Euclidean Distance. Perhitungan ini bertujuan untuk mengukur seberapa dekat setiap data terhadap pusat klaster yang telah ditetapkan.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Berikut jarak data 1 dengan setiap centroid :

$$d(1, C0)$$

$$= \sqrt{(0,455 - 0,455)^2 + (0,167 - 0,167)^2 + (1,000 - 1,000)^2 + (0,000 - 0,000)^2 + (0,056 - 0,056)^2}$$

$$= 0,000$$

$$d(1, C1)$$

$$= \sqrt{(0,455 - 1,000)^2 + (0,167 - 0,375)^2 + (1,000 - 0,214)^2 + (0,000 - 0,667)^2 + (0,056 - 0,271)^2}$$

$$= 1,204$$

$$d(1, C2)$$

$$= \sqrt{(0,455 - 0,182)^2 + (0,167 - 1,000)^2 + (1,000 - 0,786)^2 + (0,000 - 0,000)^2 + (0,056 - 0,051)^2}$$

$$= 0,902$$

$$\begin{aligned}
& d(1, C3) \\
&= \sqrt{(0,455 - 0,000)^2 + (0,167 - 0,344)^2 + (1,000 - 0,000)^2 + (0,000 - 0,333)^2 + (0,056 - 0,132)^2} \\
&= 1,164
\end{aligned}$$

Berikut jarak data 2 dengan setiap centroid :

$$\begin{aligned}
& d(2, C0) \\
&= \sqrt{(1,000 - 0,455)^2 + (0,375 - 0,167)^2 + (0,214 - 1,000)^2 + (0,667 - 0,000)^2 + (0,271 - 0,056)^2} \\
&= 1,204
\end{aligned}$$

$$\begin{aligned}
& d(2, C1) \\
&= \sqrt{(1,000 - 1,000)^2 + (0,375 - 0,375)^2 + (0,214 - 0,214)^2 + (0,667 - 0,667)^2 + (0,271 - 0,271)^2} \\
&= 0,000
\end{aligned}$$

$$\begin{aligned}
& d(2, C2) \\
&= \sqrt{(1,000 - 0,182)^2 + (0,375 - 1,000)^2 + (0,214 - 0,786)^2 + (0,667 - 0,000)^2 + (0,271 - 0,051)^2} \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
& d(2, C3) \\
&= \sqrt{(1,000 - 0,000)^2 + (0,375 - 0,344)^2 + (0,214 - 0,000)^2 + (0,667 - 0,333)^2 + (0,271 - 0,132)^2} \\
&= 1,085
\end{aligned}$$

Berikut jarak data 3 dengan setiap centroid :

$$\begin{aligned}
& d(3, C0) \\
&= \sqrt{(0,182 - 0,455)^2 + (1,000 - 0,167)^2 + (0,786 - 1,000)^2 + (0,000 - 0,000)^2 + (0,051 - 0,056)^2} \\
&= 0,902
\end{aligned}$$

$$\begin{aligned}
& d(3, C1) \\
&= \sqrt{(0,182 - 1,000)^2 + (1,000 - 0,375)^2 + (0,786 - 0,214)^2 + (0,000 - 0,667)^2 + (0,051 - 0,271)^2} \\
&= 1,371
\end{aligned}$$

$$\begin{aligned}
& d(3, C2) \\
&= \sqrt{(0,182 - 0,182)^2 + (1,000 - 1,000)^2 + (0,786 - 0,786)^2 + (0,000 - 0,000)^2 + (0,051 - 0,051)^2} \\
&= 0,000
\end{aligned}$$

$$\begin{aligned}
& d(3, C3) \\
&= \sqrt{(0,182 - 0,000)^2 + (1,000 - 0,344)^2 + (0,786 - 0,000)^2 + (0,000 - 0,333)^2 + (0,051 - 0,132)^2} \\
&= 1,095
\end{aligned}$$

Berikut jarak data 4 dengan setiap centroid :

$$\begin{aligned}
& d(4, C0) \\
&= \sqrt{(0,000 - 0,455)^2 + (0,344 - 0,167)^2 + (0,000 - 1,000)^2 + (0,333 - 0,000)^2 + (0,132 - 0,056)^2} \\
&= 1,164
\end{aligned}$$

$$\begin{aligned}
& d(4, C1) \\
&= \sqrt{(0,000 - 1,000)^2 + (0,344 - 0,375)^2 + (0,000 - 0,214)^2 + (0,333 - 0,667)^2 + (0,132 - 0,271)^2} \\
&= 1,085
\end{aligned}$$

$$\begin{aligned}
& d(4, C2) \\
&= \sqrt{(0,000 - 0,182)^2 + (0,344 - 1,000)^2 + (0,000 - 0,786)^2 + (0,333 - 0,000)^2 + (0,132 - 0,051)^2} \\
&= 1,095
\end{aligned}$$

$$\begin{aligned}
& d(4, C3) \\
&= \sqrt{(0,000 - 0,000)^2 + (0,344 - 0,344)^2 + (0,000 - 0,000)^2 + (0,333 - 0,333)^2 + (0,132 - 0,132)^2} \\
&= 0,000
\end{aligned}$$

Berikut jarak data 5 dengan setiap centroid :

$$\begin{aligned}
& d(5, C0) \\
&= \sqrt{(0,182 - 0,455)^2 + (0,156 - 0,167)^2 + (0,714 - 1,000)^2 + (0,000 - 0,000)^2 + (0,034 - 0,056)^2} \\
&= 0,396
\end{aligned}$$

$$\begin{aligned}
& d(5, C1) \\
&= \sqrt{(0,182 - 1,000)^2 + (0,156 - 0,375)^2 + (0,714 - 0,214)^2 + (0,000 - 0,667)^2 + (0,034 - 0,271)^2} \\
&= 1,211
\end{aligned}$$

$$\begin{aligned}
& d(5, C2) \\
&= \sqrt{(0,182 - 0,182)^2 + (0,156 - 1,000)^2 + (0,714 - 0,786)^2 + (0,000 - 0,000)^2 + (0,034 - 0,051)^2} \\
&= 0,847
\end{aligned}$$

$$\begin{aligned}
& d(5, C3) \\
&= \sqrt{(0,182 - 0,000)^2 + (0,156 - 0,344)^2 + (0,714 - 0,000)^2 + (0,000 - 0,333)^2 + (0,034 - 0,132)^2} \\
&= 0,836
\end{aligned}$$

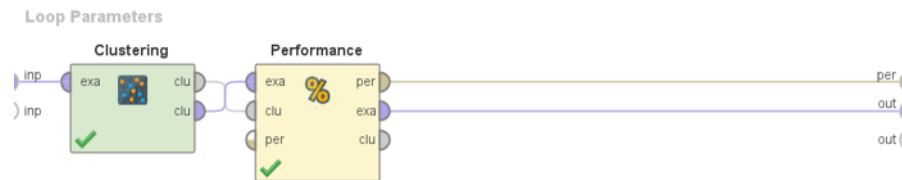
Berdasarkan hasil perhitungan jarak setiap data terhadap seluruh centroid, setiap data dikelompokkan ke dalam klaster dengan nilai jarak terkecil. Hasil pengelompokan data ke dalam klaster yang sesuai ditampilkan pada Tabel 4.16 berikut.

Tabel 4.17 Hasil Perhitungan Setiap Data ke Setiap Pusat *Cluster*

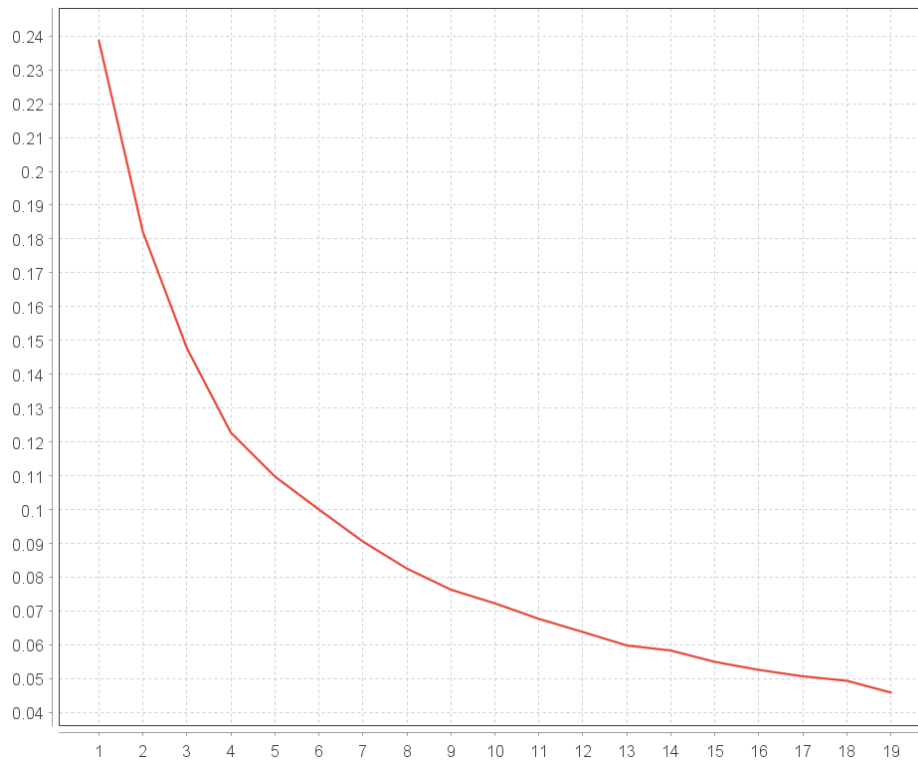
Nama Produk	Jarak Ke C0	Jarak Ke C1	Jarak Ke C2	Jarak Ke C3	Cluster
Mykonos Baby Love Parfum EDP 50 ML	0,000	1,204	0,902	1,164	C0
HINT Noble Eau de Parfum 50ml	1,204	0,000	1,371	1,085	C1
Xchange Extrait De Parfume Supernova Series 50ml	0,902	1,371	0,000	1,095	C2
HINT IDENTITY Eau De Parfume 50ml	1,164	1,085	1,095	0,000	C3
FYNE Extrait De Parfume	0,396	1,211	0,847	0,836	C0

4.4 Pengujian

Untuk menentukan jumlah kluster yang paling sesuai, penelitian ini menggunakan dua metode evaluasi, yaitu *Elbow Method* dan *Davies-Bouldin Index* (DBI), yang keduanya dijalankan melalui RapidMiner. Tahapan pengujian Elbow Method dapat dilihat pada Gambar 4.8, sedangkan hasil visualisasinya ditampilkan pada Gambar 4.9.



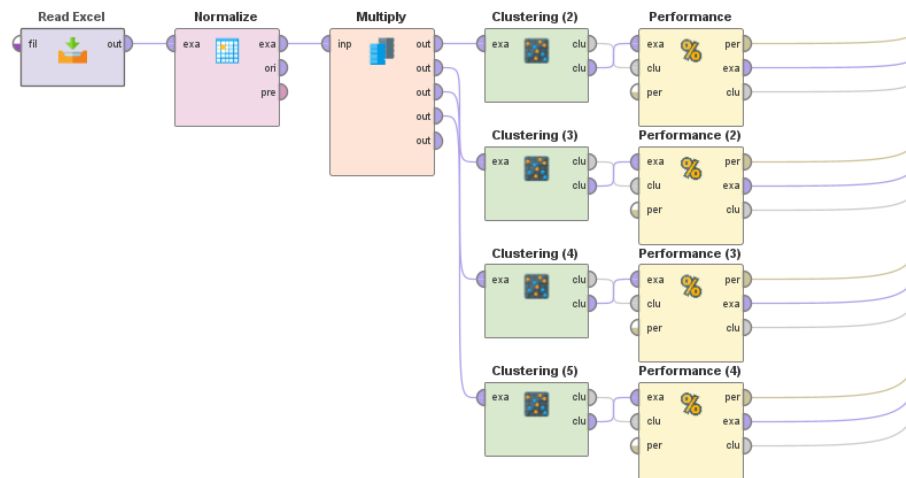
Gambar 4.8 Pengujian *Elbow Method* Menggunakan RapidMiner



Gambar 4.9 Grafik *Elbow Method*

Gambar 4.9 memperlihatkan hasil *Elbow Method* yang menggambarkan hubungan antara jumlah kluster (k) dan nilai *average within centroid distance*. Grafik menunjukkan penurunan nilai yang tajam dari kluster 2 hingga kluster 4. Setelah k=4, penurunan nilai mulai melambat dan kurva bergerak lebih landai, yang berarti penambahan kluster berikutnya tidak lagi memberikan pengurangan jarak yang berarti terhadap pusat kluster.

Dari pola tersebut, titik "siku" atau *elbow* terlihat paling jelas pada $k=4$, sehingga jumlah klaster optimal berdasarkan metode ini adalah 4 klaster.



Gambar 4.10 Pengujian Dengan *Davies Bouldien Index*

Tabel 4.18 Tabel Hasil Pengujian *Davies Bouldien Index*

Jumlah Cluster	Nilai DBI
Cluster 2	1.097
Cluster 3	1.146
Cluster 4	1.002
Cluster 5	1.077

Pengujian lanjutan dilakukan menggunakan metode *Davies-Bouldin Index* (DBI) untuk memperkuat hasil yang diperoleh. Hasil evaluasi menunjukkan nilai DBI terendah sebesar 1,002 juga ditemukan pada jumlah klaster 4, yang mencerminkan bahwa pemisahan antarklaster pada model ini paling optimal dibanding variasi jumlah klaster lainnya. Dengan demikian, baik dari sisi visualisasi Elbow maupun evaluasi kuantitatif DBI, jumlah 4 klaster ditetapkan sebagai jumlah klaster optimal dalam penelitian ini.

BAB V PENUTUP

5.1 Kesimpulan

- a. Berdasarkan hasil penelitian menggunakan RapidMiner dengan algoritma K-Means *Clustering*, model terbaik diperoleh pada jumlah 4 klaster dengan nilai *Davies-Bouldin Index* (DBI) sebesar 1,002. Nilai ini menunjukkan bahwa model yang terbentuk sudah cukup baik karena menghasilkan kelompok data yang jelas dan terpisah, sehingga dapat digunakan untuk memahami pola penjualan parfum pada GWMM SHOP berdasarkan dimensi waktu, produk, dan wilayah.
- b. Hasil pengelompokan membagi data transaksi menjadi empat klaster berdasarkan dimensi bulan, jumlah produk per transaksi, total harga transaksi, dan provinsi tujuan pengiriman. Klaster 0 berisi 1.203 transaksi dengan rata-rata bulan sebesar 6,089 yang terkonsentrasi pada bulan Juni hingga September. Rata-rata jumlah produk pada klaster ini sebesar 1,513 unit dengan rata-rata total harga transaksi sebesar Rp162.596, didominasi oleh produk Mykonos Baby Love Parfum EDP 50ML dan Fyne Extrait De Parfume dengan wilayah tujuan terbanyak Sumatera Utara. Klaster 1 merupakan klaster dengan jumlah transaksi terbanyak yaitu 1.449 transaksi dan rata-rata total harga tertinggi sebesar Rp725.445. Rata-rata bulan pada klaster ini sebesar 11,619 yang bertepatan dengan momen Harbolnas, Natal, dan Tahun Baru, dengan rata-rata jumlah produk per transaksi juga tertinggi yaitu 4,511 unit, serta wilayah tujuan terbanyak berada di DKI Jakarta, Jawa Barat, dan Jawa Timur. Klaster 2 berisi 652 transaksi dengan rata-rata bulan sebesar 5,678 yang terkonsentrasi pada bulan Mei hingga Juli. Rata-rata total harga transaksi pada klaster ini tergolong sedang yaitu Rp234.784 dengan rata-rata jumlah produk per transaksi sebesar 1,789 unit, didominasi produk Xchange Extrait De Parfume Supernova Series 50ml Terra Nova dan Sample Pack 3ml x5, dengan wilayah tujuan terbanyak berada di Papua, Nusa Tenggara Timur, dan Kalimantan Timur. Klaster 3 berisi 550 transaksi dengan rata-rata bulan sebesar 3,997 yang terkonsentrasi pada bulan Januari hingga Mei. Klaster ini memiliki rata-rata total harga transaksi paling rendah yaitu Rp93.116 dengan rata-rata jumlah produk per transaksi sebesar 1,417 unit, didominasi produk HINT *Identity* dan Ice Frost Extrait De Parfume 35ml, dengan wilayah tujuan terbanyak berada di Bali, Gorontalo, dan Bengkulu.
- c. Dari keempat klaster yang terbentuk, diperoleh beberapa rekomendasi strategi bisnis bagi GWMM SHOP. Klaster 1 menunjukkan penjualan tertinggi dengan nilai transaksi yang jauh melampaui klaster lain, sehingga GWMM SHOP perlu menyiapkan stok produk best seller dalam jumlah lebih banyak menjelang bulan Oktober hingga Desember serta memperkuat promosi pada wilayah Jawa, khususnya DKI Jakarta, Jawa Barat, dan Jawa Timur, mengingat periode ini bertepatan dengan Harbolnas, Natal, dan Tahun Baru yang menjadi puncak belanja konsumen. Klaster 0 mencerminkan penjualan yang stabil pada

pertengahan tahun dengan konsumen yang cenderung memilih produk mid-range, sehingga momentum ini dapat dimanfaatkan untuk mempertahankan loyalitas pelanggan di wilayah luar Jawa seperti Sumatera Utara melalui program promosi berkala. Klaster 2 dan Klaster 3 sama-sama mencerminkan segmen pasar dengan nilai transaksi yang lebih kecil dan didominasi produk berharga terjangkau, namun keduanya menunjukkan bahwa wilayah Indonesia timur dan tengah seperti Papua, Nusa Tenggara Timur, Kalimantan Timur, Bali, Gorontalo, dan Bengkulu tetap memiliki aktivitas transaksi yang perlu diperhatikan. GWMM SHOP dapat menjadikan wilayah-wilayah tersebut sebagai target ekspansi pasar melalui strategi diskon atau bundling produk terjangkau, sehingga potensi pasar di luar Pulau Jawa dapat dioptimalkan secara bertahap.

5.2 Saran

- a. Penelitian ini masih memiliki ruang untuk dikembangkan lebih lanjut, baik dengan menambahkan atribut lain seperti metode pembayaran atau jenis produk secara lebih rinci, menggunakan algoritma *clustering* yang berbeda, maupun menerapkan pendekatan ini pada UMKM dengan kategori produk lain guna memperoleh hasil analisis yang lebih mendalam dan sesuai konteks bisnis setiap pelaku usaha.
- b. Hasil klasterisasi ini dapat dijadikan dasar pertimbangan bagi pemilik GWMM SHOP dalam menentukan waktu, produk, dan wilayah yang perlu mendapat perhatian lebih, terutama dalam perencanaan stok, strategi promosi, dan target pemasaran ke depan.
- c. Penelitian berikutnya dapat mencoba membandingkan hasil K-Means dengan algoritma *clustering* lain selain K-Medoids, seperti DBSCAN atau Hierarchical *Clustering*, untuk melihat apakah ada pendekatan yang mampu menghasilkan nilai DBI lebih rendah pada data UMKM berbasis Shopee.
- d. penelitian berikutnya juga dapat diarahkan pada pembangunan model prediktif untuk mengestimasi lonjakan permintaan pada periode Oktober hingga Desember tahun berjalan. Pendekatan ini akan memberi nilai tambah praktis, karena hasil penelitian tidak hanya bersifat deskriptif terhadap pola yang telah terjadi, tetapi juga dapat difungsikan sebagai alat bantu perencanaan operasional bagi GWMM SHOP.

DAFTAR PUSTAKA

- Ananda, P. S. ... Sembiring, I. (2023). KMeans Clustering Menggunakan RapidMiner dalam Segmentasi Pelanggan dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal. *Jurnal BATIRSI*, 6(2).
- Dasopang, N. (2024). E-Commerce Bisnis Dan Internet. *JIEMAS: Jurnal Ilmiah Ekonomi, Manajemen Dan Syariah*, 3(1). <https://doi.org/10.55883/jiemas.v3i1>
- Dermawan, H. D. ... Wijaya, Y. A. (2024). CESS Analisis Data Penjualan Menggunakan Algoritma K-Means Clustering Pada Toko Daun Indah Di Shopee Analysis of Sales Data Using K-Means Clustering Algorithm at Daun Indah Shop in Shopee. *CESS (Journal of Computing Engineering, System and Science)*, 9(1), 175–191. <https://jurnal.unimed.ac.id/2012/index.php/cess/en/article/view/55093/pdf>
- Dona, D. ... Rahayu, D. (2025). Penerapan Data Mining Terhadap Penjualan Produk Kosmetik Menggunakan Algoritma K-Means. *Riau Jurnal Teknik Informatika*, 4(3), 502–513. <https://doi.org/10.30606/rjti.v4i3.4275>
- Eriskiannisa Febrianty Luchia Awalina, W. I. R. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions. *Jurnal Teknologi Dan Informasi*, 13(September), 122–137. <https://doi.org/10.34010/jati.v13i2>
- Fatkhudin, A. ... Zade, N. A. P. (2023). Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika). *Jurnal Minfo Polgan*, 12(2), 777–783. <https://doi.org/10.33395/jmp.v12i2.12494>
- Fitri, E. M. ... Wantoro, A. (2023). Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada PT XYZ. *Jurnal Komputasi*, 11(2), 157–168.
- Gori, T. ... Al Fatta, H. (2024). Preprocessing Data Dan Klasifikasi Untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Ilmu, G. J. M. (2024). Gudang Jurnal Multidisiplin Ilmu Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas. *Gudang Jurnal Multidisiplin Ilmu*, 2(November), 86–94. <http://gudangjurnal.com/index.php/gjmi/article/view/1040/933>
- Jollyta, D. ... Gustientiedina. (2023). Pengoptimalan Pengukuran Bregman Divergences Menggunakan Davies Bouldin Index. *JOISIE Journal Of Information System And Informatics Engineering*, 7(1), 174–180.
- Kamalludin, A., & Budianto, et al. (2023). Penerapan Algoritme Klasterisasi K-Means Kinerja Produk Dengan Analisis Recency Frequency Monetary Pada Cafe Xyz. *Prosiding Seminar ...*, 2(April), 258–266. <http://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/631%0Ahttps://senafiti.budiluhur.ac.id/index.php/senafiti/article/download/631/271>

- Khalif, A. ... Sari, B. N. (2024). Klasterisasi Tingkat Kemiskinan di Indonesia Menggunakan Algoritma K-Means. *Generation Journal*, 8(1), 54–62. <https://ojs.unpkediri.ac.id/index.php/gj/article/view/21470/3747>
- Pratiwi, D. S. ... Siregar, A. C. (2023). Peningkatan Keterampilan Pengolahan Data dengan Microsoft Excel terhadap Peserta Didik Madrasah Aliyah Al-Uswah Samarinda. *Jurnal Pengabdian Kepada Masyarakat Nusantara (JPkMN)*, 4(4), 4542–4549. <https://doi.org/10.55338/jpkmn.v4i4.2166>
- Ramadhan, M. I. ... Syafria, F. (2026). Analisis Clustering Menggunakan Metode K-Means untuk Mengidentifikasi Pola Kepuasan Alumni. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 6(1), 31–41. <https://doi.org/10.57152/malcom.v6i1.2401>
- Riadi, R., & Mesran. (2023). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume. *Journal of Informatics, Electrical and Electronics Engineering*, 2(4), 138–145. <https://doi.org/10.47065/jieeee.v2i4.1181>
- Vania, P., & Sari, B. N. (2023). Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means. *Jurnal Ilmiah Wahana Pendidikan*, 9(21), 547–558. <https://doi.org/10.5281/zenodo.10081332>
- Wahyudi, S. I., & Wibowo, A. (2022). Implementasi Metode K-Means Clustering Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI), September*, 470–478. <https://senafti.budiluhur.ac.id/index.php>
- Wilyani, F. ... Aslimar, F. (2024). Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory. *Jurnal Pengabdian Pada Masyarakat Indonesia (JPPMI)*, 3(1), 8–14. <https://doi.org/10.55606/jppmi.v3i1.1087>

**LAMPIRAN
DATASET AWAL**

Bulan	Nama Produk	Provinsi	Jumlah	Total Harga Produk
3	FYNE Extrait De Parfume	NUSA TENGGERA TIMUR	1	89000
3	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	NUSA TENGGERA TIMUR	2	218000
12	Fragrance World Cocktail Intense Unisex Eau De Parfum 100 ML	BANTEN	4	1064000
12	Verbena Night Legend Man Eau De Parfum 100 ML	JAWA BARAT	5	610000
7	BRODO - Rote EDT Parfume 40ml	BALI	2	198000
5	BRODO - Rote EDT Parfume 40ml	SUMATERA UTARA	1	94000
12	Fragrance World Suits Unisex Eau De Parfum 100 ML	DKI JAKARTA	4	968000
12	Scentplus Dominant 35ml	BANTEN	4	448000
8	Omerta The Winner Parfume Pria Eau de Toilette 100 ML	JAWA TIMUR	5	518500
2	GARVI Perfume - Naughty Friday - EDP 50ml	KALIMANTAN TIMUR	1	109000
12	HINT Noble Eau de Parfum 50ml	JAWA BARAT	3	399000
7	Kokulu Women Best Seller - Bundling 2pcs 50ML	SUMATERA UTARA	1	122000
12	NARSIS Hypnotic Desire Eau de Parfume 50ml	JAWA TIMUR	5	545000
3	Mezuca Parfume Aqua Kiss Eau de Parfume 50ml	PAPUA	1	109000
12	Omerta The Winner Parfume Pria Eau de Toilette 100 ML	JAWA BARAT	5	610000

7	HINT IDENTITY Eau de Parfum 50ml	SULAWESI SELATAN	1	109000
12	HINT Noble Eau de Parfum 50ml	JAWA TENGAH	3	399000
12	Omerta Pink Ice Woman Eau de Parfum 100 ML	DKI JAKARTA	6	715200
2	Vial Parfum 5ml	MALUKU	2	10000
12	SAFF & Co. Extrait de Parfum - CASCABEL 30ml	BANTEN	4	596000
7	Mykonos Baby Love Parfum EDP 50 ML	RIAU	2	298000
3	Vial Parfum 5ml	MALUKU	2	10000
12	Mykonos Verve Extrait De Parfum 50 ML	DKI JAKARTA	3	357600
3	GARVI Perfume - Naughty Friday - EDP 50ml	SUMATERA UTARA	2	218000
11	Scentplus Dominant 35ml	JAWA BARAT	3	336000
12	Omerta Pink Ice Woman Eau de Parfum 100 ML	JAWA TENGAH	4	596000
12	SAFF & Co. Extrait de Parfum - CASCABEL 30ml	JAWA TENGAH	4	596000
8	Dynamique Course Pour Homme Men EDP 100 ML	BALI	3	366000
12	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	JAWA TIMUR	3	1047000
4	(BUY 1 GET 1) Parfum Zara 90ML Box Laki-laki Perempuan	BANTEN	3	1047000
12	SAFF & Co. Extrait de Parfum - CASCABEL 30ml	BANTEN	6	894000
12	Russ Perfume Unomento Blue Unisex 50ml	BANTEN	7	763000

10	Verbena Night Legend Man Eau De Parfum 100 ML	DKI JAKARTA	5	610000
7	Revlon Charlie Fragrance - White EDT Parfum 100 ml	SUMATERA UTARA	1	89000
9	BRODO - Rote EDT Parfume 40ml	SULAWESI SELATAN	1	89100
12	HINT Noble Eau de Parfum 50ml	BANTEN	3	399000
12	HINT Noble Eau de Parfum 50ml	DKI JAKARTA	4	532000
4	Mykonos Milk Drips 50ml	SUMATERA UTARA	3	453000
6	Kokulu Women Best Seller - Bundling 2pcs 50ML	SUMATERA UTARA	2	231800
9	Kokulu Women Best Seller - Bundling 2pcs 50ML	BALI	1	122000
7	Onix Fragrance Eau de Perfume Senoparty 30MI	KALIMANTAN TIMUR	3	267000
12	Omerta The Winner Parfume Pria Eau de Toilette 100 ML	JAWA TIMUR	3	366000
3	FYNE Extrait De Parfume	PAPUA	1	89000
7	Fogg Perfume Scent Premium Many Flowers 100ml Eau De Perfume	KALIMANTAN TIMUR	1	109000
12	Scentplus Dominant 35ml	JAWA TIMUR	4	448000
5	Vial Parfum 5ml	BENGKULU	1	5000
12	Scentplus Dominant 35ml	DKI JAKARTA	6	571200
3	Pouch Parfum Mini	NUSA TENGGARA TIMUR	1	12000
12	Omerta The Winner Parfume Pria Eau de Toilette 100 ML	DKI JAKARTA	3	366000
11	Omerta The Winner Parfume Pria Eau de Toilette 100 ML	DKI JAKARTA	4	488000

DATASET HASIL KLASTERISASI

Bulan	Nama Produk Enc	Provinsi_Enc	Jumlah	Total Harga Produk	Cluster
0,181818	0,15625	0,714286	0	0,034454	2
0,181818	0,46875	0,714286	0,166667	0,087367	2
1	0,25	0,071429	0,5	0,434372	0
1	0,9375	0,357143	0,666667	0,248154	0
0,545455	0,0625	0	0,166667	0,079163	1
0,363636	0,0625	1	0	0,036505	2
1	0,28125	0,214286	0,5	0,394996	0
1	0,90625	0,071429	0,5	0,181706	0
0,636364	0,65625	0,5	0,666667	0,210623	0
0,090909	0,3125	0,571429	0	0,042658	2
1	0,375	0,357143	0,333333	0,161608	0
0,545455	0,4375	1	0	0,04799	2
1	0,59375	0,5	0,666667	0,221493	0
0,181818	0,46875	0,785714	0	0,042658	2
1	0,65625	0,357143	0,666667	0,248154	0
0,545455	0,34375	0,928571	0	0,042658	2
1	0,375	0,428571	0,333333	0,161608	0
1	0,625	0,214286	0,833333	0,291304	0
0,090909	0,96875	0,642857	0,166667	0,002051	3
1	0,84375	0,071429	0,5	0,242412	0
0,545455	0,5	0,857143	0,166667	0,12018	2
0,181818	0,96875	0,642857	0,166667	0,002051	3
1	0,5625	0,214286	0,333333	0,144627	0
0,181818	0,3125	1	0,166667	0,087367	2
0,909091	0,90625	0,357143	0,333333	0,135767	0
1	0,625	0,428571	0,5	0,242412	0
1	0,84375	0,428571	0,5	0,242412	0
0,636364	0,125	0	0,333333	0,148072	1
1	0	0,5	0,333333	0,4274	0
0,272727	0	0,071429	0,333333	0,4274	1
1	0,84375	0,071429	0,833333	0,364643	0
1	0,8125	0,071429	1	0,310911	0
0,818182	0,9375	0,214286	0,666667	0,248154	0
0,545455	0,78125	1	0	0,034454	3
0,727273	0,0625	0,928571	0	0,034495	2
1	0,375	0,071429	0,333333	0,161608	0
1	0,375	0,214286	0,5	0,216161	0
0,272727	0,53125	1	0,333333	0,183757	2
0,454545	0,4375	1	0,166667	0,093027	2
0,727273	0,4375	0	0	0,04799	1
0,545455	0,6875	0,571429	0,333333	0,107465	3
1	0,65625	0,5	0,333333	0,148072	0
0,181818	0,15625	0,785714	0	0,034454	2

0,545455	0,21875	0,571429	0	0,042658	2
1	0,90625	0,5	0,5	0,181706	0
0,363636	0,96875	0,142857	0	0	3
1	0,90625	0,214286	0,833333	0,23224	0
0,181818	0,71875	0,714286	0	0,002871	3
1	0,65625	0,214286	0,333333	0,148072	0
0,909091	0,65625	0,214286	0,5	0,198113	0
0,727273	0,625	0,357143	0,333333	0,181296	0
0,727273	0,8125	0,428571	0,333333	0,132075	0
1	0,90625	0,071429	0,333333	0,135767	0
1	0,84375	0,357143	0,333333	0,181296	0
0,181818	0,5	0	0	0,059065	1
0,181818	0,03125	0,642857	0	0,002051	2
0,181818	0,875	0,285714	0	0,004102	3
1	0,59375	0,5	0,5	0,176784	0
0,454545	0,4375	1	0	0,042986	2
0,636364	0,4375	1	0	0,04799	2
0,545455	0,4375	0,571429	0	0,04799	2
0,545455	0,1875	0,571429	0	0,042658	2
1	0	0,357143	0,333333	0,4274	0
0	0,1875	0,571429	0,166667	0,087367	2
0,545455	0,5	0	0,166667	0,12018	1
1	0,9375	0,5	0,333333	0,148072	0
1	0,65625	0,5	0,5	0,178097	0
0,545455	0,6875	0	0,333333	0,107465	1
0,636364	0,75	0,714286	0	0,008203	3
1	0,84375	0,5	0,5	0,230271	0
0,181818	0,15625	0,785714	0	0,034454	2
0,545455	0,6875	0,928571	0,333333	0,107465	2
1	0,625	0,428571	0,5	0,193519	0
0,090909	0,5	1	0,333333	0,162961	2
1	1	0,142857	0	0,050861	0
0,727273	0,6875	0,928571	0	0,034454	2
0,545455	0,5	0,571429	0	0,059065	2
1	0,28125	0,357143	0,5	0,394996	0
0,363636	0,5	0	0,333333	0,181296	1
0,545455	0,53125	0	0,333333	0,183757	1
1	0,90625	0,357143	0,333333	0,135767	0
0,545455	0,34375	1	0,166667	0,087367	2
0,454545	0,53125	0,857143	0	0,059885	2
0,181818	0,40625	0,142857	0	0,045119	1
0,545455	0,84375	0,071429	0,833333	0,364643	0
0,545455	0,53125	0	0	0,053692	1
0,545455	0,3125	1	0	0,042658	2
0,545455	0,34375	0,928571	0	0,042658	2
1	0,5625	0,428571	0,333333	0,181296	0

1	0,59375	0,071429	0,666667	0,221493	0
0,545455	0,3125	0,857143	0	0,042658	2
0,545455	0,0625	0,571429	0	0,038556	2
0,545455	0,125	0,928571	0,166667	0,083019	2
1	0,9375	0,5	0,666667	0,198113	0
0,909091	0,28125	0,214286	0,666667	0,494258	0
1	0	0,5	0,666667	0,7137	0
0,545455	0,3125	0	0,166667	0,087367	1
0,545455	0,34375	1	0,166667	0,078425	2
1	0,8125	0,071429	0,333333	0,132075	0
1	0,84375	0,071429	0,833333	0,364643	0

SURAT KETERANGAN RISET



GWMM SHOP

Jl. Pasar Lembang No.20, Kota Tangerang,

Telp. (089) 89747769

SURAT KETERANGAN RISET

Yang bertanda tangan dibawah ini:

Nama : Milania

Jabatan : Direksi

menerangkan bahwa:

Nama : Febri Ardana

NIM : 2212500199

telah melaksanakan riset pada bagian penjualan *E-Commerce* Shopee GWMM SHOP sejak tanggal 13 April 2026 s/d 04 Mei 2026 dengan baik.

Demikian Surat Keterangan ini dibuat untuk dapat dipergunakan semestinya.

Dibuat di : Kota Tangerang

Tanggal : 24 Juni 2026



Milania

SIMILARITY

TugasAkhir_2212500199_FebriArdana.docx

ORIGINALITY REPORT

20%
SIMILARITY INDEX

16%
INTERNET SOURCES

13%
PUBLICATIONS

9%
STUDENT PAPERS

PRIMARY SOURCES

- 1 dirdosen.budiluhur.ac.id
Internet Source
- 2 Submitted to Universitas Tarumanagara
Student Paper
- 3 jurnal.unimed.ac.id
Internet Source
- 4 repository.unissula.ac.id
Internet Source
- 5 repository.ulb.ac.id
Internet Source
- 6 docplayer.info
Internet Source
- 7 ojs.itb-ad.ac.id
Internet Source
- 8 Eunike Charina Ibrena Tarigan, Angela Wita Simanullang, Daniel S. Simbolon, Sardo Sipayung. "Clustering Penjualan Toko Retail Menggunakan Algoritma K-Means dalam Proses Penambangan Data", remik, 2026
Publication
- 9 cdn.juris.id
Internet Source
- 10 jurnal.univrab.ac.id
Internet Source
- 11 repository.uin-suska.ac.id
Internet Source
- 12 Yoel michael Sihombing, Eka Sri Rahayu. "Implementasi Fuzzy C-Means Clustering dalam Segmentasi Pelanggan dengan RFM Model", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025
Publication
- 13 repository.itpln.ac.id
Internet Source