JUPITER
Teknologi Informatika & Komputer
Vol. 7 No. 1 Bulan 2026
Universitas Muhammadiyah Bekasi Karawang
P-ISSN: 2798-3129 , E-ISSN: 2798-6187

PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING VOLUME DISTRIBUSI PAKET LION PARCEL JABODETABEK

Modesta Rika Pertiwi[1], Betha Nurina Sari[2], Dadang Yusup[3]
[1]Universitas Singaperbangsa Karawang, 1910631170098@student.unsika.ac.id,
[2]Universitas Singaperbangsa Karawang, betha.nurina@staff.unsika.ac.id,
[3]Universitas Singaperbangsa Karawang, dadang.dyf@staff.unsika.ac.id

ABSTRAK

Peningkatan aktivitas pengiriman paket menuntut perusahaan untuk memahami pola distribusi agar dapat meningkatkan efektivitas operasional. Penelitian ini bertujuan menerapkan algoritma KMeans Clustering untuk mengelompokkan volume distribusi paket Lion Parcel di wilayah Jabodetabek. Penelitian menggunakan data distribusi paket periode Oktober–Desember 2025 yang mencakup 183 kecamatan. Metode penelitian mengacu pada tahapan Knowledge Discovery in Databases (KDD) yang meliputi data selection, data cleaning, data transformation, data mining, dan evaluation. Penentuan jumlah cluster dilakukan menggunakan Elbow Method, kemudian proses pengelompokan dilakukan dengan algoritma K-Means menggunakan 5 cluster. Hasil pengelompokan menghasilkan lima kategori, yaitu Sangat Rendah, Rendah, Sedang, Tinggi, dan Sangat Tinggi. Evaluasi menggunakan Silhouette Coefficient menghasilkan nilai 0,6313, sedangkan Davies–Bouldin Index menghasilkan nilai 0,4336, yang menunjukkan bahwa hasil pengelompokan memiliki kualitas yang baik dengan pemisahan antar-cluster yang cukup jelas. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengidentifikasi pola volume distribusi paket pada setiap kecamatan sehingga dapat menjadi dasar dalam mendukung pengambilan keputusan dan penyusunan strategi distribusi yang lebih efektif.

Kata Kunci: Data Mining, Distribusi Paket, KDD, K-Means Clustering, Lion Parcel.

ABSTRACT

The increasing volume of package deliveries requires companies to understand distribution patterns to improve operational efficiency. This study aims to apply the K-Means Clustering algorithm to classify Lion Parcel package distribution volumes in the Greater Jakarta (Jabodetabek) area. The dataset consists of package distribution data from October to December 2025 covering 183 subdistricts. The research follows the Knowledge Discovery in Databases (KDD) process, including data selection, data cleaning, data transformation, data mining, and evaluation. The optimal number of clusters was determined using the Elbow Method, and the clustering process was performed using five clusters. The results produced five distribution categories: Very Low, Low, Moderate, High, and Very High. The clustering performance achieved a Silhouette Coefficient of 0.6313 and a Davies–Bouldin Index of 0.4336, indicating good cluster quality with satisfactory separation between clusters. The findings demonstrate that the K-Means algorithm effectively identifies package distribution patterns across sub-districts and can support decision-making in developing more efficient distribution strategies.

Keywords: *Data Mining, KDD, K-Means Clustering, Lion Parcel, Package Distribution.*

Naskah diterima : 20 Februari 2026, Naskah dipublikasikan : 24 Februari 2026

PENDAHULUAN

Kemajuan teknologi digital dan kehadiran marketplace telah mengubah pola transaksi masyarakat Indonesia secara signifikan, yang berdampak langsung pada meningkatnya volume pengiriman barang di industri logistik. Wilayah JABODETABEK menjadi pusat aktivitas ekonomi dengan volume distribusi paket yang sangat tinggi setiap harinya. Dalam operasional perusahaan logistik, perbedaan volume distribusi antarwilayah memengaruhi kebutuhan armada, kurir, tenaga operasional, kapasitas gudang, dan penjadwalan proses sortir. Tanpa informasi mengenai karakteristik distribusi setiap wilayah, perencanaan operasional cenderung mengandalkan estimasi sehingga berisiko menimbulkan ketidakseimbangan beban kerja, penumpukan paket, keterlambatan pengiriman, serta peningkatan biaya operasional.

Di sisi lain, analisis data distribusi pada Lion Parcel Jabodetabek masih dilakukan secara manual melalui tabel atau laporan transaksi, hal ini kurang efektif karena memerlukan waktu yang lama dan sulit mengungkap pola distribusi yang tersembunyi. Padahal, informasi mengenai pengelompokan wilayah berdasarkan tingkat volume distribusi sangat penting sebagai dasar penyusunan strategi operasional yang lebih tepat. Kondisi tersebut menjadi tantangan bagi perusahaan ekspedisi seperti Lion Parcel, karena pengelolaan data secara manual dapat menghambat proses pengambilan keputusan, meningkatkan risiko ketidakseimbangan distribusi, menyebabkan penumpukan paket di hub tertentu, serta memperlambat penentuan prioritas operasional.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan metode *data mining* menggunakan algoritma *K-Means Clustering*. Metode ini dipilih karena terbukti sederhana, cepat, dan efektif dalam mengelompokkan data berskala besar menjadi klaster volume tinggi, sedang, dan rendah berdasarkan kemiripan karakteristik wilayah. Penelitian ini diharapkan dapat membantu perusahaan mengoptimalkan kapasitas operasional, meminimalkan keterlambatan, dan meningkatkan efisiensi layanan pengiriman. Rumusan masalah dalam penelitian ini berfokus pada bagaimana penerapan algoritma *K-Means Clustering* untuk mengelompokkan wilayah berdasarkan volume distribusi paket, apa saja karakteristik yang terbentuk dari setiap klaster tersebut, serta bagaimana hasil akhir dari pemetaan wilayah pada masing-masing klaster.

Agar tidak melebar, penelitian ini dibatasi hanya pada analisis volume distribusi pengiriman wilayah tujuan JABODETABEK menggunakan data POS (Point of Sale) Lion Parcel Desa Burangkeng, Kecamatan Setu, Kabupaten Bekasi, selama periode Oktober–Desember 2025. Proses segmentasi wilayah dilakukan menggunakan algoritma K-Means Clustering dengan penentuan jumlah cluster menggunakan Elbow Method, sedangkan kualitas hasil clustering dievaluasi menggunakan Silhouette Coefficient dan Davies-Bouldin Index. Seluruh proses pengolahan dan analisis data dilakukan menggunakan bahasa pemrograman Python beserta library data mining yang relevan.

Dalam penelitian sebelumnya, Mauludin et al. (2025) menggunakan Algoritma KMeans Clustering untuk menganalisis kinerja pengiriman paket Shopee Express di Hub Transit Kedawung dimana menggunakan 359 data pengiriman dan menghasilkan jumlah cluster optimal K=10 dengan nilai DaviesBouldin Index (DBI) 0,288. Clustering mampu mengidentifikasi kelompok kinerja pengiriman paket secara efektif.

Olivia (2023) juga menggunakan K-Means dan Hierarchical Clustering dalam penelitiannya dimana ia berhasil membentuk 12 cluster titik pengiriman untuk optimasi distribusi barang. Evaluasi menunjukkan Silhouette Score 0,756 dan DBI 0,36, sehingga clustering dinilai efektif dalam mendukung pengambilan keputusan distribusi.

Penelitian Danjiah et al. (2025) juga menggunakan K-Means Clustering, dimana K-Means efektif untuk mengevaluasi efisiensi lokasi depot last-mile delivery dan membantu menentukan strategi penempatan depot yang lebih optimal untuk perusahaan logistik.

Siahaan et al. (2024) juga menerapkan Metode K-Means Clustering untuk pengelompokan minat konsumen terhadap pengguna jasa layanan pada Kantor Pos Binjai, Hasil penelitian menunjukkan bahwa clustering dapat membantu perusahaan memahami pola pelanggan dan meningkatkan kualitas layanan.

Adhitama & Wibisono (2024) juga menggunakan K-Means, Tabu Search dan ACS dalam penelitiannya dalam perencanaan alokasi dan rute distribusi beras di Kota Yogyakarta, K-Means menghasilkan 3 kelompok lokasi distribusi berdasarkan kapasitas angkut dan permintaan pasar. Hasilnya mampu menekan waktu distribusi, jarak tempuh, emisi, dan biaya operasional.

Berdasarkan penelitian terdahulu, algoritma K-Means telah banyak diterapkan pada analisis di bidang logistik dan distribusi. Pada penelitian yang dilakukan saat ini akan berfokus pada analisis pola distribusi pengiriman paket Lion Parcel pada setiap kecamatan di area JABODETABEK. Metode yang akan digunakan pada penelitian ini yaitu K-Means Clustering yang bertujuan mengidentifikasi kelompok wilayah dengan volume distribusi tinggi, sedang, dan rendah untuk mendukung pengambilan keputusan logistik. Hasil clustering nantinya akan dievaluasi menggunakan beberapa metode seperti Elbow Method (Metode Siku), Silhouette Coefficient, dan Davies-Bouldin Index (DBI), guna mengukur kualitas cluster yang terbentuk.

Tujuan dari penelitian ini adalah untuk menerapkan algoritma *K-Means Clustering* dalam mengelompokkan wilayah Jabodetabek berdasarkan volume distribusi paket Lion Parcel, menganalisis karakteristik setiap klaster yang terbentuk guna mengategorikan tingkatan volumenya, serta menghasilkan pemetaan wilayah yang dapat memberikan gambaran persebaran distribusi sebagai pendukung pengambilan keputusan dalam menyusun strategi distribusi yang lebih efektif.

TINJAUAN PUSTAKA

Data Mining

Data mining merupakan pendekatan analisis komputasi yang digunakan untuk menggali informasi tersembunyi dari kumpulan data sehingga menghasilkan pola yang dapat dimanfaatkan sebagai dasar pengambilan Keputusan. Menurut Dias et al. (2021) Tujuan utama *data mining* adalah:

1. Menemukan pola tersembunyi dalam data.
2. Melakukan prediksi terhadap kejadian di masa mendatang.
3. Mengelompokkan data berdasarkan karakteristik tertentu.
4. Mengidentifikasi hubungan antar variabel.
5. Mendukung pengambilan keputusan yang lebih efektif dan berbasis data.

K-Means

Menurut Setyawan et al. (2025) Clustering merupakan teknik analisis data yang digunakan untuk mengelompokkan sekumpulan data berdasarkan kemiripan karakteristik tertentu tanpa adanya informasi kelas sebelumnya. Prinsip utama clustering adalah memaksimalkan kemiripan antar objek dalam satu kelompok (intra-cluster 19 similarity) dan meminimalkan kemiripan antar kelompok (inter-cluster dissimilarity). Menurut Setyaningtyas et al. (2022), tujuan utama *clustering* adalah mengelompokkan data dengan karakteristik serupa ke dalam satu klaster guna menemukan pola atau struktur tersembunyi serta mengidentifikasi hubungan antarobjek yang sebelumnya tidak diketahui. Melalui pendekatan ini, proses analisis data berukuran besar dapat disederhanakan sehingga mampu mendukung pengambilan keputusan yang lebih efektif berdasarkan hasil pengelompokan tersebut.

Menurut Ikotun et al. (2023), K-Means bekerja dengan menentukan sejumlah cluster terlebih dahulu, kemudian memilih titik pusat (centroid) untuk setiap cluster. Selanjutnya, setiap data akan dikelompokkan ke centroid yang memiliki jarak paling dekat. Proses tersebut dilakukan secara berulang hingga posisi centroid stabil dan tidak mengalami perubahan yang signifikan.

Metode Elbow

Metode *Elbow* adalah teknik yang digunakan untuk menentukan jumlah klaster (k) optimal pada algoritma K-Means dengan cara mengevaluasi penurunan nilai *Sum of Squared Error* (SSE). Semakin banyak klaster, nilai SSE akan terus menurun secara signifikan hingga mencapai titik tertentu di mana penurunan tersebut mulai melambat dan membentuk pola menyerupai siku (*elbow*). Titik siku inilah yang dipilih sebagai jumlah klaster terbaik karena mampu memberikan keseimbangan optimal antara kualitas pengelompokan dan kompleksitas model. (Maori & Evanita, 2023).

Silhouette Score

Silhouette Score merupakan metode validasi internal untuk mengevaluasi kualitas hasil *clustering* berdasarkan tingkat kedekatan antaranggota dalam klaster yang sama (*cohesion*) dan tingkat keterpisahan antarkluster (*separation*). Nilai evaluasi ini berkisar antara -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan pengelompokan yang sangat baik dan terpisah jelas, nilai mendekati 0 mengindikasikan posisi data yang kurang tegas di antara dua klaster, sedangkan nilai negatif menandakan adanya kesalahan penempatan objek. Dalam penerapannya, jumlah klaster (k) yang menghasilkan nilai rata-rata (*Average Silhouette Score*) tertinggi akan dipilih sebagai solusi terbaik karena menghasilkan struktur klaster yang paling kompak.

Dalam praktik algoritma K-Means, *Silhouette Score* sering digunakan sebagai pelengkap metode *Elbow* agar penentuan jumlah klaster optimal tidak hanya bertumpu pada penurunan nilai SSE, melainkan juga mempertimbangkan kualitas strukturnya secara objektif. Namun, metode ini memiliki keterbatasan karena nilainya sangat dipengaruhi oleh jenis pengukuran jarak yang digunakan serta menjadi kurang representatif pada dataset dengan kepadatan berbeda atau bentuk klaster yang tidak beraturan. Oleh karena itu, hasil evaluasi menggunakan *Silhouette Score* sebaiknya tetap dikombinasikan dengan metode validasi lain, seperti *Davies–Bouldin Index*, agar hasil penentuan klaster menjadi lebih akurat.

Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) merupakan metode validasi internal yang digunakan untuk mengukur kualitas hasil *clustering* berdasarkan tingkat kekompakan (*compactness*) anggota di dalam klaster dan tingkat pemisahan (*separation*) antarkluster. Diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979, indeks ini dihitung dengan membandingkan rasio penyebaran data dalam suatu klaster terhadap jarak antarpusat (*centroid*) klaster tersebut. Nilai DBI yang dihasilkan kemudian dirata-ratakan, di mana nilai yang semakin kecil atau mendekati 0 menunjukkan kualitas pengelompokan yang semakin baik karena klasternya bersifat homogen, kompak, dan terpisah dengan jelas tanpa tumpang tindih.

Dalam penerapan algoritma K-Means, DBI sering digunakan sebagai metode evaluasi objektif untuk menentukan jumlah klaster (k) yang paling optimal dengan cara memilih nilai DBI yang paling kecil. Melalui pendekatan ini, penentuan jumlah klaster menjadi lebih seimbang karena mempertimbangkan kekompakan sekaligus pemisahan antarkluster. Oleh karena itu, untuk menghasilkan proses evaluasi yang lebih komprehensif dan akurat, DBI sangat disarankan untuk digunakan bersama dengan metode *Elbow* dan *Silhouette Score*.

METODE PENELITIAN

Objek Penelitian

Objek penelitian yang digunakan adalah data distribusi paket Lion Parcel pada wilayah JABODETABEK. Data penelitian berupa 2.828 data volume pengiriman paket pada periode Oktober–Desember 2025. Data tersebut akan dianalisis menggunakan algoritma K-Means Clustering untuk mengelompokkan kecamatan berdasarkan karakteristik volume distribusi paket, sehingga dapat diketahui kelompok wilayah dengan tingkat distribusi tinggi, sedang, dan rendah.

Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penelitian ini adalah metode Knowledge Discovery in Database (KDD). Metode KDD dipilih karena dapat membantu proses pengolahan data secara sistematis untuk menemukan pola atau informasi baru dari kumpulan data yang besar. Pada penelitian ini, metode KDD digunakan untuk melakukan analisis dan pengelompokan volume distribusi paket Lion Parcel di wilayah JABODETABEK menggunakan algoritma K-Means Clustering yang meliputi seleksi, pembersihan, transformasi, data mining, serta evaluasi. Menggunakan data transaksi periode Oktober hingga Desember 2025 sebanyak 2.828 transaksi, diperoleh 183 kecamatan yang kemudian dikelompokkan untuk mengidentifikasi pola distribusi berdasarkan karakteristik volumenya. Hasil pengelompokan tersebut dievaluasi menggunakan Elbow Method, Silhouette Score, dan Davies-Bouldin Index (DBI) untuk menentukan jumlah klaster optimal serta kualitasnya, dengan harapan dapat memberikan gambaran persebaran volume distribusi sekaligus mendukung pengambilan keputusan operasional dan strategi distribusi perusahaan.

Selection Data

Tahap seleksi data (*selection*) dilakukan dengan memilih atribut yang relevan dari dataset awal untuk disesuaikan dengan fokus penelitian, yaitu pengelompokan wilayah berdasarkan volume distribusi paket. Terdapat tiga atribut utama yang

digunakan, meliputi Kecamatan Tujuan sebagai unit wilayah pengelompokan, Volume Pengiriman sebagai variabel utama *clustering*, dan Kota Tujuan sebagai informasi pendukung interpretasi. Setelah dilakukan seleksi dan pengelompokan berdasarkan kecamatan untuk memperoleh total volume distribusi yang lebih ringkas, dari total 2.828 transaksi pengiriman awal dihasilkan 183 kecamatan tujuan pengiriman yang siap digunakan sebagai objek penelitian pada tahap *preprocessing* selanjutnya.

Preprocessing Data

Proses pembersihan data dalam penelitian ini dilakukan melalui empat langkah utama untuk memastikan kualitas data sebelum dianalisis. Langkah pertama adalah pemeriksaan data kosong (*missing value*) pada setiap atribut, yang hasilnya menunjukkan tidak adanya data kosong sehingga seluruh data dapat diproses. Selanjutnya, dilakukan pemeriksaan dan penghapusan data duplikat guna menghindari kesalahan penghitungan volume pengiriman yang dapat memengaruhi hasil *clustering*.

Langkah ketiga adalah standarisasi penulisan nama wilayah tujuan untuk menyeragamkan nama kecamatan agar memiliki identitas yang konsisten. Setelah data dinyatakan bersih, langkah terakhir adalah agregasi data berdasarkan kecamatan tujuan untuk menggabungkan dan menghitung total volume pengiriman, sehingga menghasilkan data yang lebih ringkas dan siap digunakan oleh algoritma K-Means.

Tabel 1. Hasil Pembersihan Data

Keterangan	Jumlah
Total Transaksi	2.828
Data Kosong	0
Data Duplikat	0
Total Kecamatan	183

Kemudian diperoleh sebanyak 183 kecamatan tujuan pengiriman yang siap digunakan pada tahap transformasi dan clustering. Hasil pembersihan data menunjukkan bahwa dataset memiliki kualitas yang baik karena tidak ditemukan data kosong maupun duplikasi yang signifikan.

Transformation Data

Tahap transformasi data dilakukan untuk memperbaiki struktur data, mengatasi ketidaksesuaian format, dan menyiapkan data volume distribusi di 183 kecamatan wilayah Jabodetabek agar dapat diolah secara optimal oleh algoritma *clustering*. Berdasarkan pemeriksaan, ditemukan 24 data (sekitar 13,04%) yang bernilai nol karena tidak memiliki aktivitas distribusi pada periode pengamatan. Nilai nol ini tetap dipertahankan dalam penelitian karena dianggap sebagai informasi relevan yang merepresentasikan kondisi aktual di lapangan, bukan sebagai data hilang (*missing value*) atau kesalahan input.

Meskipun dipertahankan, keberadaan data bernilai nol berpotensi memengaruhi posisi *centroid* dengan menarik pusat klaster ke nilai yang lebih rendah. Untuk meminimalkan pengaruh perbedaan skala yang terlalu besar tersebut, dilakukan proses normalisasi menggunakan metode *StandardScaler* yang mengubah data ke skala standar dengan rata-rata mendekati nol dan standar deviasi sebesar satu. Melalui proses ini,

seluruh data berhasil disusun dalam format yang lebih terstruktur dan siap digunakan pada tahap *K-Means Clustering* guna menghasilkan pengelompokan yang lebih representatif.

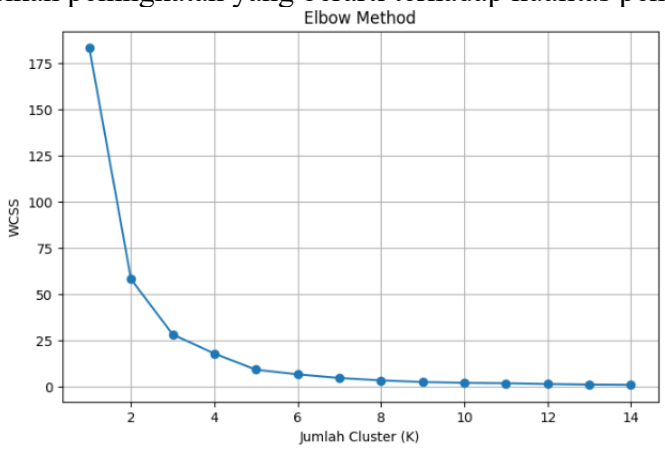
Kota	Kecamatan	Volume
Jakarta Timur	Cakung, Jakarta Timur	74
	Cipayung, Jakarta Timur	26
	Ciracas, Jakarta Timur	26
	Duren Sawit, Jakarta Timur	41
	Jatinegara, Jakarta Timur	14
	Kramatjati, Jakarta Timur	40
	Makasar, Jakarta Timur	30
	Matraman, Jakarta Timur	8
	Pasar Rebo, Jakarta Timur	31
	Pulogadung, Jakarta Timur	37
Jakarta Barat	Cengkareng, Jakarta Barat	44
	Grogol Petamburan, Jakarta Barat	26
	Taman Sari, Jakarta Barat	10
	Tambora, Jakarta Barat	7
	Kebon Jeruk, Jakarta Barat	21
	Kalideres, Jakarta Barat	22
	Pal Merah, Jakarta Barat	17
	Kembangan, Jakarta Barat	20

Gambar 1. Tampilan Data Setelah Transformasi

Data Mining

Penentuan jumlah *cluster* optimal pada penelitian ini dilakukan menggunakan *Elbow Method*. Metode ini digunakan untuk menentukan jumlah *cluster* terbaik dengan melihat perubahan nilai *Within Cluster Sum of Squares* (WCSS) atau *Sum of Squared Error* (SSE) pada setiap penambahan jumlah *cluster*. Nilai WCSS menunjukkan jumlah variasi data terhadap pusat *cluster* (*centroid*). Semakin kecil nilai WCSS, maka semakin baik tingkat homogenitas data dalam suatu *cluster*.

Prinsip Elbow Method adalah memilih jumlah cluster pada titik yang membentuk pola menyerupai siku (elbow), yaitu ketika penurunan nilai WCSS mulai melambat secara signifikan. Titik tersebut menunjukkan bahwa penambahan jumlah cluster berikutnya tidak lagi memberikan peningkatan yang berarti terhadap kualitas pengelompokan.



Gambar 2. Metode Elbow Dengan Within Cluster Sum of Square (WCSS)

Penelitian ini memilih jumlah lima klaster (K = 5) sebagai kompromi terbaik antara kualitas model dan kemudahan interpretasi, karena setelah titik tersebut penurunan nilai WCSS sudah tidak lagi signifikan. Penggunaan lima klaster ini menghasilkan pembagian wilayah kecamatan yang lebih rinci berdasarkan tingkat volume distribusinya, yaitu kategori sangat rendah, rendah, sedang, tinggi, dan sangat tinggi. Untuk menjaga stabilitas model, digunakan parameter random_state=42 agar

hasil pengelompokan tetap konsisten setiap kali program dijalankan, serta n_init=10 agar algoritma melakukan 10 kali percobaan inisialisasi *centroid* untuk memilih hasil yang terbaik.

Selanjutnya dilakukan proses pembentukan cluster menggunakan fungsi fit_predict(). Fungsi tersebut digunakan untuk menjalankan proses K-Means sekaligus menghasilkan label cluster pada setiap data. Algoritma akan menghitung jarak antar data terhadap titik pusat cluster (centroid) menggunakan Euclidean Distance, kemudian menempatkan data ke dalam cluster dengan jarak terdekat. Setelah proses clustering selesai, dilakukan perhitungan ratarata volume pada masing-masing cluster yang bertujuan untuk mengetahui urutan tingkat volume pada setiap cluster. Pengurutan dilakukan dari nilai rata-rata terkecil hingga terbesar agar cluster dapat diinterpretasikan secara lebih mudah. Kemudian dilakukan proses mapping atau pemberian label kategori, Label numerik hasil K-Means seperti 0, 1, 2, 3, dan 4 tidak memiliki arti tingkatan tertentu. Oleh karena itu, dilakukan pemberian label berdasarkan rata-rata volume masing-masing cluster agar hasil pengelompokan lebih mudah dipahami

```
# Mapping cluster
label_cluster = {
    cluster_mean.index[0]: 'Sangat Rendah',
    cluster_mean.index[1]: 'Rendah',
    cluster_mean.index[2]: 'Sedang',
    cluster_mean.index[3]: 'Tinggi',
    cluster_mean.index[4]: 'Sangat Tinggi'
}

data_cluster['Kategori'] = (
    data_cluster['Cluster']
    .map(label_cluster)
)
```

Gambar 3. Source Code untuk Labeling Kategori Cluster

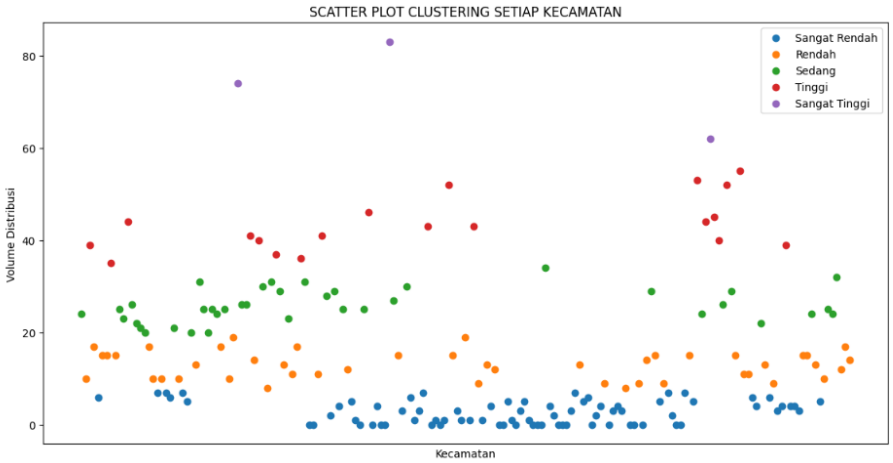
Terakhir, dilakukan pengurutan data berdasarkan volume distribusi. Pengurutan ini dilakukan untuk mempermudah proses analisis hasil sehingga kecamatan dengan volume distribusi tertinggi dapat terlihat dengan lebih jelas.

```
=== HASIL CLUSTERING K-MEANS ===
              Kota                 Kecamatan  Volume       Kategori
73      Kab. Bekasi   Tambun Selatan, Bekasi      83  Sangat Tinggi
37    Jakarta Timur    Cakung, Jakarta Timur      74  Sangat Tinggi
149     Kota Bekasi     Bekasi Utara, Bekasi      62  Sangat Tinggi
156     Kota Bekasi        Rawalumbu, Bekasi      55         Tinggi
146     Kota Bekasi     Bekasi Barat, Bekasi      53         Tinggi
..              ...                      ...     ...           ...
142  Kab. Tangerang     Sukamulya, Tangerang       0  Sangat Rendah
125  Kab. Tangerang        Kemiri, Tangerang       0  Sangat Rendah
130  Kab. Tangerang          Mauk, Tangerang       0  Sangat Rendah
133  Kab. Tangerang      Pakuhaji, Tangerang       0  Sangat Rendah
121  Kab. Tangerang  Gunung Kaler, Tangerang       0  Sangat Rendah

[183 rows x 4 columns]

=== Jumlah Kecamatan per Cluster ===
Kategori
Sangat Rendah    77
Rendah           46
Sedang           38
Tinggi           19
Sangat Tinggi     3
Name: count, dtype: int64
```

Gambar 4. Hasil Clustering



Gambar 5. Scatter Plot Clustering

Hasil akhir berupa pengelompokan kecamatan berdasarkan tingkat volume distribusi yang dapat digunakan untuk mendukung pengambilan keputusan distribusi.

Cluster Sangat Rendah memiliki nilai centroid sebesar 2,71. Nilai tersebut menunjukkan bahwa kecamatan pada kelompok ini memiliki rata-rata volume distribusi sekitar tiga paket selama periode penelitian. Kecamatan dalam kelompok ini merupakan wilayah dengan aktivitas distribusi paling rendah dibandingkan cluster lainnya. Hal ini mengindikasikan bahwa permintaan pengiriman paket di wilayah tersebut relatif kecil.

Cluster Rendah memiliki centroid sebesar 12,91. Artinya rata-rata volume distribusi pada kelompok ini berada di sekitar 13 paket. Kecamatan dalam cluster ini memiliki aktivitas distribusi yang lebih tinggi dibandingkan kategori Sangat Rendah, namun masih berada di bawah rata-rata keseluruhan data.

Cluster Sedang mempunyai centroid sebesar 25,82. Nilai ini menunjukkan bahwa rata-rata volume distribusi kecamatan dalam kelompok tersebut berada pada kisaran 26 paket. Kecamatan dalam cluster ini dapat dianggap sebagai wilayah dengan tingkat distribusi yang stabil sehingga dapat dijadikan dasar dalam perencanaan kapasitas operasional harian.

Cluster Tinggi memiliki centroid sebesar 43,42. Kecamatan dalam kelompok ini mempunyai rata-rata volume distribusi sekitar 43 paket. Volume tersebut menunjukkan bahwa aktivitas pengiriman relatif tinggi sehingga wilayah pada 35 cluster ini membutuhkan perhatian yang lebih besar dalam pengelolaan kapasitas distribusi dibandingkan cluster sebelumnya.

Cluster Sangat Tinggi memiliki centroid sebesar 73,00, yang merupakan nilai centroid terbesar di antara seluruh cluster. Nilai ini menunjukkan bahwa kecamatan pada kelompok tersebut memiliki rata-rata volume distribusi paling tinggi. Kecamatan seperti Tambun Selatan, Cakung, dan Bekasi Utara berada pada kelompok ini karena memiliki volume pengiriman yang jauh lebih besar dibandingkan kecamatan lainnya. Dengan demikian, cluster ini dapat diprioritaskan sebagai wilayah utama dalam penyusunan strategi distribusi paket.

Berdasarkan nilai centroid yang diperoleh, terlihat adanya kenaikan nilai ratarata volume distribusi secara bertahap dari 2,71 hingga 73,00. Hal ini menunjukkan bahwa algoritma K-Means berhasil memisahkan kecamatan ke dalam kelompok yang memiliki karakteristik volume distribusi yang relatif homogen di dalam cluster, namun memiliki perbedaan yang cukup jelas antarkluster. Semakin besar nilai centroid, semakin tinggi

rata-rata volume distribusi pada kelompok tersebut. Oleh karena itu, centroid dapat digunakan sebagai representasi tingkat aktivitas distribusi pada masing-masing kategori dan menjadi dasar dalam pemberian label Sangat Rendah, Rendah, Sedang, Tinggi, dan Sangat Tinggi.

Interpretation / Evaluation

Evaluasi hasil clustering pada penelitian ini dilakukan menggunakan metode Silhouette Coefficient dan Davies–Bouldin Index (DBI) untuk mengetahui kualitas pengelompokan yang dihasilkan oleh algoritma K-Means dengan jumlah cluster sebanyak 5 cluster (K=5).

Berdasarkan hasil pengujian, diperoleh nilai Silhouette Score sebesar 0,6313 pada K=5. Nilai tersebut menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik karena berada mendekati nilai 1. Semakin tinggi nilai Silhouette Score, maka semakin baik tingkat kedekatan data pada cluster yang sama (cohesion) serta semakin besar perbedaan dengan cluster lainnya (separation). Nilai sebesar 0,6313 menunjukkan bahwa sebagian besar data telah berada pada cluster yang sesuai dan memiliki tingkat pemisahan yang cukup jelas antar kelompok.

```
from sklearn.metrics import silhouette_score

silhouette_5_clusters = silhouette_score(X_scaled, data_cluster['Cluster'])

print(f"Silhouette Score : {silhouette_5_clusters}")
```

```
Silhouette Score : 0.6312976576615115
```

Gambar 6. Source Code dan Hasil Perhitungan Silhoutte Score

Selain itu, hasil evaluasi menggunakan Davies–Bouldin Index (DBI) menunjukkan hasil nilai 0,4336 pada K=5. Nilai DBI yang lebih kecil menunjukkan kualitas cluster yang lebih baik karena mengindikasikan tingkat kemiripan antar cluster yang rendah serta tingkat homogenitas data dalam cluster yang tinggi. Nilai DBI sebesar 0,4336 menunjukkan bahwa hasil pengelompokan pada penelitian ini memiliki tingkat pemisahan antar cluster yang cukup baik.

```
from sklearn.metrics import davies_bouldin_score

dbi_score = davies_bouldin_score(X_scaled, data_cluster['Cluster'])
print(f"Davies-Bouldin Index: {dbi_score:.5f}")
```

```
Davies-Bouldin Index: 0.43355
```

Gambar 7. Source Code dan Hasil Perhitungan Davies-Bouldin Index (DBI)

Dengan demikian, hasil evaluasi menggunakan Silhouette Coefficient dan Davies–Bouldin Index menunjukkan bahwa penerapan K=5 mampu menghasilkan kualitas clustering yang baik dan dapat digunakan untuk mengelompokkan volume distribusi kecamatan ke dalam kategori Sangat Rendah, Rendah, Sedang, Tinggi, dan Sangat Tinggi

PEMBAHASAN

Penelitian ini dilakukan untuk menganalisis pola volume distribusi pada setiap kecamatan di wilayah Jabodetabek menggunakan metode K-Means Clustering. Penerapan metode data mining bertujuan untuk mengidentifikasi karakteristik distribusi berdasarkan tingkat volume sehingga dapat menghasilkan informasi yang lebih mudah dipahami serta mendukung proses pengambilan keputusan terkait strategi distribusi.

Data Understanding

Tahapan penelitian diawali dengan proses pengumpulan dan pemahaman data. Dataset yang digunakan terdiri dari data volume distribusi pada tingkat kecamatan di wilayah Jabodetabek dengan total 183 data kecamatan. Pada tahap pemeriksaan data ditemukan beberapa nilai kosong pada atribut wilayah akibat format merged cells pada file Excel. Selain itu ditemukan 24 data dengan nilai volume sebesar 0 atau sekitar 13,04% dari total data. Nilai nol tersebut dipertahankan dalam penelitian karena dianggap merepresentasikan kondisi distribusi yang sebenarnya, yaitu wilayah yang tidak memiliki aktivitas distribusi selama periode pengamatan.

Data Preparation

Tahap ini meliputi pembersihan dan transformasi data. Proses pembersihan dilakukan dengan mengisi nilai kosong menggunakan metode forward fill agar setiap kecamatan tetap memiliki informasi wilayah yang lengkap. Setelah itu dilakukan agregasi data berdasarkan kota dan kecamatan untuk memperoleh total volume distribusi pada masing-masing wilayah. Tahap berikutnya dilakukan normalisasi menggunakan metode StandardScaler agar data memiliki skala yang seragam sehingga tidak terjadi dominasi nilai tertentu pada proses perhitungan jarak dalam algoritma K-Means.

Clustering

Berdasarkan hasil evaluasi dan pertimbangan interpretasi, penelitian ini menggunakan jumlah cluster sebanyak lima kelompok (K=5). Kelima kelompok tersebut diklasifikasikan menjadi kategori Sangat Rendah, Rendah, Sedang, Tinggi, dan Sangat Tinggi. Penggunaan lima cluster dipilih karena mampu memberikan pengelompokan yang lebih rinci dibandingkan penggunaan jumlah cluster yang lebih sedikit.

Hasil clustering menunjukkan bahwa persebaran volume distribusi antar kecamatan memiliki perbedaan yang cukup signifikan. Kecamatan yang termasuk kategori Sangat Tinggi memiliki volume distribusi paling besar dibandingkan wilayah lainnya. Beberapa kecamatan yang termasuk kategori tersebut antara lain Tambun Selatan dengan volume distribusi sebesar 83, Cakung sebesar 74, dan Bekasi Utara sebesar 62. Wilayah-wilayah tersebut menunjukkan tingkat aktivitas distribusi yang tinggi dan dapat dikategorikan sebagai pusat distribusi utama.

Pada kategori Tinggi, terdapat beberapa kecamatan seperti Rawalumbu, Bekasi Barat, Cileungsi, dan Mustikajaya dengan volume distribusi berkisar antara 35–55. Wilayah pada kelompok ini menunjukkan aktivitas distribusi yang cukup besar, namun masih berada di bawah kelompok Sangat Tinggi.

Kategori Sedang terdiri dari kecamatan dengan volume distribusi yang relatif stabil pada rentang menengah. Sementara itu kategori Rendah dan Sangat Rendah didominasi oleh wilayah dengan volume distribusi yang kecil, termasuk kecamatan

dengan volume distribusi bernilai nol. Keberadaan nilai nol pada kelompok Sangat Rendah menunjukkan bahwa beberapa wilayah tidak memiliki aktivitas distribusi selama periode penelitian.

Untuk mengetahui kualitas hasil pengelompokan, dilakukan evaluasi menggunakan Silhouette Coefficient dan Davies–Bouldin Index (DBI). Hasil evaluasi pada penggunaan K=5 menunjukkan nilai Silhouette Score sebesar 0,6313 yang mengindikasikan bahwa data dalam masing-masing cluster memiliki tingkat kemiripan yang cukup tinggi dan pemisahan antar cluster berjalan dengan baik. Selain itu diperoleh nilai DBI sebesar 0,4336 yang menunjukkan bahwa tingkat kemiripan antar 46 cluster relatif rendah sehingga hasil pengelompokan memiliki kualitas yang cukup baik.

Secara keseluruhan, hasil penelitian menunjukkan bahwa metode K-Means Clustering mampu mengelompokkan wilayah berdasarkan karakteristik volume distribusi secara efektif. Hasil pengelompokan yang diperoleh dapat digunakan sebagai dasar untuk mengetahui pola persebaran distribusi, menentukan prioritas wilayah, serta membantu penyusunan strategi distribusi yang lebih efektif dan efisien.

PENUTUP

Simpulan

Berdasarkan hasil penelitian, algoritma K-Means Clustering berhasil mengelompokkan wilayah berdasarkan volume distribusi paket Lion Parcel di Jabodetabek melalui tahapan Knowledge Discovery in Databases (KDD). Penentuan jumlah cluster menggunakan Elbow Method menghasilkan lima cluster dengan kualitas yang baik, ditunjukkan oleh nilai Silhouette Coefficient sebesar 0,6313 dan Davies–Bouldin Index (DBI) sebesar 0,4336. Kelima cluster tersebut dikategorikan menjadi Sangat Rendah, Rendah, Sedang, Tinggi, dan Sangat Tinggi, yang mencerminkan perbedaan karakteristik volume distribusi pada setiap wilayah, sehingga mampu memberikan gambaran yang jelas mengenai tingkat aktivitas distribusi paket di masing-masing kecamatan.

Hasil pemetaan wilayah berdasarkan cluster menunjukkan bahwa persebaran volume distribusi paket di wilayah Jabodetabek tidak merata. Wilayah dengan volume distribusi tinggi terkonsentrasi pada beberapa kecamatan tertentu, sedangkan sebagian besar wilayah berada pada kategori sedang hingga sangat rendah. Hasil pemetaan ini dapat dimanfaatkan sebagai informasi pendukung dalam menentukan prioritas wilayah distribusi, pengalokasian sumber daya, serta penyusunan strategi distribusi yang lebih efektif dan efisien.

Saran

Penelitian selanjutnya disarankan untuk menggunakan variabel tambahan selain volume distribusi, seperti jumlah pelanggan, jarak dan waktu pengiriman, kepadatan wilayah, maupun jumlah transaksi, agar hasil pengelompokan dapat menggambarkan kondisi distribusi secara lebih komprehensif. Selain itu, penelitian juga dapat membandingkan algoritma K-Means dengan metode clustering lain, seperti Hierarchical Clustering, DBSCAN, atau Fuzzy C-Means, serta menggunakan data dengan periode yang lebih panjang untuk memperoleh pola dan tren distribusi yang lebih akurat.

Hasil pengelompokan pada penelitian ini juga dapat dikembangkan dalam bentuk dashboard atau sistem pemetaan geografis (*mapping system*) untuk

mempermudah visualisasi, analisis, dan pengambilan keputusan. Selain itu, hasil penelitian diharapkan dapat dimanfaatkan oleh perusahaan sebagai dasar dalam menentukan prioritas wilayah distribusi, mengalokasikan sumber daya, serta menyusun strategi operasional yang lebih efektif dan efisien.

REFERENSI

Adhitama, L., & Wibisono, M. A. (2024). *Perencanaan alokasi dan rute distribusi beras Kota Yogyakarta dengan metode K-Means clustering, Tabu Search dan Ant Colony System*. Logistik, 17(2). https://doi.org/10.21009/logistik.v17i02.37436.

Danjiah, M., Sofitra, M., & Batubara, H. (2025). *Evaluation of last-mile delivery depot locations for third-party logistics companies using the K-Means algorithm*. Jurnal Manajemen Industri dan Logistik. https://doi.org/10.30988/jmil.v9i1.1574.

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909.

Dias, J. L., Sott, M. K., Ferrão, C. C., Furtado, J. C., & Moraes, J. A. R. (2021). *Data mining and knowledge discovery in databases for urban solid waste management: A scientific literature review*. Waste Management & Research, 39(11), 1331–1340. https://doi.org/10.1177/0734242X211042276.

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Jia, H. (2023). *K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data*. *Information Sciences, 622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139.

Maori, N. A., & Evanita. (2023). *Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering*. Simetris: Jurnal Teknik Mesin, Industri, Elektro dan Ilmu Komputer, 14(2). DOI: https://doi.org/10.24176/simet.v14i2.9630.

Mauludin, M.R., Nurdiawan, O., dan Basysyar, F.M. 2024. Penerapan Algoritma K-Means Clustering untuk Analisis Kinerja Pengiriman Paket Shopee Express di Hub Transit Kedawung. *Jurnal Informatika dan Teknik Elektro Terapan*.

Olivia, L. (2023). *Optimalisasi pengiriman barang dengan mengelompokkan titik pengiriman menggunakan metode K-Means clustering dan hierarchical clustering*. Jurnal Komputer dan Informatika.

Setyaningtyas, S., Nugroho, B. I., & Arif, Z. (2022). *Tinjauan pustaka sistematis pada data mining: Studi kasus algoritma K-Means clustering*. Jurnal Teknoif Teknik Informatika Institut Teknologi Padang, 10(2), 52–61. https://doi.org/10.21063/jtif.2022.V10.2.52-61.

Setyawan, R., & Murtiyasa, B. (2025). *Systematic literature review: A comparison of clustering methods in data mining*. Journal of Computer Networks, Architecture and High Performance Computing, 8(1). https://doi.org/10.47709/cnahpc.v8i1.7333.

Siahaan, A. S., Saragih, R., & Simanjuntak, M. (2024). *Penerapan metode K-Means clustering untuk pengelompokan minat konsumen terhadap pengguna jasa layanan pada Kantor Pos Binjai*. Polygon: Jurnal Ilmu Komputer dan Ilmu Pengetahuan Alam. https://doi.org/10.62383/polygon.v2i5.236.