

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS KINERJA PENGIRIMAN PAKET SHOPEE EXPRESS DI HUB TRANSIT KEDAWUNG

Muhammad Rifqi Mauludin^{1*}, Odi Nurdiawan², Fadhil Muhammad Basysyar³

^{1,2}STIMIK IKMI CIREBON; Jl. Perjuangan No. 10 B Majasem Kota Cirebon No Tlp. 0231490480

Received: 18 Desember 2024

Accepted: 14 Januari 2025

Published: 20 Januari 2025

Kata Kunci : K-Means Clustering, Kinerja Pengiriman, Shopee Express, Davies-Bouldin Index, Analisis Data

Keywords : K-Means Clustering, Delivery Performance, Shopee Express, Davies-Bouldin Index, Data Analysis

Correspondent Email:
qbollicimun@gmail.com

Abstrak- Penelitian ini bertujuan menganalisis kinerja pengiriman Shopee Express (SPX) di Hub Transit Kedawung menggunakan algoritma K-Means Clustering. Data sebanyak 359 pengiriman dengan 12 atribut dikumpulkan dari operator SPX. Model Knowledge Discovery in Databases (KDD) digunakan dalam penelitian, meliputi pemilihan data, pra-pemrosesan, transformasi, penerapan algoritma K-Means, dan evaluasi model menggunakan Davies-Bouldin Index (DBI). Tahapan pra-pemrosesan mencakup pembersihan data, pemilihan atribut relevan, dan normalisasi data, sementara transformasi dilakukan untuk mengubah atribut nominal menjadi numerik. Hasil evaluasi menunjukkan nilai DBI terbaik sebesar 0.288 dengan jumlah cluster optimal $K = 10$. Cluster 4 dan Cluster 6 menunjukkan performa terbaik dengan pengiriman tercepat, sedangkan Cluster 7 dan Cluster 9 memiliki tingkat on-hold tertinggi, disebabkan penerima tidak tersedia atau alamat tidak valid. Atribut seperti Driver ID, Zone ID, dan On-hold Reason menjadi faktor signifikan dalam pengelompokan. Penelitian ini memberikan wawasan bagi manajemen logistik SPX untuk meningkatkan efisiensi operasional dengan strategi seperti optimalisasi rute, peningkatan SOP, dan validasi alamat. Hasilnya diharapkan menjadi dasar untuk penerapan lebih lanjut algoritma clustering dalam manajemen logistik skala besar.

Abstract - This study analyzes the delivery performance of Shopee Express (SPX) at the Kedawung Transit Hub using the K-Means Clustering algorithm. A total of 359 shipment data points with 12 attributes were collected from SPX operators. The Knowledge Discovery in Databases (KDD) model was applied, including data selection, preprocessing, transformation, K-Means implementation, and evaluation using the Davies-Bouldin Index (DBI). Preprocessing involved data cleaning, attribute selection, and normalization, while transformation converted nominal attributes into numerical ones. The evaluation identified the optimal number of clusters ($K = 10$) with the best DBI score of 0.288. Cluster 4 and Cluster 6 demonstrated the best performance with the fastest delivery times, while Cluster 7 and Cluster 9 had the highest on-hold rates, often due to unavailable recipients or invalid addresses. Key attributes such as Driver ID, Zone ID, and On-hold Reason significantly influenced clustering. This research provides valuable insights for SPX logistics management to enhance operational efficiency through strategies like route optimization, SOP improvement, and address validation. The findings aim to serve as a foundation for further applications of clustering algorithms in large-scale logistics management.

1. PENDAHULUAN

Perkembangan pesat di bidang informatika telah mengubah berbagai aspek kehidupan, termasuk

teknologi, bisnis, dan pendidikan. Salah satu penerapan informatika yang signifikan adalah dalam analisis data menggunakan algoritma K-Means Clustering. Algoritma ini telah banyak digunakan untuk berbagai tujuan, seperti segmentasi pasar, analisis perilaku pengguna, hingga pengelompokan data pendidikan untuk memahami pola dan tren [1]

Dalam konteks logistik, K-Means Clustering efektif untuk mengelompokkan data pengiriman berdasarkan atribut spesifik, seperti waktu pengiriman, lokasi, dan jenis paket. Sebagai contoh, penelitian oleh [2] menunjukkan bahwa digitalisasi dan analisis data berbasis clustering memberikan kontribusi besar pada manajemen logistik. Penelitian lain oleh [3] mengembangkan condensed silhouette untuk meningkatkan proses penentuan jumlah cluster. Sementara itu, [4] menggabungkan K-Means dengan teknik lain untuk segmentasi dan identifikasi pola dalam data yang kompleks.

Meskipun algoritma K-Means telah diterapkan dalam berbagai studi, penerapan spesifiknya untuk analisis kinerja pengiriman paket Shopee Express (SPX) di Hub Transit Kedawung masih jarang ditemukan. Sebagian besar penelitian lebih terfokus pada sektor pendidikan atau kesehatan [5]. Selain itu, atribut seperti alasan on-hold dan identitas pengemudi, yang secara signifikan memengaruhi efisiensi operasional, belum dieksplorasi secara mendalam. Oleh karena itu, penelitian ini menawarkan kebaruan dalam mengaplikasikan K-Means Clustering untuk mengelompokkan data logistik dan menganalisis atribut yang relevan secara lebih komprehensif.

Penelitian ini bertujuan untuk menganalisis kinerja pengiriman paket Shopee Express (SPX) di Hub Transit Kedawung dengan menerapkan algoritma K-Means Clustering. Tujuan utama penelitian meliputi menentukan jumlah cluster optimal menggunakan Davies-Bouldin Index (DBI), mengidentifikasi atribut-atribut signifikan yang memengaruhi pengelompokan data pengiriman, seperti alasan paket tertahan (on-hold) dan identitas pengemudi, serta mengevaluasi karakteristik dari setiap cluster yang terbentuk. Hasil dari penelitian ini diharapkan dapat memberikan wawasan strategis kepada manajemen logistik Shopee Express dalam meningkatkan efisiensi operasional, mengoptimalkan proses pengiriman, serta meminimalkan kendala yang terjadi dalam distribusi paket. Penelitian ini bertujuan untuk menganalisis kinerja pengiriman paket Shopee Express (SPX) di Hub Transit Kedawung dengan menerapkan algoritma K-Means Clustering. Tujuan utama penelitian meliputi menentukan jumlah cluster optimal menggunakan Davies-Bouldin Index (DBI), mengidentifikasi atribut-atribut signifikan yang memengaruhi pengelompokan data pengiriman, seperti alasan paket tertahan (on-hold) dan identitas

pengemudi, serta mengevaluasi karakteristik dari setiap cluster yang terbentuk. Hasil dari penelitian ini diharapkan dapat memberikan wawasan strategis kepada manajemen logistik Shopee Express dalam meningkatkan efisiensi operasional, mengoptimalkan proses pengiriman, serta meminimalkan kendala yang terjadi dalam distribusi paket.

2. TINJAUAN PUSTAKA

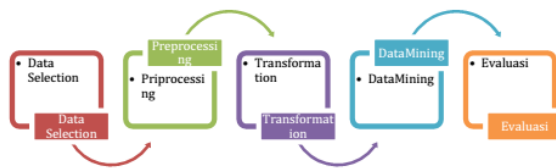
K-Means Clustering adalah algoritma pengelompokan data yang digunakan untuk membagi dataset ke dalam sejumlah cluster berdasarkan kesamaan atribut. Algoritma ini bekerja dengan menentukan centroid secara iteratif hingga mencapai hasil optimal. Penelitian [6] menunjukkan bahwa K-Means efektif dalam mengelompokkan data dan dapat dievaluasi menggunakan Davies-Bouldin Index (DBI) untuk menentukan kualitas cluster. Namun, pemilihan jumlah cluster (K) menjadi tantangan karena secara langsung memengaruhi hasil pengelompokan.

DBI adalah metrik evaluasi yang digunakan untuk mengukur kualitas cluster dalam algoritma K-Means. Nilai DBI dihitung berdasarkan rasio antara jarak antar cluster dan kepadatan intra-cluster. Semakin kecil nilai DBI, semakin baik kualitas cluster. Penelitian [3] menyoroti pentingnya DBI dalam menentukan jumlah cluster optimal untuk mencapai hasil yang lebih akurat.

Model KDD merupakan kerangka kerja yang sistematis untuk menemukan pola dalam dataset besar. Model ini terdiri dari tahapan seperti pemilihan data, pra-pemrosesan, transformasi data, data mining, dan evaluasi hasil. Dalam penelitian ini, model KDD digunakan untuk mengelola data pengiriman Shopee Express secara terstruktur sehingga analisis dapat dilakukan dengan optimal. Penelitian oleh [2] menunjukkan bahwa digitalisasi dalam manajemen logistik mampu meningkatkan efisiensi operasional melalui pengelompokan data. Sementara itu, penelitian [7] menggunakan K-Means untuk menganalisis tren penjualan di berbagai wilayah, yang relevan dengan pengelompokan data logistik berdasarkan atribut pengiriman.

3. METODE PENELITIAN

Penelitian ini menggunakan model *Knowledge Discovery in Databases (KDD)* dengan pendekatan algoritma *K-Means clustering* untuk menganalisis kinerja pengiriman paket Shopee Express (SPX) di HUB Transit Kedawung.



Gambar 1 Tahapan KDD

3.1 Rancangan Penelitian

Penelitian ini dirancang menggunakan model Knowledge Discovery in Databases (KDD), yang terdiri dari lima tahap utama:

1. **Pemilihan Data:** Tahap awal yang bertujuan untuk menentukan data relevan dari database Shopee Express (SPX) Hub Transit Kedawung, seperti waktu penerimaan, waktu pengiriman, alasan paket tertahan (*on-hold*), dan identitas pengemudi.
2. **Pra-Pemrosesan Data:** Proses pembersihan data dilakukan dengan menghapus data duplikat, menangani nilai yang hilang, dan normalisasi atribut agar sesuai dengan input algoritma.
3. **Transformasi Data:** Data nominal seperti *on-hold reason* dan *driver ID* dikonversi menjadi bentuk numerik untuk keperluan clustering.
4. **Penerapan Algoritma K-Means:** Algoritma ini diterapkan untuk membentuk cluster berdasarkan atribut yang telah ditransformasi.
5. **Evaluasi Hasil:** Davies-Bouldin Index (DBI) digunakan untuk menentukan kualitas cluster dan jumlah cluster optimal.

3.2 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh melalui operator Shopee Express di Hub Transit Kedawung. Data mencakup 359 entri pengiriman selama satu bulan dengan 12 atribut, termasuk waktu pengiriman, alasan *on-hold*, dan ID pengemudi. Data dikumpulkan dalam format file Excel dan disesuaikan dengan kebutuhan analisis.

3.3 Sumber Data

Sumber data berasal dari sistem operasional Shopee Express Hub Transit Kedawung. Data ini menggambarkan kinerja pengiriman dalam periode tertentu dan mencakup seluruh proses, mulai dari penerimaan paket hingga pengiriman ke penerima.

3.4 Teknik Analisis Data

Proses analisis data dilakukan melalui tahapan berikut:

1. **Pra-Pemrosesan:** Dataset yang diperoleh diproses untuk memastikan kelengkapan

dan akurasi data. Normalisasi dilakukan agar atribut memiliki skala yang seragam.

2. **Penerapan Algoritma K-Means:** Algoritma K-Means digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan atribut yang relevan. Jumlah cluster optimal ditentukan melalui nilai DBI.
3. **Evaluasi Hasil Clustering:** Evaluasi dilakukan menggunakan metrik DBI untuk menilai kualitas cluster. Semakin kecil nilai DBI, semakin baik pemisahan antar cluster.

3.5 Implementasi Algoritma

Algoritma K-Means diterapkan menggunakan perangkat lunak AI Studio versi 2024.1.0. Setiap atribut data yang relevan ditetapkan perannya, seperti *driver ID* sebagai identitas pengemudi dan *on-hold reason* sebagai penyebab keterlambatan. Tahap ini menghasilkan 10 cluster, dengan nilai DBI terbaik sebesar 0.288 yang menunjukkan kualitas cluster yang baik.

3.6 Visualisasi Hasil

Proses clustering divisualisasikan melalui grafik hasil DBI untuk memudahkan interpretasi pola dan karakteristik dari setiap cluster yang terbentuk.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis data yang diperoleh dari penerapan algoritma K-Means Clustering pada data pengiriman Shopee Express (SPX) di Hub Transit Kedawung. Analisis difokuskan pada pembentukan cluster, evaluasi kualitas cluster menggunakan Davies-Bouldin Index (DBI), serta karakteristik yang dihasilkan dari masing-masing cluster. Selain itu, pembahasan akan menginterpretasikan temuan penelitian, menjawab permasalahan yang diajukan, membandingkan hasil dengan penelitian sebelumnya, serta menjelaskan implikasi teoritis dan implementasi praktis dari hasil penelitian ini.

Tabel 1 Nilai DBI Setiap Cluster

K-means	
K	DBI
2	0.536
3	0.520
4	0.434
5	0.314
6	0.333

7	0.319
8	0.328
9	0.375
10	0.288

4.1 Hasil Penelitian

A. Clustering dan Evaluasi Model

Proses clustering menghasilkan 10 cluster optimal berdasarkan analisis Davies-Bouldin Index (DBI). Nilai DBI terbaik tercatat sebesar 0.288, yang menunjukkan kualitas cluster yang sangat baik dengan pemisahan antar cluster yang signifikan dan homogenitas dalam cluster. Hasil clustering disajikan dalam Tabel 4.1 di bawah ini:

Tabel 2 Nilai DBI Setiap Cluster

Cluster	Jumlah Data	Karakteristik Utama
0	54	Waktu on-hold rendah; penyebab keterlambatan sedikit.
4	39	Waktu pengiriman tercepat; tingkat keterlambatan rendah.
6	41	Efisiensi tinggi, minimal alasan on-hold.
7	37	Waktu on-hold tertinggi; banyak alasan penerima tidak tersedia.
9	29	Paket sering tertahan dengan alasan alamat tidak valid.

B. Visualisasi Evaluasi DBI

Hasil DBI untuk masing-masing jumlah cluster digambarkan dalam Tabel 1. Tabel ini menunjukkan bahwa nilai DBI terus menurun hingga mencapai nilai optimal pada $K = 10$.

C. Atribut Penting Dalam Clustering

Atribut yang memiliki peran signifikan dalam pembentukan cluster adalah Driver ID, Zone ID, dan On-hold Reason. Atribut-atribut ini memberikan kontribusi utama dalam membedakan cluster berdasarkan kinerja pengiriman.

4.2 Pembahasan

Setiap cluster yang dihasilkan memiliki karakteristik tersendiri. Cluster 4 dan Cluster 6 mencatat performa terbaik, dengan waktu pengiriman yang lebih singkat dan frekuensi keterlambatan yang rendah. Hal ini menunjukkan

bahwa manajemen waktu yang efektif dan minimnya hambatan operasional menjadi kunci keberhasilan dalam kelompok ini. Sebaliknya, Cluster 7 dan Cluster 9 memiliki performa paling rendah, ditandai dengan waktu on-hold tertinggi dan frekuensi alasan keterlambatan yang dominan. Alasan utama keterlambatan pada cluster ini adalah alamat penerima yang tidak valid, penerima tidak tersedia, serta keterbatasan waktu pengiriman. Hal ini mengindikasikan perlunya perbaikan dalam validasi alamat pengiriman dan koordinasi yang lebih baik antara kurir dan pelanggan.

Selain itu, atribut Driver ID dan Zone ID menjadi faktor yang paling berpengaruh dalam pembentukan cluster. Perlu diperhatikan pentingnya peran pengemudi dan wilayah operasional dalam menentukan efisiensi pengiriman. Dengan memahami atribut-atribut kunci ini, manajemen dapat merancang strategi yang lebih spesifik, seperti mengalokasikan driver berpengalaman ke wilayah dengan performa rendah atau mengoptimalkan rute pengiriman untuk meminimalkan waktu tunggu.

Hasil penelitian ini selaras dengan studi terdahulu seperti yang dilakukan oleh [8] dan [6] yang menunjukkan bahwa algoritma *K-Means* efektif dalam mengelompokkan data operasional untuk memberikan wawasan yang lebih dalam. Namun, penelitian ini juga menekankan pentingnya tahapan pra-pemrosesan data dan evaluasi cluster untuk memastikan hasil clustering yang akurat.

Secara keseluruhan, penelitian ini membuktikan bahwa algoritma *K-Means* dapat digunakan sebagai alat analisis kinerja operasional logistik. Temuan ini dapat menjadi dasar bagi manajemen Shopee Express dalam mengambil langkah-langkah strategis untuk meningkatkan efisiensi pengiriman, seperti optimalisasi SOP pengiriman, pelatihan driver, serta pemanfaatan teknologi dalam memvalidasi alamat penerima.

5. KESIMPULAN

Berdasarkan hasil penelitian ini, penerapan algoritma *K-Means Clustering* dalam menganalisis kinerja pengiriman paket Shopee Express (SPX) di Hub Transit Kedawung menunjukkan efektivitas yang signifikan. Berikut adalah tujuan yang di capai pada penelitian ini :

1. Nilai *Davies Bouldin Index* (DBI) sebesar 0.288 mengindikasikan bahwa jumlah cluster optimal yang terbentuk adalah 10 cluster, dengan kualitas pemisahan yang baik antar cluster.
2. Atribut Driver ID, Zone ID, dan On-Hold Reason terbukti memiliki pengaruh signifikan dalam pengelompokan data, menegaskan peran penting validasi alamat,

efisiensi pengemudi, serta manajemen operasional dalam menentukan kualitas pengiriman.

3. Setiap cluster memiliki karakteristik spesifik yang mengidentifikasi perbedaan performa pengiriman. *Cluster 4* dan *Cluster 6* menunjukkan performa optimal dengan waktu pengiriman yang lebih cepat dan tingkat keterlambatan yang rendah. Sebaliknya, *Cluster 7* dan *Cluster 9* memiliki performa rendah yang ditandai dengan waktu *on-hold* tertinggi dan penyebab keterlambatan dominan, seperti alamat penerima yang tidak valid serta penerima yang tidak tersedia.

Dengan demikian, hasil penelitian ini menunjukkan bahwa *algoritma K-Means Clustering* dapat menjadi alat yang efektif untuk mengidentifikasi pola kinerja pengiriman paket dan memberikan dasar bagi perbaikan operasional logistik.

UCAPAN TERIMA KASIH

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas rahmat dan hidayah-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga ingin menyampaikan rasa terima kasih yang mendalam kepada pihak-pihak yang telah memberikan dukungan, bimbingan, dan kontribusi selama proses penelitian ini:

1. **Bapak Odi Nurdiawan, M.Kom** selaku Dosen Pembimbing Utama dan **Bapak Fadhil Muhammad Basysyar, M.Kom** selaku Dosen Pembimbing Kedua, yang telah memberikan arahan, masukan, dan motivasi dalam setiap tahap penyusunan penelitian ini.
2. **STMIK IKMI Cirebon**, khususnya Program Studi Teknik Informatika, yang telah memberikan fasilitas dan kesempatan untuk melaksanakan penelitian ini.
3. Tim operasional Shopee Express Hub Transit Kedawung, yang telah menyediakan data penelitian dan mendukung kelancaran pengumpulan data.
4. Keluarga tercinta, terutama kedua orang tua, yang selalu memberikan doa, dukungan moral, dan semangat dalam menyelesaikan penelitian ini.
5. Rekan-rekan mahasiswa Teknik Informatika, yang telah memberikan dukungan dan berbagi pengalaman selama proses penyusunan skripsi ini.

Penulis menyadari bahwa penelitian ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang

membangun sangat diharapkan untuk perbaikan di masa depan. Semoga penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan peningkatan efisiensi operasional dalam bidang logistik.

DAFTAR PUSTAKA

- [1] M. M. Fard, T. Thonet, and E. Gaussier, "Pattern Recognition Letters Deep k-Means: Jointly clustering with k-Means and learning representations," *Elsevier*, no. 2016, pp. 1–8, 2020.
- [2] E. Calderon-Monge and D. Ribeiro-Soriano, *The role of digitalization in business and management: a systematic literature review*, vol. 18, no. 2. Springer Berlin Heidelberg, 2024.
- [3] A. Naghizadeh and D. N. Metaxas, "Condensed silhouette: An optimized filtering process for cluster selection in K-means," *Procedia Comput. Sci.*, vol. 176, pp. 205–214, 2020.
- [4] N. Altini *et al.*, "Segmentation and identification of vertebrae in ct scans using cnn, k-means clustering and k-nn," *Informatics*, vol. 8, no. 2, pp. 1–17, 2021.
- [5] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," *IEEE Access*, vol. 8, pp. 151847–151855, 2020.
- [6] M. Wahyudi, S. Solikhun, and L. Pujiastuti, "Komparasi K-Means Clustering dan K-Medoids Clustering dalam Mengelompokkan Produksi Susu Segar di Indonesia Berdasarkan Nilai DBI," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 243–254, 2022.
- [7] E. M. Fitri, R. R. Suryono, and A. Wantoro, "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz," *J. Komputasi*, vol. 11, no. 2, pp. 157–168, 2023.
- [8] A. Adiyanto and Y. Arie Wijaya, "Penerapan Algoritma K-Means Pada Pengelompokan Data Set Bahan Pangan Indonesia Tahun 2022-2023," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1344–1350, 2023.