

# Penerapan Algoritma *K-Means* Untuk Klusterisasi Penerima Dana Bantuan Pendidikan KJP Plus Pada DKI Jakarta

<sup>1</sup>Tops Fari Bima Pratama, <sup>2</sup>Rifiana Arief

<sup>1\*,2</sup>Magister Manajemen Sistem Informasi, Universitas Gunadarma, Depok, Indonesia

\*Korespondensi: [topsfaribima@gmail.com](mailto:topsfaribima@gmail.com)

Submit : 06 Agust 2026 | Diterima : 08 Okt 2026 | Terbit : 08 Okt 2026

## ABSTRACT

*This study aims to build a clustering model for recipients of the Jakarta Smart Card (KJP) Plus Phase 2 for the 2025 Fiscal Year based on their ownership of transportation assets and school status to support accurate targeting of educational aid distribution. The analysis was conducted in response to findings from the Audit Board of Indonesia (BPK RI) regarding the need to strengthen recipient verification. The research uses the K-Means algorithm on Orange software. The data consists of 707,513 KJP Plus recipients, which were integrated with 8,958,396 motor vehicle tax records by matching the guardian's National Identification Number (NIK), resulting in 398,590 recipients identified as owning vehicles. The variables used include vehicle weight and school status, with school status transformed using one-hot encoding. The optimal number of clusters was determined using the Silhouette coefficient and Davies-Bouldin Index (DBI). The results show that two clusters ( $k = 2$ ) are the best configuration with a Silhouette score of 0.836 and a DBI of 0.41. Cluster 1 contains 393,909 recipients (98.8%), mostly owners of light vehicles, especially motorcycles, while Cluster 2 contains 4,681 recipients (1.2%), mostly owners of four-wheeled vehicles and commercial vehicles. School status doesn't play a significant role in differentiating the clusters. The study concludes that owning four-wheeled and commercial vehicles is a stronger indicator for identifying potential ineligibility of aid recipients, so it should be prioritized in the verification process.*

**Keywords:** clustering, *K-Means*, KJP Plus, Silhouette coefficient, Orange.

## ABSTRAK

Penelitian ini bertujuan membangun model klusterisasi penerima Kartu Jakarta Pintar (KJP) Plus Tahap 2 Tahun Anggaran 2025 berdasarkan kepemilikan aset transportasi dan status sekolah guna mendukung ketepatan sasaran penyaluran bantuan pendidikan. Analisis dilakukan sebagai respons terhadap temuan BPK RI terkait perlunya penguatan verifikasi penerima bantuan. Penelitian menggunakan algoritma *K-Means* pada perangkat lunak Orange. Data terdiri atas 707.513 penerima KJP Plus yang diintegrasikan dengan 8.958.396 data pajak kendaraan bermotor melalui pencocokan Nomor Induk Kependudukan (NIK) wali, menghasilkan 398.590 penerima yang teridentifikasi memiliki kendaraan. Variabel yang digunakan meliputi bobot kendaraan dan status sekolah, dengan status sekolah ditransformasikan menggunakan one-hot encoding. Jumlah klaster optimal ditentukan menggunakan koefisien Silhouette dan Davies-Bouldin Index (DBI). Hasil menunjukkan bahwa dua klaster ( $k = 2$ ) merupakan konfigurasi terbaik dengan nilai Silhouette 0,836 dan DBI 0,41. Klaster 1 berisi 393.909 penerima (98,8%) yang didominasi pemilik kendaraan ringan, terutama sepeda motor, sedangkan Klaster 2 berisi 4.681 penerima (1,2%) yang didominasi pemilik kendaraan roda empat dan kendaraan niaga. Status sekolah tidak berperan signifikan dalam membedakan klaster. Penelitian menyimpulkan bahwa kepemilikan kendaraan roda empat dan niaga merupakan indikator yang lebih kuat untuk mengidentifikasi potensi ketidaklayakan penerima bantuan sehingga layak diprioritaskan dalam proses verifikasi.

**Kata Kunci:** Klusterisasi, *K-Means*, KJP Plus, Koefisien Silhouette, Orange.

## PENDAHULUAN

Implementasi program perlindungan sosial yang ideal menuntut pengawasan yang ketat demi menjamin pemenuhan asas keadilan, akuntabilitas, dan ketepatan sasaran. Berdasarkan Peraturan Gubernur Provinsi DKI Jakarta Nomor 44 Tahun 2022 tentang Pemberian Bantuan Sosial Dalam Rangka Perlindungan Sosial, sirkulasi penyaluran bantuan melibatkan multisektor yang kompleks, mulai dari tingkat kedinasan, aparat kelurahan, hingga institusi perbankan yang bertindak sebagai penyalur. Namun, kompleksitas birokrasi dan besarnya volume data di lapangan sering kali memicu berbagai hambatan, meliputi ketidakakuratan data basis penerima, keterlambatan distribusi, hingga potensi penyalahgunaan alokasi dana di tingkat implementasi (Azzahra, Musyaffi dan Khairunnisa, 2025).

Fenomena salah sasaran dalam distribusi jaminan sosial dipertegas oleh temuan audit. Hasil pemeriksaan Badan Pemeriksa Keuangan (BPK) RI atas Laporan Keuangan Pemerintah Provinsi DKI Jakarta Tahun 2023 menunjukkan bahwa akumulasi dana senilai Rp13.177.499.800,00 telah disalurkan kepada 5.623 penerima yang secara faktual tidak memenuhi kriteria melalui Dinas Sosial dan Dinas Pendidikan Provinsi DKI Jakarta. Salah satu pos penyimpangan terbesar terjadi pada program Kartu Jakarta Pintar (KJP) Plus Tahap 2, di mana bantuan biaya personal pendidikan didistribusikan kepada 4.341 penerima yang tidak memenuhi syarat, dengan taksiran penyaluran tidak sesuai kriteria senilai Rp10,27 miliar (BPK RI, 2024; Azzahra, Musyaffi dan Khairunnisa, 2025). Besaran temuan tersebut memperlihatkan bahwa persoalan ketepatan sasaran bukan sekadar isu administratif, melainkan telah berimplikasi langsung pada efisiensi fiskal Anggaran Pendapatan dan Belanja Daerah.

Program KJP Plus merupakan instrumen kebijakan strategis yang diinisiasi sejak tahun 2013 dan diperkuat melalui Peraturan Gubernur Provinsi DKI Jakarta Nomor 141 Tahun 2016 tentang Bantuan Biaya Personal Pendidikan Bagi Peserta Didik Dari Keluarga Tidak Mampu Melalui Kartu Jakarta Pintar. Kebijakan ini dirancang untuk mendukung wajib belajar dua belas tahun, meningkatkan mutu pendidikan, serta menjamin hak aksesibilitas pendidikan yang setara bagi seluruh lapisan masyarakat (Zikri, Ismanto dan Candradewini, 2022). Namun, pemrosesan data calon penerima KJP Plus saat ini masih bertumpu pada mekanisme seleksi manual administratif. Kelemahan inheren dari sistem konvensional ini adalah rendahnya kecepatan komputasi ketika dihadapkan pada basis data berdimensi besar, tingginya risiko kesalahan input, serta kerentanan terhadap bias subjektivitas pengambil keputusan. Akibatnya, terjadi diskrepansi distribusi di mana peserta didik dari keluarga mampu secara ekonomi lolos verifikasi, sedangkan kelompok masyarakat rentan yang membutuhkan justru tereliminasi dari sistem penyaluran.

Persoalan ketepatan sasaran pada dasarnya bermuara pada satu titik, yaitu bagaimana kriteria keluarga tidak mampu diterjemahkan menjadi parameter yang dapat diukur dan diverifikasi. Kemampuan ekonomi sebuah keluarga tidak hanya tecermin dari penghasilan yang dilaporkan, melainkan juga dari kepemilikan aset bernilai ekonomi yang relatif lebih stabil dan lebih sulit disembunyikan. Kepemilikan kendaraan bermotor merupakan salah satu penanda yang paling mudah ditelusuri karena tercatat secara resmi dalam basis data perpajakan daerah. Persoalannya, penelusuran indikator semacam ini menuntut pemeriksaan silang terhadap basis data instansi lain, suatu pekerjaan yang praktis mustahil dikerjakan secara manual terhadap ratusan ribu penerima dalam satu tahap penyaluran. Pada titik inilah teknik data mining menemukan relevansinya: alih-alih memeriksa satu per satu, seluruh penerima dipetakan sekaligus ke dalam kelompok-kelompok berdasarkan kemiripan karakteristik, sehingga sumber daya verifikasi dapat diarahkan pada kelompok yang paling berpotensi bermasalah.

Urgensi pemanfaatan teknologi dalam mereduksi kelemahan seleksi manual telah banyak ditekankan pada riset terdahulu. Dwiguna dan Bahtiar (2024) mengonfirmasi bahwa proses manual penentuan penerima Bantuan Langsung Tunai (BLT) di Desa Pamulihan dinilai tidak transparan, kurang efisien, serta berisiko menimbulkan ketidakpuasan; melalui penerapan Algoritma K-Means, penerima bantuan dapat diidentifikasi secara lebih tepat sasaran, efisien, dan transparan dibandingkan proses manual. Temuan serupa dilaporkan Rainaya dan Dewi (2025) yang menyoroti bahwa ketergantungan verifikasi pada rekomendasi manual di tingkat RT dan RW tanpa dukungan teknologi membutuhkan waktu lama dan berisiko memicu mal-distribusi. Sejalan dengan itu, Hidayatullah, Prihartono dan Fathurrohman (2025) membuktikan bahwa klasterisasi *K-Means* atas data penerima bantuan pangan di Kota dan Kabupaten Cirebon mampu meningkatkan efektivitas distribusi, sedangkan Siti Paridah dan Martanto (2024)

menunjukkan bahwa metode yang sama dapat meningkatkan akurasi penetapan penerima Program Keluarga Harapan sekaligus mengurangi kesalahan sasaran.

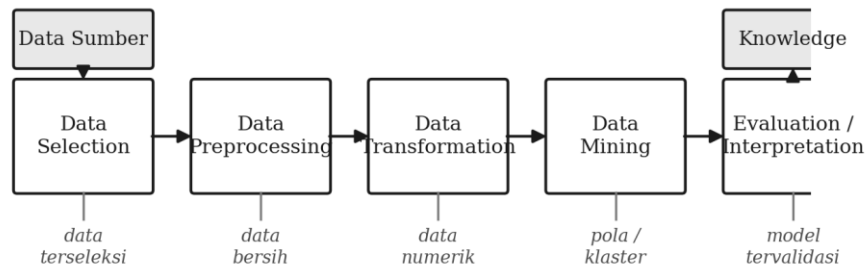
Meskipun demikian, sebagian besar penelitian terdahulu tersebut dijalankan pada lingkup wilayah yang relatif sempit, yaitu tingkat desa atau kelurahan, dengan volume data berkisar puluhan hingga ratusan record, dan atribut yang umumnya bersumber dari pendataan lapangan atau pernyataan mandiri responden. Penelitian ini menempatkan diri pada celah tersebut melalui dua pembeda. Pertama, cakupan analisis berada pada tingkat provinsi dengan volume data ratusan ribu record dalam satu tahap penyaluran, sehingga sekaligus menguji keterterapan Algoritma *K-Means* pada skala administratif yang jauh lebih besar. Kedua, atribut kepemilikan aset tidak diperoleh melalui pendataan mandiri, melainkan ditarik langsung dari basis data administratif resmi berupa catatan pajak kendaraan bermotor, sehingga keandalannya tidak bergantung pada pelaporan mandiri calon penerima.

Meskipun berbagai penelitian telah membuktikan efektivitas Algoritma *K-Means* dalam proses pengelompokan data pada berbagai bidang, sebagian besar kajian sebelumnya masih berfokus pada klasifikasi penerima bantuan sosial secara umum atau pada pengelompokan data pendidikan dengan variabel yang terbatas (Suhardinata, Ningsih dan Kasyidi, 2023; Andayani dan Nababan, 2024; Hidayatullah, Prihartono dan Fathurrohman, 2025). Selain itu, penelitian terdahulu umumnya lebih menitikberatkan pada aspek akurasi pembentukan klaster, sementara integrasi hasil klasterisasi ke dalam sistem pendukung keputusan yang mendukung transparansi dan akuntabilitas penyaluran bantuan pendidikan masih relatif terbatas. Belum banyak penelitian yang secara spesifik mengkaji penerapan *K-Means* dalam proses verifikasi dan klasifikasi penerima Kartu Jakarta Pintar (KJP) Plus dengan memanfaatkan platform visual data science yang dapat dipahami oleh pemangku kepentingan non-teknis. Kondisi ini menunjukkan adanya kesenjangan penelitian (research gap), yaitu minimnya model klasifikasi berbasis data mining yang tidak hanya mampu mengelompokkan calon penerima bantuan secara objektif, tetapi juga menyajikan hasil analisis dalam bentuk visual yang mendukung proses pengambilan keputusan yang transparan dan mudah diinterpretasikan.

Guna menjembatani kesenjangan tersebut, penelitian ini mengusulkan implementasi teknik data mining dengan menerapkan Algoritma *K-Means* sebagai basis pendukung keputusan dalam klasifikasi penerima bantuan pendidikan. Algoritma *K-Means* dipilih karena implementasinya mudah dipahami (Suhardinata, Ningsih dan Kasyidi, 2023; Hidayatullah, Prihartono dan Fathurrohman, 2025), memiliki performa komputasi yang relatif cepat dan sederhana, serta mempunyai kemampuan besar dalam mengolah data berskala besar dengan waktu yang lebih efisien (Andayani dan Nababan, 2024). Integrasi algoritma ini disimulasikan menggunakan perangkat komputasi visual berbasis data science, yaitu Orange, untuk menghasilkan visualisasi klaster yang interaktif dan mudah dipahami oleh pemangku kepentingan non-teknis. Dengan demikian, penerapan model klasterisasi *K-Means* ini diharapkan dapat mentransformasi sistem verifikasi KJP Plus di Provinsi DKI Jakarta menjadi lebih presisi, akuntabel, dan transparan.

## METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan data mining untuk membangun dan mengevaluasi model klasterisasi (Setiawan dkk., 2023) (Ariansyah, 2024) penerima dana bantuan pendidikan pada program KJP Plus. Kerangka kerja yang digunakan adalah Knowledge Discovery in Databases (KDD) (Han, Kamber dan Pei, 2012) (Juliawati, Buaton dan Saragih, 2023). dipilih dibanding CRISP-DM dengan pertimbangan sebagaimana diuraikan pada Tinjauan Pustaka yaitu proses menemukan pola bermanfaat dari data yang besar dan kompleks melalui lima tahapan utama yang saling berurutan, yaitu data selection, data preprocessing, data transformation, data mining, serta evaluation atau interpretasi (Wororomi dkk., 2024). Setiap tahapan menghasilkan keluaran antara yang menjadi masukan bagi tahapan berikutnya, sebagaimana disajikan pada Gambar 1.



Gambar 1. Tahapan Proses Knowledge Discovery in Databases (KDD)  
Sumber: Diadaptasi dari Wororomi dkk. (2024)

### Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh secara resmi dari dua instansi Pemerintah Provinsi DKI Jakarta, yaitu Dinas Pendidikan dan Badan Pendapatan Daerah (Bapenda). Basis data penerima bantuan pendidikan KJP Plus Tahap 2 Tahun Anggaran 2025 berjumlah 707.513 record dengan atribut meliputi nama siswa, jenis kelamin, NIK siswa, nomor KK siswa, kabupaten dan kota, nama sekolah, status sekolah, serta nomor KK wali. Adapun data pajak kendaraan bermotor dari Bapenda berjumlah 8.958.396 record yang terdiri atas atribut nomor plat kendaraan, nomor KK pemilik, tipe kendaraan, dan NIK pemilik. Kedua basis data sama-sama memuat nomor induk kependudukan sebagai pengenal tunggal sehingga secara teknis dapat dipertemukan. Rincian atribut beserta perannya dalam analisis disajikan pada Tabel 1.

Tabel 1. Atribut Data Penerima KJP Plus dan Perannya dalam Analisis

| No | Atribut                                     | Tipe Data      | Peran dalam Analisis                                                              |
|----|---------------------------------------------|----------------|-----------------------------------------------------------------------------------|
| 1  | Nama Siswa                                  | Teks (nominal) | Meta/pengenal, tidak diikutsertakan dalam perhitungan jarak                       |
| 2  | NIK Siswa/Wali                              | Teks (nominal) | Meta/pengenal; NIK wali menjadi kunci pencocokan dengan data kendaraan            |
| 3  | Nomor KK Siswa/Wali                         | Teks (nominal) | Meta/pengenal                                                                     |
| 4  | Status Sekolah                              | Kategorik      | Fitur klasterisasi (setelah transformasi numerik)                                 |
| 5  | Kepemilikan Aset (bobot kendaraan)          | Numerik        | Fitur klasterisasi utama; diperoleh dari penggabungan dengan data pajak kendaraan |
| 6  | Jenis Kelamin, Kabupaten/Kota, Nama Sekolah | Kategorik      | Tidak digunakan (di luar batasan masalah)                                         |

### Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh penerima bantuan pendidikan KJP Plus Tahap 2 Tahun Anggaran 2025 di Provinsi DKI Jakarta yang berjumlah 707.513 record. Teknik sampling yang digunakan adalah sampling jenuh (total sampling), yaitu penentuan sampel dengan menjadikan seluruh anggota populasi sebagai sampel, sehingga jumlah sampel sama dengan jumlah populasi. Teknik ini dipilih karena keseluruhan data populasi tersedia dan layak dianalisis, sehingga tidak diperlukan penarikan sebagian sampel yang berpotensi menghilangkan pola pada kelompok minoritas. Seluruh populasi diikutsertakan pada tahap penggabungan data, namun atribut kepemilikan kendaraan hanya melekat pada penerima yang walinya tercatat pada basis data pajak kendaraan bermotor, sehingga objek yang dapat diklasterkan berdasarkan bobot kendaraan dan status sekolah berjumlah 398.590 penerima atau sekitar 56% dari populasi. Penyaringan ini bukan penarikan sampel, melainkan konsekuensi ketersediaan atribut.

### Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Dokumentasi, yaitu pengumpulan data sekunder berupa basis data penerima KJP Plus dan data pajak kendaraan bermotor yang diperoleh secara resmi dari Dinas Pendidikan dan Badan Pendapatan Daerah Provinsi DKI Jakarta.
2. Studi kepustakaan, yaitu penelusuran literatur berupa jurnal ilmiah, buku, dan peraturan perundang-undangan yang relevan dengan penyaluran bantuan sosial pendidikan dan penerapan algoritma klusterisasi.
3. Observasi dokumen kebijakan, yaitu penelaahan Peraturan Gubernur Provinsi DKI Jakarta Nomor 141 Tahun 2016 dan Nomor 44 Tahun 2022 serta Laporan Hasil Pemeriksaan BPK RI Tahun 2023 sebagai dasar penetapan kriteria kelayakan penerima.

### Teknik Analisis Data

Analisis data mengikuti lima tahapan KDD. Pada tahap data mining, klusterisasi dilakukan dengan Algoritma *K-Means* yang diimplementasikan pada perangkat lunak Orange melalui rangkaian widget File, Data Table, Preprocess, k-Means, Scatter Plot, dan Silhouette Plot. Konfigurasi parameter model yang digunakan adalah jumlah kluster  $k = 2$ , metode inisialisasi centroid *K-Means++* (Initialize with KMeans++), ukuran jarak Euclidean Distance, jumlah iterasi maksimum 300, serta pengulangan (re-runs) sebanyak 10 kali yang menyimpan solusi dengan total jarak dalam kluster terkecil, sehingga hasil klusterisasi tetap stabil ketika proses dijalankan ulang pada data yang sama.

Setiap tahap mencakup prosedur perhitungan yang spesifik sebagai berikut. Nilai centroid awal dihitung menggunakan persamaan:

$$C_k = (1/n_k) \sum d \quad (1)$$

dengan  $C_k$  adalah centroid pada kluster ke- $k$ ,  $n_k$  banyaknya data pada kluster ke- $k$ , dan  $d$  nilai data pada kluster tersebut. Selanjutnya, jarak setiap data terhadap centroid dihitung menggunakan Euclidean Distance:

$$d(x, c) = \sqrt{(\sum (x_i - c_i)^2)} \quad (2)$$

Kualitas kluster dinilai melalui validasi internal. Davies-Bouldin Index dihitung dari kohesi dan separasi kluster. Sum of Square Within Cluster (SSW) mengukur kohesi dalam kluster ke- $i$ :

$$SSW_i = (1/m_i) \sum d(x_i, c_i) \quad (3)$$

Sum of Square Between Cluster (SSB) mengukur separasi antarkluster:

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

Rasio antara kluster ke- $i$  dan ke- $j$  kemudian dihitung sebagai:

$$R_{i,j} = (SSW_i + SSW_j) / SSB_{i,j} \quad (5)$$

Nilai rasio digunakan untuk memperoleh DBI melalui persamaan:

$$DBI = (1/K) \sum \max_{i \neq j} (R_{i,j}) \quad (6)$$

dengan  $K$  jumlah kluster yang digunakan. Semakin kecil nilai DBI, semakin baik kluster yang diperoleh. Sebagai metrik utama pemilihan jumlah kluster digunakan koefisien Silhouette:

$$s(i) = (b(i) - a(i)) / \max(a(i), b(i)) \quad (7)$$

dengan  $a(i)$  rata-rata jarak objek  $i$  terhadap objek lain dalam kluster yang sama dan  $b(i)$  rata-rata jarak minimum objek  $i$  terhadap objek pada kluster tetangga terdekat. Koefisien Silhouette diperoleh langsung dari keluaran widget *K-Means* dan Silhouette Plot pada Orange (Mustika dkk., 2021) (Hartama, Wanayumini dan Damanik, 2024). Sedangkan nilai DBI dihitung terpisah atas data berlabel kluster pada ruang fitur terstandarisasi yang sama (Sari dkk., 2024). Karena perhitungan koefisien Silhouette menuntut jarak antarpasangan objek sehingga kebutuhan memorinya tumbuh kuadratik terhadap jumlah record, pengujian jumlah kluster dijalankan atas subset acak berukuran 5.000 record dengan penyemaian tetap agar dapat direplikasi, sementara pemodelan *K-Means* tetap dijalankan atas seluruh 398.590 objek analisis.

### Algoritma K-Means

*K-Means* merupakan salah satu algoritma dalam teknik klustering non-hierarki yang berusaha mempartisi data ke dalam satu atau lebih kluster, sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam kluster yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam kelompok lainnya (Siti Paridah dan Martanto, 2024). *K-Means* adalah metode klustering berbasis jarak yang membagi data ke dalam sejumlah kluster dan hanya bekerja pada atribut numerik. Algoritma ini melakukan dua proses utama, yakni pendeteksian lokasi pusat tiap kluster dan pencarian anggota dari tiap kluster, dengan

tujuan meminimalkan variasi di dalam suatu kumpulan dan memaksimalkan keragaman antarkumpulan (Hamu dan Talakua, 2024; Hutagaol, Anggraeni dan Nata, 2025).

Cara kerja Algoritma *K-Means* secara umum adalah sebagai berikut (Sari dkk., 2024; Ufriani, Jasmir dan Arvita, 2023):

1. Tentukan  $k$  sebagai jumlah klaster yang ingin dibentuk.
2. Bangkitkan  $k$  centroid (titik pusat klaster) awal.
3. Hitung jarak setiap data ke masing-masing centroid.
4. Setiap data memilih centroid yang terdekat.
5. Tentukan posisi centroid baru dengan menghitung nilai rata-rata data pada centroid yang sama.
6. Ulangi langkah ketiga sampai posisi centroid tidak lagi berubah (konvergen).

Bentuk dasar *K-Means* membangkitkan centroid awal secara acak. Penelitian ini menerapkan varian penyempurnaannya, yaitu *K-Means++*, yang memilih centroid pertama secara acak seragam kemudian memilih centroid berikutnya dengan peluang sebanding dengan kuadrat jarak setiap objek terhadap centroid terdekat yang telah terpilih, sehingga posisi centroid awal tersebut dan risiko algoritma berhenti pada optimum lokal menurun (Arthur dan Vassilvitskii, 2007).

## HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metode Knowledge Discovery in Database (KDD) yang terdiri atas tahapan data selection, pre-processing, transformation, data mining, dan evaluation untuk menganalisis efektivitas Algoritma *K-Means* dalam mengelompokkan calon penerima KJP Plus Tahap 2 Tahun Anggaran 2025 di Provinsi DKI Jakarta. Pendekatan ini dipilih karena mampu mengidentifikasi pola tersembunyi dalam data secara objektif sehingga dapat menghasilkan kelompok penerima bantuan berdasarkan tingkat kemiripan karakteristik sosial ekonomi yang dimiliki.

### *Selection*

Tahap seleksi diarahkan untuk mempertemukan basis data penerima KJP Plus Tahap 2 dengan basis data pajak kendaraan bermotor guna menelusuri atribut kepemilikan aset transportasi. Penggabungan dilakukan melalui pencocokan nomor induk kependudukan (NIK) wali penerima terhadap NIK pemilik kendaraan; setiap penerima yang walinya teridentifikasi memiliki kendaraan selanjutnya ditautkan ke tabel referensi untuk memperoleh jenis kendaraan beserta bobotnya. Melalui proses penyaringan ini, diperoleh 398.590 penerima yang walinya tercatat memiliki kendaraan bermotor dari total populasi 707.513 record (sekitar 56%), dan himpunan inilah yang menjadi objek klasterisasi. Setiap penerima hanya diwakili oleh dua atribut pembentuk klaster, yaitu bobot kendaraan sebagai representasi kelas kepemilikan aset dan status sekolah sebagai penanda jenjang pendidikan, sedangkan seluruh atribut pengenalan diperlakukan sebagai data pelengkap yang tidak ikut dalam perhitungan jarak.

### *Preprocessing*

Pada tahap praproses dilakukan pemeriksaan kualitas data yang mencakup identifikasi nilai kosong (missing value), data duplikat, serta data pencilan (outlier). Berkas sumber tidak mengandung nilai kosong pada kolom-kolom kunci. Nilai kosong yang muncul setelah penggabungan dengan data pajak kendaraan bukanlah cacat data, melainkan penanda ketidakcocokan bagi penerima yang walinya tidak tercatat memiliki kendaraan; baris semacam ini ditangani melalui penyaringan (Select Rows dengan kondisi `KODE_KEND` terdefinisi), bukan melalui imputasi, karena imputasi justru akan mengarang sinyal aset yang sesungguhnya tidak ada dan membiaskan klasterisasi. Selain itu, karena satu wali dapat memiliki lebih dari satu kendaraan, satu NIK yang berulang pada data pajak kendaraan disaring menjadi satu catatan per NIK dengan aturan last instance (widget Unique). Konsekuensinya, jumlah pemilik kendaraan berat yang terbaca bersifat batas bawah (lower bound) dan cenderung konservatif.

### *Transformation*

Tahap transformasi mengubah data hasil praproses menjadi format yang sesuai untuk analisis *K-Means* yang hanya bekerja pada data numerik. Pada widget Preprocess, ketiga penapis dijalankan berurutan dan urutannya tidak boleh dipertukarkan, yaitu standardisasi (Normalize), imputasi sebagai pengaman (Impute), lalu one-hot encoding (Continue). Karena standardisasi dilakukan sebelum one-hot encoding, penyetaraan skala (rata-rata 0 dan ragam

1) hanya diterapkan pada bobot kendaraan, sedangkan 16 kolom biner status sekolah yang terbentuk kemudian tetap bernilai mentah 0 atau 1. Pemilihan one-hot encoding dilandasi kenyataan bahwa jenis sekolah tidak memiliki urutan intrinsik, sehingga pengkodean angka berurutan (label encoding) akan memaksakan relasi tingkat yang keliru; algoritma bertipe campuran seperti K-Modes (Huang, 1998) dapat menangani data kategorik secara langsung, namun tidak dipilih demi menjaga kesederhanaan, dukungan penuh Orange, dan keterbandingan dengan penelitian terdahulu yang juga memakai K-Means. Konsekuensi dari skema penskalaan ini bersifat dapat diramalkan: karena hanya bobot kendaraan yang distandardisasi sementara kolom status sekolah tetap 0 atau 1, jarak Euclidean antarobjek akan lebih banyak ditentukan oleh bobot kendaraan, sehingga status sekolah diperkirakan tidak akan efektif membedakan kluster. Rincian pemetaan nilai asal terhadap nilai hasil transformasi disajikan pada Tabel 2.

Tabel 2. Transformasi Nilai Atribut Kategorik menjadi Numerik

| No | Atribut                            | Nilai Asal (Kategorik)                                                   | Metode/Nilai Hasil Transformasi                                                         |
|----|------------------------------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| 1  | Status Sekolah                     | SDN, SDS, MIN, MIS, MTSN, MTSS, SMPN, SMPS, SMAN, SMAS, SMKN, SMKS, dll. | One-hot encoding: satu kolom biner (0/1) untuk setiap kategori                          |
| 2  | Kepemilikan Aset (jenis kendaraan) | Sepeda motor, sedan, micro/minibus, pick up, jeep, truk, dan lainnya     | Direpresentasikan oleh bobot kendaraan (numerik, mis. 1.000–1.400) dari tabel referensi |

## Data Mining

Pada tahap data mining, klasterisasi dijalankan dengan Algoritma *K-Means* (Siti Paridah dan Martanto, 2024) pada Orange melalui rangkaian widget yang telah dirancang. Inisialisasi centroid menggunakan metode K-Means++ dan jarak antardata dihitung menggunakan Euclidean Distance, dengan iterasi berulang sampai posisi centroid konvergen. Rangkaian widget model klasterisasi disajikan pada Gambar 2.

Data Table (2) - Orange (Not Responding)

File Edit View Window Help

Info

396590 instances  
17 features  
No target variable  
11 meta attributes (9.1 % missing data)

Variables

☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

Selection

☐ Select full rows

|    | nama_siswa       | nik_s            | nama_sekolah    | nik_w            | no_kk_w          | HURUF_TNKB | KODE_KEND | NO_KK            | JENIS         | Cluster | Silhouette | BOBOT  | status_ssk |
|----|------------------|------------------|-----------------|------------------|------------------|------------|-----------|------------------|---------------|---------|------------|--------|------------|
| 1  | TELUKFAIZDAR...  | 1111090211070... | SMKINANJUSA...  | 1111096105820... | 3175091305151... | TYS        | 98        | 3175091305151... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 2  | TELUKHATIMW...   | 1111090802130... | SDNSUSUKAND...  | 1111096105820... | 3175091305151... | TYS        | 98        | 3175091305151... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 3  | CUTMUTIAAZZ...   | 1111136304110... | SMPNEGERI96J... | 1111134309790... | 3174042404200... | SWP        | 98        | 3174042404200... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 4  | MUHAMMADA...     | 3173011509151... | SDNEGERIKAPU... | 3173014406830... | 3173011001094... | BZK        | 98        | 3173011001094... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 5  | MUHAMMADIL...    | 3173011509160... | MISALFURQON...  | 3173015601850... | 3173012009160... | BVA        | 98        | 3173012009160... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 6  | MUHAMMADH...     | 1305052407140... | SDNSUSUKAND...  | 1305057001900... | 3175093012150... | TJK        | 98        | 3175093012150... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 7  | MUHAMMADFI...    | 3173011510101... | MTSSALMUNA...   | 3173015005670... | 3173010601094... | BJO        | 98        | 3173010601094... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 8  | IRSYADHAMA...    | 3173011510111... | SMPNEGERI108... | 3173016909771... | 3173010901090... | BTS        | 98        | 3173010901090... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 9  | AZKARIBRAH...    | 3173011510121... | MTSSANNIDAA...  | 3173012412830... | 3173011801091... | BXS        | 98        | 3173011801091... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 10 | MUHAMMADA...     | 3173011510121... | SDNEGERIKAPU... | 3173014504910... | 3173011401098... | BTB        | 98        | 3173011401098... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 11 | AIRLANGSARJ...   | 3173011510131... | MIN7JAKARTA...  | 3173015402870... | 3173011102100... | BSN        | 98        | 3173011102100... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 12 | MUHAMMADH...     | 3173011510131... | SDNCEGKARE...   | 3173015612850... | 3173011909131... | BTB        | 98        | 3173011909131... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 13 | ALVAROFABIA...   | 3173011511150... | SDNEGERIKEDA... | 3173015106870... | 3173010608100... | BMI        | 98        | 3173010608100... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 14 | MUHAMMADIM...    | 3173011511150... | SDNEGERICEN...  | 3173012707691... | 3173010510100... | BLI        | 98        | 3173010510100... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 15 | AZKHADWIPUT...   | 3173011511160... | SDNEGERICEN...  | 3173016809730... | 3173011304180... | BTU        | 98        | 3173011304180... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 16 | ARVANPRAMA...    | 3173011511170... | SDNEGERIKAPU... | 3173011006811... | 3173011105111... | BTH        | 98        | 3173011105111... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 17 | RAFFARDHANU...   | 3173011511170... | SDNEGERITEGA... | 3173016806930... | 3173011109141... | BWM        | 98        | 3173011109141... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 18 | LOVELYBAHQIS...  | 1372015106080... | SMKN22JAKAR...  | 1372015912830... | 3175040309180... | TUF        | 98        | 3175040309180... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 19 | AFRATULHOLLI...  | 1373014411100... | SMPNEGERI24J... | 1373016510900... | 3175082305170... | TKM        | 98        | 3175082305170... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 20 | KANIANPUTRI...   | 1373016708150... | SDPINANGRA...   | 1373016510900... | 3175082305170... | TKM        | 98        | 3175082305170... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 21 | ADITYAARTANDI... | 3173011601100... | SMPDARROSTA...  | 3173016909750... | 3173010806151... | BOJ        | 98        | 3173010806151... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 22 | NOVALALDJIRA...  | 3173011601101... | SMPNEGERI249... | 3173015909760... | 3173010609100... | BXO        | 98        | 3173010609100... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 23 | RIALIJANMESSA... | 3173011601101... | SMPMERDEKAS...  | 3173012410720... | 3173010801092... | PCL        | 98        | 3173010801092... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 24 | DENDYRANUKU...   | 3173011601101... | MAN17JAKARTA... | 3173014304740... | 3173011001095... | BKK        | 98        | 3173011001095... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 25 | MUHAMMADZ...     | 3173011601101... | SMAN57JAKAR...  | 3173014803810... | 3173011312100... | BMB        | 98        | 3173011312100... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 26 | RAIJUDITHAL...   | 3173011603111... | MTSN11JAKAR...  | 3173015010860... | 3173010802111... | BIB        | 98        | 3173010802111... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 27 | ALVINNDOWIC...   | 3173011603121... | SMPALMUGHILL... | 3173015703740... | 3173012101091... | PS         | 98        | 3173012101091... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 28 | BINTANGPRAT...   | 3173011603121... | SMPNEGERI292... | 3173046804920... | 3173041805210... | BOM        | 98        | 3173041805210... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 29 | DIMASADITIVA...  | 3173011603121... | MTSSALMUNA...   | 3173015106810... | 3173010801090... | BDJ        | 98        | 3173010801090... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 30 | MUSYARADI...     | 3173011603121... | MTSSNURULLO...  | 3173015806780... | 3173011701095... | BGW        | 98        | 3173011701095... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 31 | FATHANDWIYU...   | 3173011603121... | SMPAL-MANS...   | 3173016608860... | 3173012101099... | BZI        | 98        | 3173012101099... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 32 | YUKIYONOSAH...   | 3173011603121... | SMPMITRAREF...  | 3173012711840... | 3173010610110... | SCD        | 98        | 3173010610110... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 33 | ANGGULIANDA...   | 1605045311070... | SMKN44JAKAR...  | 1605042612800... | 3175062101160... | TDV        | 98        | 3175062101160... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 34 | LIANAVANESSA...  | 1605075801120... | SMPANNUR...     | 1605074705740... | 317302111141...  | BSV        | 98        | 317302111141...  | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |
| 35 | NOVITAMARSY...   | 1607106301120... | SDNPLUIT03...   | 1607106908840... | 3172012102100... | UCY        | 98        | 3172012102100... | SPO-MTR >SOCC | C1      | ?          | -0.089 |            |

Restore Original Order

☒ Send Automatically

399K | 399K

Gambar 2. Rangkaian Widget Model Klasterisasi *K-Means* pada Orange

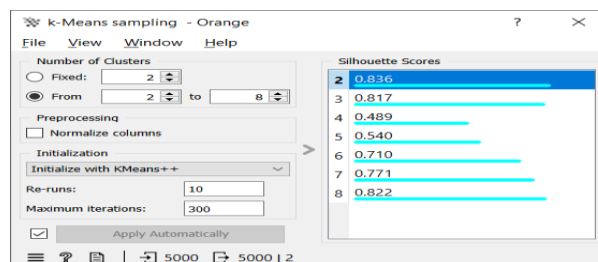
Sebelum penetapan model final, jumlah kluster (*k*) diuji secara objektif menggunakan koefisien Silhouette pada rentang  $k = 2$  hingga  $k = 8$ . Mengingat komputasi Silhouette pada keseluruhan data melampaui ambang kelayakan perangkat, pengujian dijalankan pada subset

acak berukuran 5.000 record dengan penyemaian tetap, sementara konfigurasi terbaik selanjutnya diterapkan pada keseluruhan data. Hasil pengujian dirangkum pada Tabel 3 dan divisualisasikan pada Gambar 3.

Tabel 3. Skor Koefisien Silhouette pada Berbagai Jumlah Kluster

| Jumlah Kluster (k) | Koefisien Silhouette | Keterangan          |
|--------------------|----------------------|---------------------|
| 2                  | 0,836                | Tertinggi — dipilih |
| 3                  | 0,817                | Pembanding          |
| 4                  | 0,489                | Pembanding          |
| 5                  | 0,540                | Pembanding          |
| 6                  | 0,710                | Pembanding          |
| 7                  | 0,771                | Pembanding          |
| 8                  | 0,822                | Pembanding          |

Sumber: Panel Silhouette Scores pada widget K-Means Orange untuk  $k = 2$  sampai  $k = 8$ .



Gambar 3. Hasil Silhouette Scores

Nilai koefisien Silhouette tertinggi diperoleh pada  $k = 2$  sebesar 0,836, diikuti  $k = 8$  sebesar 0,822 dan  $k = 3$  sebesar 0,817. Berdasarkan kriteria nilai tertinggi, konfigurasi dua kluster ditetapkan sebagai model final. Penetapan ini sekaligus selaras dengan kebutuhan pengambilan keputusan yang membagi penerima ke dalam dua kelompok, yaitu kelompok yang layak dan tidak layak. Perlu dicermati bahwa seluruh nilai berada di atas 0,8 — tergolong struktur kuat — dengan selisih antar-konfigurasi yang sangat tipis dan pola menurun lalu menaik kembali. Pola semacam ini merupakan ciri data yang atribut pembentuknya bersifat diskret atau kategorikal, sehingga objek terkonsentrasi pada sejumlah titik kombinasi alih-alih menyebar merata. Dengan karakteristik demikian, pemilihan  $k = 2$  tidak hanya didasarkan pada skor tertinggi, tetapi juga pada kesederhanaan dan ketertafsiran hasil.

### Evaluation & Interpretation

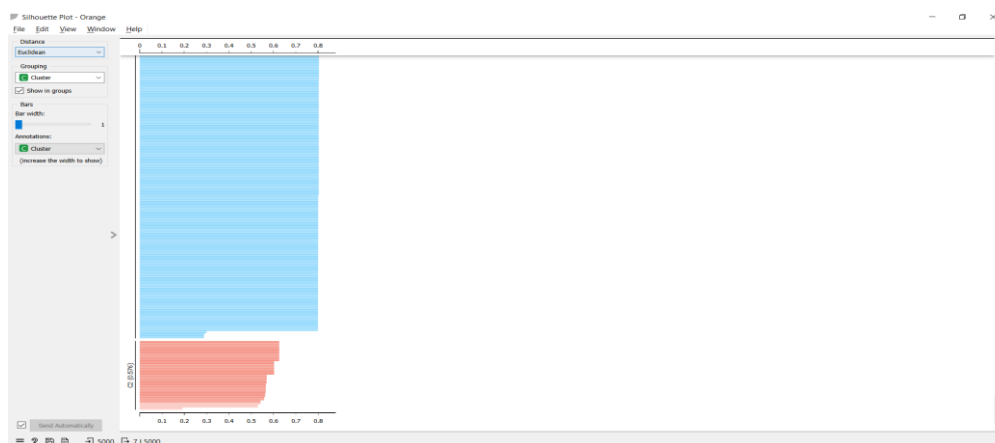
Kualitas model dua kluster dinilai melalui dua indikator validasi internal, yaitu koefisien Silhouette dan Davies-Bouldin Index (DBI). Hasil evaluasi disajikan pada Tabel 5.

Tabel 4. Hasil Evaluasi Internal Model Kluster ( $k = 2$ )

| Indikator            | Nilai | Kriteria           |
|----------------------|-------|--------------------|
| Koefisien Silhouette | 0,836 | Mendekati 1 (baik) |
| Davies-Bouldin Index | 0,41  | Mendekati 0 (baik) |

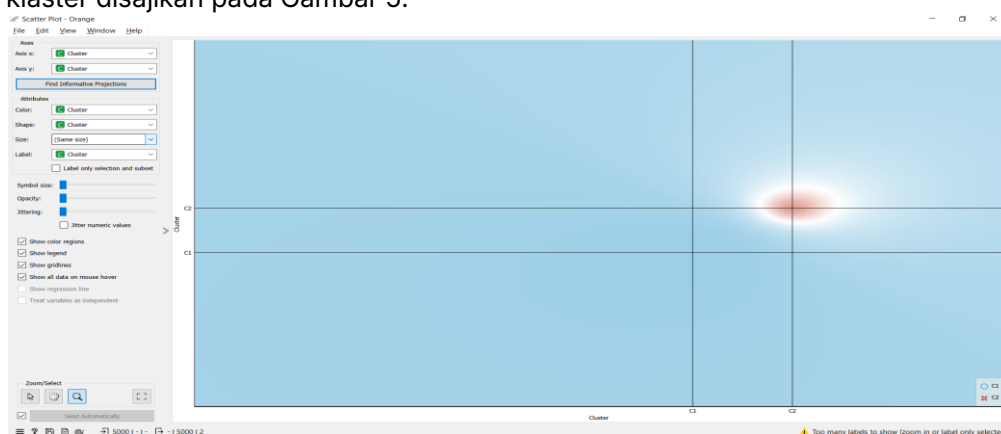
Sumber: Silhouette dari Orange; DBI dihitung dari penetapan kluster pada ruang fitur terstandarisasi.

Koefisien Silhouette sebesar 0,836 berada pada rentang 0,71–1,00 yang menandakan struktur kluster yang kuat, yaitu tingginya kohesi objek di dalam klasternya sendiri sekaligus separasi yang jelas terhadap kluster lain. Nilai Davies-Bouldin Index sebesar 0,41 yang mendekati nol memperkuat kesimpulan tersebut. Karena widget *K-Means* pada Orange tidak menghasilkan DBI secara langsung, nilai tersebut dihitung dari data berlabel kluster pada ruang fitur yang sama; sebagai verifikasi, koefisien Silhouette pada ruang fitur tersebut (0,855) mendekati nilai yang dilaporkan Orange (0,836), sehingga kedua indikator konsisten. Konsistensi penempatan objek juga tampak pada Silhouette Plot yang disajikan pada Gambar 4.



Gambar 4. Hasil Silhouette Plot

Selain kedua indikator numerik, mutu pemisahan kluster ditinjau secara visual melalui scatter plot yang memetakan setiap objek menurut bobot kendaraan dan label kluster. Sebaran kedua kluster disajikan pada Gambar 5.



Gambar 5. Hasil Scatter Plot Sebaran Dua Kluster

Gambar 5 memperlihatkan dua kelompok titik yang terpisah tegas tanpa wilayah tumpang tindih. Kluster 1 memusat rapat pada nilai bobot terendah, sedangkan Kluster 2 menempati rentang bobot yang lebih tinggi dengan sebaran yang jauh lebih lebar. Tidak adanya objek yang menggantung di antara kedua kelompok menjelaskan tingginya koefisien Silhouette yang diperoleh, karena setiap objek berjarak jauh lebih dekat terhadap centroid klasternya sendiri dibandingkan terhadap centroid kluster lain. Pola visual ini memperkuat temuan bahwa pemisahan digerakkan oleh atribut bobot kendaraan.

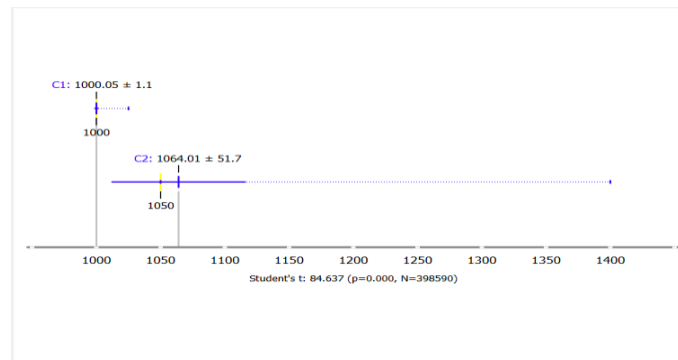
Karakteristik masing-masing kluster diperoleh dengan menempelkan kembali label kluster pada data asli, kemudian menelaah sebaran bobot kendaraan, jenis kendaraan, dan status sekolah pada tiap kelompok. Ringkasan profil disajikan pada Tabel 6.

Tabel 5. Profil Ringkas Dua Kluster

| Kluster   | Jumlah (%)      | Bobot (rata-rata/rentang) | Jenis Kendaraan Dominan      | Jenjang Dominan |
|-----------|-----------------|---------------------------|------------------------------|-----------------|
| Kluster 1 | 393.909 (98,8%) | 1.000 (1.000–1.025)       | Sepeda motor (SPD-MTR >50CC) | SDN, SMPN       |
| Kluster 2 | 4.681 (1,2%)    | 1.064 (1.050–1.400)       | Micro/minibus, pick up, jeep | SDN, SMPN       |

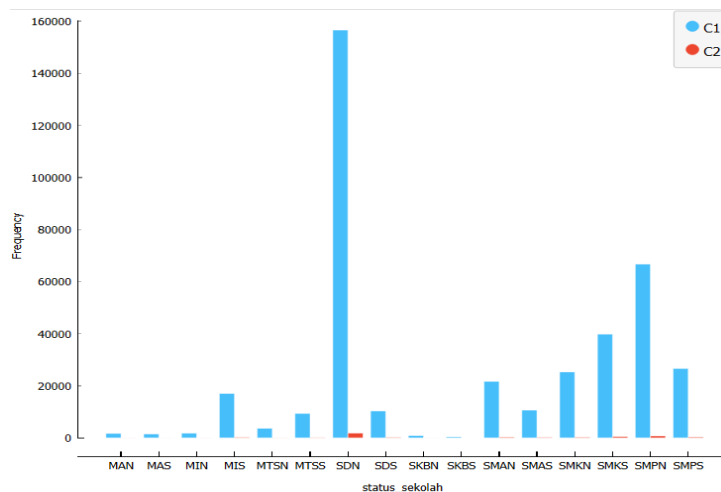
Kluster 1 menghimpun 393.909 penerima atau 98,8% dari total, dengan bobot kendaraan yang sangat homogen di kisaran 1.000–1.025. Kelompok ini hampir seluruhnya terdiri atas pemilik kendaraan roda dua, yakni jenis SPD-MTR >50CC sebanyak 392.399 unit atau sekitar 99,6% dari anggota kluster, sehingga dapat dimaknai sebagai kelompok penerima dengan

kepemilikan kendaraan ringan. Sebaliknya, Klaster 2 berukuran jauh lebih kecil, yakni 4.681 penerima atau 1,2% dari total, namun memiliki bobot kendaraan yang lebih tinggi dan lebih bervariasi (1.050–1.400). Kelompok ini didominasi kendaraan roda empat dan niaga, yaitu micro/minibus sebanyak 3.608 unit (sekitar 77% anggota klaster), diikuti pick up, jeep, blind/delivery van, mikrolet, serta sejumlah truk dan kendaraan berat lainnya, sehingga dapat dimaknai sebagai kelompok penerima dengan kepemilikan kendaraan berat atau bernilai ekonomi lebih tinggi. Perbedaan bobot antara kedua kelompok dapat dibaca lebih ringkas melalui box plot pada Gambar 6.



Gambar 6. Hasil Box Plot Sebaran Bobot Kendaraan Tiap Klaster

Gambar 6 memperlihatkan kotak Klaster 1 yang sangat pendek dan nyaris berimpit pada nilai 1.000, yang menandakan bobot hampir seragam di seluruh anggotanya. Sebaliknya, kotak Klaster 2 berada pada posisi yang lebih tinggi dengan rentang antarkuartil yang jauh lebih lebar serta ekor yang memanjang sampai 1.400. Posisi median kedua klaster tidak beririsan sama sekali, sehingga perbedaan bobot antarkelompok tidak dapat dijelaskan sebagai variasi biasa di dalam satu populasi yang sama. Adapun komposisi jenis kendaraan pada masing-masing klaster ditampilkan pada Gambar 7.



Gambar 7. Hasil Distribusi Jenis Kendaraan per Klaster

Temuan yang menonjol adalah bahwa pemisahan kedua klaster ditentukan oleh jenis dan bobot kendaraan, bukan oleh status sekolah. Sebaran jenjang pendidikan pada kedua klaster hampir identik: SDN merupakan jenjang terbanyak di Klaster 1 (156.607 atau 39,8%) maupun Klaster 2 (1.813 atau 38,7%), diikuti pola yang serupa pada SMPN dan SMKS. Kesamaan pola ini menunjukkan bahwa status sekolah tersebar merata di seluruh kelompok dan tidak berkontribusi membedakan klaster, sehingga struktur pengelompokan sepenuhnya digerakkan oleh atribut kepemilikan kendaraan.

Hasil di atas mengarah pada satu simpulan analitis yang penting bagi konteks ketepatan sasaran KJP Plus. Karena kepemilikan sepeda motor di Indonesia bersifat nyaris merata dan bukan penanda kemapanan ekonomi yang kuat, keberadaan Klaster 1 yang sangat besar tidak serta-merta mengindikasikan ketidaklayakan. Sebaliknya, Klaster 2 yang berukuran kecil justru

merepresentasikan penerima yang walinya memiliki kendaraan roda empat atau niaga — penanda kapasitas ekonomi yang jauh lebih kuat dan karenanya lebih patut dicurigai tidak memenuhi kriteria kelayakan bantuan. Dengan kata lain, analisis ini menyarankan agar pengetahuan verifikasi difokuskan pada kelompok kecil pemilik kendaraan berat (Klaster 2), alih-alih memperlakukan seluruh pemilik kendaraan secara seragam.

Sebelum dimaknai lebih jauh, perlu ditegaskan sifat atribut bobot kendaraan yang menjadi penggerak utama pemisahan. Bobot kendaraan merupakan koefisien pada penghitungan pajak kendaraan bermotor, bukan ukuran berat fisik maupun nilai jual aset. Konsekuensinya, pemisahan yang terbentuk mencerminkan kelas administratif kendaraan, sehingga sebagian kendaraan niaga ringan seperti pikap dan mikrolet dapat menempati klaster berbobot tinggi, sementara sebagian sedan pribadi berada pada klaster berbobot rendah. Oleh karena itu, keanggotaan Klaster 2 sebaiknya dijadikan indikasi awal yang menandai prioritas verifikasi, bukan vonis ketidaklayakan.

Lebih lanjut, keanggotaan Klaster 2 tidak dapat diperlakukan sebagai kelompok yang seragam dalam hal kelayakan karena komposisi jenis kendaraannya beragam secara sosial-ekonomi. Sebagian merupakan kendaraan penumpang pribadi seperti jeep (280 unit), sementara sebagian besar lainnya justru kendaraan niaga atau angkutan, yaitu micro/minibus sebanyak 3.608 unit (sekitar 77% anggota klaster), diikuti pikap 371 unit, mikrolet 94 unit, dan blind/delivery van 99 unit. Kepemilikan kendaraan niaga perlu dibaca lebih berhati-hati karena tidak serta-merta menandakan kemapanan, melainkan berpotensi merupakan aset produktif atau sarana mencari nafkah bagi keluarga yang justru rentan secara ekonomi. Oleh karena itu, keanggotaan Klaster 2 sebaiknya tidak diterjemahkan menjadi satu tingkat prioritas verifikasi yang seragam, melainkan secara berjenjang, yaitu prioritas tertinggi bagi pemilik kendaraan penumpang pribadi bernilai tinggi, dan prioritas kedua yang menuntut verifikasi lapangan bagi pemilik kendaraan niaga yang statusnya masih perlu dikonfirmasi. Verifikasi berjenjang ini penting karena mayoritas anggota Klaster 2 justru berada pada kelompok kendaraan niaga yang ambigu, sehingga menyamaratakan seluruh anggota klaster sebagai indikasi ketidaklayakan berpotensi menimbulkan kesalahan sasaran dalam arah sebaliknya, yaitu mencurigai penerima yang sesungguhnya layak.

Temuan penelitian ini dapat dihubungkan dengan sejumlah penelitian terdahulu. Suhardinata, Ningsih dan Kasyidi (2023) memperoleh koefisien Silhouette yang tergolong lemah, sekitar 0,44, dan menduga bahwa penambahan data serta atribut dapat meningkatkan mutu pemisahan; koefisien 0,836 yang diperoleh dalam penelitian ini, dengan volume ratusan ribu objek, memberikan dukungan empiris terhadap dugaan tersebut. Konfigurasi dua klaster yang terpilih juga sejalan dengan Mirna dkk. (2025) yang menemukan struktur dua klaster pada data penerima bantuan sosial, sedangkan penggunaan perangkat Orange, inisialisasi K-Means++, dan validasi berbasis koefisien Silhouette pada penelitian ini konsisten dengan praktik yang ditempuh Timanta dkk. (2025) serta Astafah, Habibba dan Prayoga (2025).

Secara substansi, penelitian ini menindaklanjuti diagnosis Azzahra, Musyaffi dan Khairunnisa (2025) yang menyoroti akar persoalan ketidaktepatan sasaran KJP Plus pada mekanisme seleksi manual. Menariknya, jumlah 4.681 penerima pada Klaster 2 sebanding tatanan besarnya dengan 4.341 penerima tidak tepat sasaran yang dilaporkan BPK RI. Perbandingan ini perlu dibaca dengan hati-hati karena kedua angka berasal dari tahun anggaran, kriteria, dan metode penetapan yang berbeda, dan keanggotaan klaster bukan vonis kelayakan; kesamaan tatanan besaran hanya memperkuat indikasi bahwa persoalan kepemilikan aset di kalangan penerima berada pada skala ribuan. Perlu ditegaskan pula bahwa model ini tidak berbicara mengenai 308.923 penerima (sekitar 44%) yang NIK walinya tidak tercocokkan dengan basis data pajak kendaraan; ketidakcocokan tersebut tidak identik dengan kelayakan karena kelompok ini berada di luar cakupan analisis.

Perlu disampaikan secara terbuka keterbatasan metodologis penelitian. Keterbatasan utama terletak pada mekanisme pencocokan berbasis NIK wali: kendaraan yang terdaftar atas nama anggota keluarga lain atau pihak ketiga tidak tertangkap sehingga berpotensi menimbulkan indikasi negatif palsu, sedangkan kendaraan yang telah dijual namun belum dibaliknamakan dapat menimbulkan indikasi positif palsu. Oleh sebab itu, keluaran model diposisikan sebagai instrumen penyaringan awal yang mengarahkan prioritas verifikasi manual, bukan penentu status akhir kelayakan. Selain itu, tingginya nilai Silhouette yang relatif seragam pada seluruh jumlah klaster menegaskan bahwa atribut pembentuk bersifat kategorikal/diskret, sehingga klasterisasi pada data ini pada dasarnya mendekati segmentasi berdasarkan kombinasi bobot kendaraan dan jenjang sekolah; karena status sekolah tidak berperan

membedakan kelompok, dimensi analisis secara efektif bertumpu pada satu atribut kepemilikan kendaraan. Seluruh keterbatasan ini tidak menggugurkan kesahihan model — yang telah dikonfirmasi oleh koefisien Silhouette dan Davies-Bouldin Index — namun menjadi catatan penting bagi penafsiran hasil dan pengembangan penelitian lanjutan.

### KESIMPULAN

Berdasarkan hasil penelitian, penggabungan data penerima KJP Plus dengan data pajak kendaraan bermotor berhasil mengidentifikasi 398.590 penerima yang memiliki keterkaitan dengan kepemilikan kendaraan dan selanjutnya dikelompokkan menggunakan algoritma K-Means. Model terbaik diperoleh pada dua klaster ( $k = 2$ ) dengan nilai Silhouette sebesar 0,836 dan Davies-Bouldin Index sebesar 0,41, yang menunjukkan kualitas klasterisasi yang baik. Hasil pengelompokan memperlihatkan bahwa Klaster 1 (98,8%) didominasi pemilik kendaraan ringan, terutama sepeda motor, sedangkan Klaster 2 (1,2%) didominasi pemilik kendaraan roda empat dan kendaraan niaga. Status sekolah tidak memberikan kontribusi signifikan dalam pembentukan klaster, sehingga kepemilikan kendaraan menjadi faktor utama yang membedakan karakteristik penerima. Temuan ini menunjukkan bahwa kelompok pemilik kendaraan roda empat dan kendaraan niaga memiliki indikasi ketidaklayakan yang lebih kuat dibandingkan kelompok lainnya dan layak menjadi prioritas dalam proses verifikasi. Oleh karena itu, Dinas Pendidikan Provinsi DKI Jakarta disarankan untuk mengintegrasikan data kepemilikan kendaraan ke dalam sistem seleksi dan evaluasi penerima KJP Plus, menerapkan mekanisme verifikasi berbasis risiko dengan memprioritaskan pemeriksaan pada penerima yang termasuk dalam Klaster 2, serta memperkuat integrasi data lintas instansi secara berkala agar proses penyaluran bantuan menjadi lebih tepat sasaran, efisien, dan akuntabel. Selain itu, penelitian selanjutnya dapat menambahkan variabel sosial-ekonomi lain, seperti tingkat pendapatan, kepemilikan properti, dan konsumsi listrik, guna menghasilkan model klasterisasi yang lebih informatif dan mampu menggambarkan kondisi ekonomi penerima secara lebih akurat.

### REFERENSI

- Amna, S. W., dkk. (2022). *Data Mining*. PT Global Eksekutif Teknologi.
- Andayani, S., & Nababan, L. (2024). Algoritma *K-Means* Clustering untuk Program Kartu Indonesia Pintar bagi Siswa SMP Alwasliyah. *JID (Jurnal Info Digit)*, 2(3), 1689–1702.
- Ariansyah, R. (2024). Sistem Pendukung Keputusan Penentuan Penerima Bantuan Langsung Tunai (BLT) dengan Kombinasi Metode Clustering *K-Means* dan VIKOR. *ADA: Journal of Information System Research*, 1(3), 129–140. <https://doi.org/10.64366/adajisr.v1i3.51>
- Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. Dalam *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07)* (hlm. 1027–1035). SIAM.
- Astafah, A. A., Habibba, I. S., & Prayoga, S. T. (2025). Pemetaan Efektivitas Anggaran Kementerian pada Tahun 2023 melalui Klasterisasi dengan Orange Data Mining. *Integrative Perspectives of Social and Science Journal (IPSSJ)*, 2(3), 6236–6254.
- Azzahra, A., Musyaffi, A. M., & Khairunnisa, H. (2025). Analisis Akuntabilitas Penyaluran Dana Bantuan Sosial pada Pemerintah Provinsi DKI Jakarta. *Indonesia Economic Journal (IEJ)*, 1(2), 811–829. <https://doi.org/10.63822/rn7bb605>
- Badan Pemeriksa Keuangan Republik Indonesia. (2024). *Laporan Hasil Pemeriksaan atas Laporan Keuangan Pemerintah Daerah Provinsi DKI Jakarta Tahun 2023 Nomor 12/LHP/XVIII.JKT/7/2024*. BPK RI.
- Chapman, P., dkk. (2000). *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. SPSS Inc.
- Dwiguna, E., & Bahtiar, A. (2024). Penerapan Data Mining untuk Menentukan Penerima Bantuan BLT Menggunakan Metode Clustering *K-Means* pada Desa Pamulihan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1382–1388. <https://doi.org/10.36040/jati.v8i2.9029>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.
- Hamu, M. A. M., & Talakua, A. C. (2024). Optimalisasi Manajemen Bantuan Sosial dengan Implementasi Data Mining Menggunakan Algoritma *K-Means* Clustering. *Jurnal Ilmu Komputer Universitas Katolik Widya Mandira*, 1(1), 7–12.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hartama, D., Wanayumini, & Damanik, I. S. (2024). *Optimasi Algoritma Clustering: Studi Kasus Lokasi Bencana Alam di Indonesia*. PT Adab Indonesia.

- Hidayatullah, N. A., Prihartono, W., & Fathurrohman. (2025). Clustering Penerima Bantuan Pangan Berbasis Algoritma *K-Means* untuk Meningkatkan Efektivitas Program Sosial di Kota/Kabupaten Cirebon. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5692>
- Huang, Z. (1998). Extensions to the *K-Means* algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3), 283–304.
- Hutagaol, V., Anggraeni, D., & Nata, A. (2025). Penerapan Data Mining untuk Clustering Masyarakat Kurang Mampu pada Desa Rawang Pasar VI dengan Algoritma K-Means. *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 10(1), 304–312. <https://doi.org/10.29100/jipi.v10i1.5741>
- Juliawati, F., Buaton, R., & Saragih, R. (2023). Pengelompokan Data Mining Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan Metode Clustering. *Explorer*, 3(2), 69–76. <https://doi.org/10.47065/explorer.v3i2.793>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Mirna, Martanto, Dikananda, A. R., & Rifa'i, A. (2025). Algoritma *K-Means* untuk Mengelompokan Bantuan Sosial dalam Penentuan Strategi Distribusi Desa Sigong. *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, 7(2), 482–491.
- Mustika, dkk. (2021). *Data Mining dan Aplikasinya*. Widina Bhakti Persada.
- Peraturan Gubernur Provinsi DKI Jakarta Nomor 44 Tahun 2022 tentang Pemberian Bantuan Sosial dalam Rangka Perlindungan Sosial. (2022). Pemerintah Provinsi DKI Jakarta.
- Peraturan Gubernur Provinsi DKI Jakarta Nomor 141 Tahun 2016 tentang Bantuan Biaya Personal Pendidikan bagi Peserta Didik dari Keluarga Tidak Mampu Melalui Kartu Jakarta Pintar. (2016). Pemerintah Provinsi DKI Jakarta.
- Rainaya, & Dewi, S. (2025). Penerapan Data Mining Penyaluran Dana Bantuan Sosial Menggunakan Metode *K-Means* Clustering di Desa Rancaekek Kulon. *Jurnal Dimamu*, 4(3), 434–443.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Sari, R. M., Rizka, A., Putri, N. A., & Efriana, A. (2024). *Perhitungan Metode Clustering*. Serasi Media Teknologi.
- Setiawan, Z., dkk. (2023). *Buku Ajar Data Mining*. PT Sonpedia Publishing Indonesia.
- Siti Paridah, N., & Martanto, M. (2024). Klasterisasi Penerima Dana Bantuan Program Keluarga Harapan Menggunakan Metode *K-Means* pada Desa Gereba. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 1036–1043. <https://doi.org/10.36040/jati.v8i1.8873>
- Suhardinata, D., Ningsih, A. K., & Kasyidi, F. (2023). Klasterisasi Data Penduduk untuk Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan K-Means. *IJESPG*, 1(3), 221–228.
- Timanta, D., dkk. (2025). Klasterisasi Provinsi di Indonesia Berdasarkan Sumber Pendanaan Daerah Menggunakan K-Means. *Integrative Perspectives of Social and Science Journal (IPSSJ)*, 2(3), 6203–6219.
- Ufriani, S., Jasmir, & Arvita, Y. (2023). Penerapan Algoritma Clustering *K-Means* untuk Menentukan Prioritas Penerima Bantuan Dana Sosial PKH di Kelurahan Kampung Singkep. *JAKAKOM*, 3(1), 342–350.
- Wororomi, J. K., dkk. (2024). *Data Mining: Memahami Pola di Balik Angka*. Eureka Media Aksara.
- Zikri, D., Ismanto, S. U., & Candradewini. (2022). Upaya dalam Pengelolaan Program Kartu Jakarta Pintar Plus oleh UPT P4OP Wilayah Jakarta Selatan. *JANE (Jurnal Administrasi Negara)*, 13(2), 367–371.