

PENERAPAN METODE K-MEDOIDS DALAM ANALISIS POLA KEPADATAN PENUMPANG DI SETIAP HALTE SUROBOYO BUS

Thalita Syahlani Putri ¹⁾, I Gede Susrama Mas Diyasa ¹⁾, Achmad Junaidi ¹⁾

1) Program Studi Informatika Universitas Pembangunan Nasional Veteran, Jawa Timur

E-mail: 21081010203@student.upnjatim.ac.id, igsusrama.if@upnjatim.ac.id,
achmadjunaidi.if@upnjatim.ac.id

Abstrak

Transportasi publik merupakan elemen penting dalam mendukung mobilitas masyarakat perkotaan, termasuk di Kota Surabaya dengan layanan Suroboyo Bus. Penelitian ini bertujuan untuk menganalisis pola kepadatan penumpang menggunakan metode K-Medoids, yang dikenal efektif dalam menangani data dengan keberadaan outlier. Data yang digunakan berasal dari Dinas Perhubungan Kota Surabaya yang mencakup jumlah penumpang berdasarkan waktu dan halte. Tahapan penelitian meliputi pra-pemrosesan data seperti seleksi atribut, deteksi outlier dengan metode Interquartile Range (IQR), serta normalisasi menggunakan Z-Score. Proses klasterisasi dilakukan untuk mengelompokkan halte berdasarkan tingkat kepadatan penumpang, sedangkan evaluasi hasil menggunakan metrik Silhouette Score. Hasil penelitian menunjukkan bahwa metode K-Medoids mampu membentuk klaster dengan tingkat pemisahan yang baik dan nilai Silhouette Score sebesar 0,4222, yang menandakan kualitas klasterisasi cukup kuat. Penelitian ini memberikan kontribusi terhadap optimalisasi operasional transportasi publik, khususnya dalam menentukan kebijakan penambahan rute atau halte baru berdasarkan hasil klasterisasi kepadatan penumpang.

Kata kunci: K-Medoids, Klasterisasi, Suroboyo Bus, Silhouette Score, Data Mining.

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mendorong peningkatan signifikan dalam pengelolaan data di berbagai sektor, termasuk bidang transportasi publik [1]. Kota Surabaya sebagai salah satu kota metropolitan di Indonesia memiliki tingkat mobilitas masyarakat yang tinggi, yang menuntut ketersediaan transportasi publik yang efisien dan merata. Salah satu transportasi andalan yang digunakan masyarakat adalah Suroboyo Bus, yang memiliki konsep ramah lingkungan dan terus mengalami peningkatan jumlah penumpang setiap tahunnya. Peningkatan ini menuntut pengelola untuk memahami pola kepadatan penumpang di setiap halte agar distribusi armada dan jadwal operasional dapat diatur secara optimal.

Analisis terhadap pola kepadatan penumpang menjadi penting sebagai dasar perencanaan strategis dan pengambilan keputusan oleh Dinas Perhubungan Kota Surabaya. Namun, jumlah data penumpang yang besar dan dinamis memerlukan pendekatan analitis yang mampu mengelompokkan data secara efektif. Salah satu pendekatan yang relevan adalah data mining yaitu proses untuk menemukan pola dan informasi tersembunyi dari kumpulan data besar, salah satunya melalui teknik klasterisasi [2]. Klasterisasi digunakan untuk membagi data ke dalam beberapa kelompok berdasarkan tingkat kesamaan antar data, sehingga dapat diperoleh pola distribusi penumpang yang lebih terstruktur [3].

Salah satu metode yang efektif digunakan adalah K-Medoids, karena mampu menghasilkan klaster yang representatif dan tahan terhadap data ekstrem (outlier). Penelitian ini menerapkan metode K-Medoids untuk mengidentifikasi pola kepadatan penumpang Suroboyo Bus berdasarkan data jumlah penumpang dan waktu operasional [4]. Hasil analisis diharapkan dapat membantu Dinas Perhubungan Kota Surabaya dalam pengambilan keputusan strategis, seperti perencanaan rute dan penentuan prioritas penambahan halte atau armada di titik dengan tingkat kepadatan tinggi.

Studi Pustaka

Data mining merupakan proses untuk menemukan pola dan informasi tersembunyi dari kumpulan data besar dengan tujuan mendukung pengambilan keputusan berbasis data, salah satunya melalui teknik klasterisasi [5]. Klasterisasi yaitu metode pengelompokan data ke dalam beberapa kelompok berdasarkan tingkat kesamaan karakteristik antar data. Melalui klasterisasi, pola-pola

tersembunyi dalam data dapat diidentifikasi dan divisualisasikan secara lebih terstruktur, sehingga memudahkan interpretasi dan analisis lanjutan.

Salah satu algoritma klusterisasi yang banyak digunakan adalah K-Medoids, yang termasuk dalam algoritma berbasis partisi (*partitioning-based clustering*). Metode K-Medoids merupakan salah satu algoritma klusterisasi berbasis partisi yang menggunakan medoid sebagai pusat kluster, di mana medoid dipilih dari data yang memiliki jarak total terpendek terhadap anggota klusternya [4]. Berbeda dengan K-Means yang menggunakan rata-rata sebagai pusat kluster, K-Medoids lebih tahan terhadap data ekstrem (outlier) dan menghasilkan pembagian kluster yang lebih stabil [6].

Evaluasi hasil klusterisasi biasanya dilakukan dengan menggunakan metrik Silhouette Score, yang mengukur tingkat keseragaman data dalam satu kluster serta seberapa baik data terpisah antar kluster [7]. Beberapa penelitian sebelumnya menunjukkan bahwa metode K-Medoids efektif digunakan untuk analisis pola dan segmentasi data dalam berbagai bidang, termasuk transportasi publik, karena kemampuannya dalam mengidentifikasi kelompok data yang homogen dan representatif. Dengan demikian, kombinasi metode K-Medoids dan evaluasi Silhouette Score dapat menghasilkan analisis pola kepadatan penumpang yang akurat dan reliabel.

Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan metode K-Medoids untuk mengelompokkan halte Suroboyo Bus berdasarkan tingkat kepadatan penumpang. Data diperoleh dari Dinas Perhubungan Kota Surabaya yang mencakup jumlah penumpang, waktu keberangkatan, dan lokasi halte pada rute Purabaya–Perak. Tahapan penelitian meliputi proses pengumpulan data, pra-pemrosesan, klusterisasi, dan evaluasi hasil yang dijelaskan sebagai berikut:

1. Pra-pemrosesan Data: Meliputi seleksi atribut relevan, transformasi waktu, penanganan data kosong, deteksi outlier menggunakan Interquartile Range (IQR), serta normalisasi menggunakan Z-Score agar seluruh variabel berada pada skala yang sebanding.
2. Proses Klusterisasi: Metode K-Medoids digunakan untuk membentuk sejumlah kluster dengan medoid sebagai pusat representatif tiap kluster. Jumlah kluster ditentukan berdasarkan metode elbow.
3. Evaluasi Hasil: Kualitas hasil klusterisasi diukur menggunakan metrik Silhouette Score, yang menunjukkan tingkat keseragaman dalam kluster dan pemisahan antar kluster. Nilai yang mendekati 1 menunjukkan klusterisasi yang baik.
4. Visualisasi dan Interpretasi: Hasil klusterisasi divisualisasikan dalam bentuk scatter plot dan tabel karakteristik tiap kluster, untuk menggambarkan pola kepadatan penumpang pada setiap halte.

Tabel 1. Hasil Klusterisasi Halte Suroboyo Bus Menggunakan Metode K-Medoids

Kluster	Kategori Kepadatan Penumpang	Rata-rata Penumpang	Contoh Halte
1	Sedang	14.20	Halte Alun-Alun Contong, Halte Barunawati B, Halte Blauran, dst.
2	Rendah	14.19	Halte Gozco, Halte Intermodal Joyoboyo, Halte Kerto Menanggal, dst.
3	Sangat Tinggi	17.02	Halte Basra, Halte Dukuh Menanggal, Halte Ikan Kerapu B, dst.
4	Tinggi	16.97	Halte Gubernur Suryo, Halte Ikan Kerapu A, Halte Indrapura, dst.

Metodologi ini diharapkan dapat menghasilkan pengelompokan halte berdasarkan tingkat kepadatan penumpang yang akurat, sehingga dapat menjadi dasar bagi pengambilan keputusan strategis dalam perencanaan rute dan pengaturan armada Suroboyo Bus.

Hasil dan Pembahasan

Inialisasi K-Medoids

Proses inialisasi pada metode K-Medoids diawali dengan penentuan jumlah kluster optimal, yaitu $k = 4$, berdasarkan hasil evaluasi sebelumnya. Selanjutnya, nilai random seed diatur untuk memastikan hasil pemilihan medoid bersifat konsisten dan dapat direplikasi. Setelah itu, sistem secara acak memilih empat data unik dari keseluruhan dataset sebagai medoid awal, yang akan digunakan sebagai titik representatif setiap kluster.

```

--- 3. PROSES UTAMA K-MEDOIDS ---
Inialisasi medoid awal untuk k=4:
Medoid 1: index data = 654, No = 656
Medoid 2: index data = 114, No = 116
Medoid 3: index data = 25, No = 27
Medoid 4: index data = 281, No = 283

```

Gambar 1. Inialisasi K-Medoids

Gambar 1. menunjukkan tahap awal K-Medoids dimulai dengan inialisasi medoid acak dari data normalisasi, jumlah kluster (k), batas iterasi, dan random_state untuk memastikan konsistensi. Medoid awal dipilih dari data aktual, kemudian jarak Euclidean dihitung untuk menentukan label kluster. Proses iteratif meliputi perhitungan SSE dan pertukaran medoid (swap) untuk menurunkan SSE hingga konvergen. Hasil akhir berupa medoid optimal, label kluster, dan SSE_final digunakan untuk mengevaluasi kualitas klusterisasi dan membandingkan performa dengan metode lain seperti GMM dan AHC.

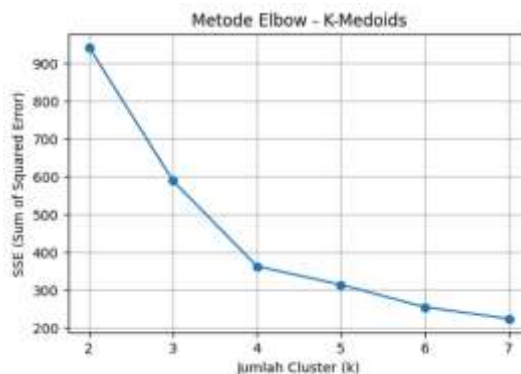
```

Iterasi 1
Total Cost (SSE) saat ini = 836.5047
Coba tukar Medoid 1 dengan data No=2: b=749.0604, S=-87.4443
= Swap dilakukan: Medoid 1 diganti dengan data No=2

```

Gambar 2. Hasil proses utama hingga konvergen

Gambar 2. menunjukkan Pada iterasi awal K-Medoids, sistem melakukan pertukaran medoid untuk menurunkan SSE, menetapkan medoid baru jika menghasilkan perbaikan. Proses iteratif ini terus dilakukan hingga konvergen pada iterasi ke-34 dengan SSE akhir 362,3469, menunjukkan konfigurasi medoid yang stabil. Selanjutnya, jumlah kluster optimal ditentukan menggunakan Metode Elbow untuk menyeimbangkan jumlah kluster dan kualitas pemisahan data.



Gambar 3. Hasil kluster optimal dengan metode elbow

Hasil analisis menunjukkan bahwa SSE menurun tajam dari $k = 2$ hingga $k = 4$ dan melandai setelahnya, sehingga titik elbow berada pada $k = 4$. Hal ini menandakan bahwa pembagian data menjadi empat klaster memberikan keseimbangan terbaik antara variasi dalam klaster dan kompleksitas model. Dengan demikian, jumlah klaster optimal ditetapkan empat, dan pada iterasi ke-34 algoritma K-Medoids mencapai konvergensi dengan SSE akhir 362,3469, menunjukkan klasterisasi yang efisien dan stabil. Selanjutnya, setiap data diberi label akhir berdasarkan medoid terdekat menggunakan jarak Euclidean, dengan penomoran klaster mulai dari 1 dan disimpan dalam kolom "KMedoids" pada dataset utama.

Label K-Medoids (5 data pertama):

No	Total passenger	time_numeric	KMedoids
2	19	6.5	3
3	10	7.5	1
4	4	8.5	1
5	17	9.5	3
6	12	10.5	1

Gambar 4. Hasil label K-Medoids

Pelabelan menunjukkan bahwa data dengan jumlah penumpang tinggi masuk ke klaster 3, sedangkan data dengan penumpang rendah termasuk klaster 1, sehingga memudahkan identifikasi pola kepadatan penumpang Suroboyo Bus. Selanjutnya, dilakukan analisis karakteristik klaster melalui perhitungan statistik deskriptif, termasuk jumlah data, rata-rata penumpang, dan rata-rata waktu operasional untuk setiap klaster, yang disajikan secara terurut untuk memudahkan perbandingan dan interpretasi pola tiap klaster.

Klaster 1:
Jumlah data: 227
Rata-rata penumpang: 8.90
Rata-rata waktu: 9.63
Klaster 2:
Jumlah data: 164
Rata-rata penumpang: 6.92
Rata-rata waktu: 17.20
Klaster 3:
Jumlah data: 172
Rata-rata penumpang: 23.84
Rata-rata waktu: 9.62
Klaster 4:
Jumlah data: 193
Rata-rata penumpang: 22.37
Rata-rata waktu: 17.32

Gambar 5. Hasil analisis karakteristik

Analisis karakteristik menunjukkan bahwa K-Medoids membagi penumpang Suroboyo Bus menjadi empat klaster: Klaster 1 (kepadatan rendah pagi), Klaster 2 (sedang sore-malam), Klaster 3 (tinggi pagi), dan Klaster 4 (sangat tinggi sore). Selanjutnya, medoid akhir ditentukan sebagai titik pusat tiap klaster, diambil dari dataset berdasarkan indeksinya, dengan informasi nomor data, jumlah penumpang, dan waktu operasional, sehingga mewakili karakteristik utama masing-masing klaster.

Medoid Akhir:

Medoid 1: No=221, Total Passenger=9, Time=9.5
Medoid 2: No=173, Total Passenger=7, Time=17.5
Medoid 3: No=491, Total Passenger=24, Time=9.5
Medoid 4: No=497, Total Passenger=22, Time=17.5

Gambar 6. Hasil medoid akhir

Hasil identifikasi medoid menunjukkan empat pusat kluster yang mewakili kepadatan penumpang Suroboyo Bus, dengan Medoid 1–2 mencerminkan kepadatan rendah hingga sedang pada pagi dan sore, sedangkan Medoid 3–4 menunjukkan kepadatan tinggi hingga sangat tinggi pada jam sibuk. Evaluasi klusterisasi menggunakan metode Silhouette dilakukan untuk menilai kualitas pemisahan antar kluster dan keseragaman anggota, dengan data normalisasi dan rentang kluster 2–7. Setiap iterasi melibatkan pemilihan medoid awal, perhitungan jarak Euclidean, penetapan label kluster, dan pembaruan medoid hingga konvergensi, kemudian nilai SSE dianalisis menggunakan metode Elbow untuk menentukan jumlah kluster optimal, sehingga memberikan dasar analitis dan matematis yang kuat sebelum evaluasi Silhouette lebih lanjut.

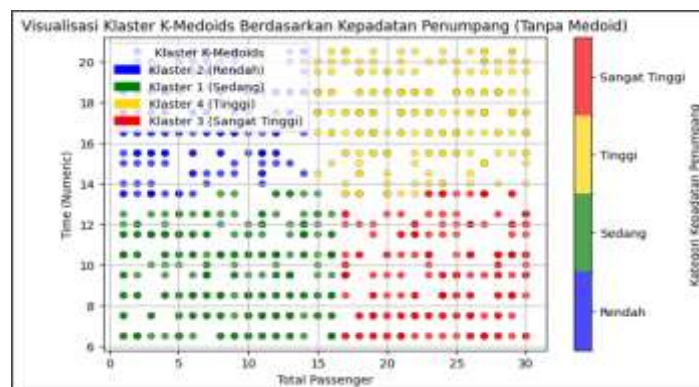
```

--- 4. EVALUASI JUMLAH KLASER OPTIMAL ---
K=2 → Silhouette Manual=0.3569, SSE=940.42
K=3 → Silhouette Manual=0.3819, SSE=587.95
K=4 → Silhouette Manual=0.4222, SSE=362.35
K=5 → Silhouette Manual=0.3718, SSE=313.60
K=6 → Silhouette Manual=0.3621, SSE=254.54
K=7 → Silhouette Manual=0.3523, SSE=223.60

```

Gambar 7. Hasil silhouette tiap kluster

Hasil evaluasi dengan Silhouette Manual dan SSE menunjukkan bahwa jumlah kluster optimal pada metode K-Medoids adalah empat, dengan nilai Silhouette tertinggi 0,4222 dan SSE 362,35, menandakan keseimbangan terbaik antara keseragaman internal dan pemisahan antar kluster. Penambahan kluster lebih dari empat menurunkan nilai Silhouette meskipun SSE lebih rendah, sehingga empat kluster menjadi representasi paling ideal untuk pola kepadatan penumpang berdasarkan jumlah penumpang dan waktu operasional. Tahap visualisasi menampilkan sebaran data menggunakan scatter plot, di mana kluster diurutkan berdasarkan rata-rata penumpang dan diberi kategori “Rendah”, “Sedang”, “Tinggi”, dan “Sangat Tinggi”. Titik-titik diwarnai sesuai indeks kluster dengan legenda, colorbar, judul, label sumbu, dan grid, sehingga memudahkan interpretasi pola kepadatan penumpang secara visual sekaligus mendukung evaluasi numerik klusterisasi.



Gambar 8. Visualisasi Scatterplot K-Medoids

Hasil visualisasi scatterplot klusterisasi K-Medoids memperlihatkan sebaran data penumpang Suroboyo Bus berdasarkan jumlah penumpang dan waktu operasional. Setiap kluster ditandai dengan warna berbeda: kluster rendah (biru), sedang (hijau), tinggi (kuning), dan sangat tinggi (merah). Kluster rendah umumnya berada di sisi kiri bawah dengan jumlah penumpang kecil, sedangkan kluster sangat tinggi terletak di sisi kanan bawah dengan jumlah penumpang besar. Pola ini menunjukkan pemisahan kluster yang jelas dan menegaskan bahwa K-Medoids berhasil mengelompokkan data sesuai tingkat kepadatan penumpang secara efektif.

Kesimpulan

Penelitian ini berhasil menerapkan metode K-Medoids untuk menganalisis pola kepadatan penumpang di setiap halte Suroboyo Bus. Berdasarkan hasil klusterisasi, diperoleh empat klaster utama, yaitu kategori rendah, sedang, tinggi, dan sangat tinggi, dengan jumlah klaster optimal ditentukan menggunakan metode Elbow. Nilai Silhouette Score sebesar 0,4222 menunjukkan bahwa kualitas klusterisasi berada pada tingkat baik dan memiliki pemisahan yang jelas antar klaster. Pola hasil analisis menunjukkan bahwa halte dengan tingkat kepadatan tinggi umumnya berada di kawasan pusat kota dan wilayah transit utama, seperti Purabaya, Joyoboyo, dan Siola. Hasil penelitian ini diharapkan dapat dimanfaatkan oleh Dinas Perhubungan Kota Surabaya sebagai dasar dalam pengambilan keputusan strategis terkait pengaturan jadwal, penambahan armada, dan perencanaan pembangunan halte baru di area dengan kepadatan penumpang tinggi.

Daftar Pustaka

- [1] Y.-H. Kuo, J. M. Y. Leung, and Y. Yan, "Public transport for smart cities: Recent innovations and future challenges," *Eur. J. Oper. Res.*, vol. 306, no. 3, pp. 1001–1026, 2023, doi: <https://doi.org/10.1016/j.ejor.2022.06.057>.
- [2] A. N. B. Prasetyo, M. Maimunah, and P. Sukmasetya, "K-Means Clustering Method for Determining Waste Transportation Routes to Landfill," *J. Ris. Inform.*, vol. 5, no. 3 SE-Articles, pp. 277–284, Jun. 2023, doi: 10.34288/jri.v5i3.219.
- [3] R. S. M. L. Patibandla and V. N., "Clustering Algorithms: An Exploratory Review," N. Tang, Ed., London: IntechOpen, 2021. doi: 10.5772/intechopen.100376.
- [4] C. Wang, S. J. Zhao, Z. Q. Ren, and Q. Long, "Place-Centered Bus Accessibility Time Series Classification with Floating Car Data: An Actual Isochrone and Dynamic Time Warping Distance-Based k-Medoids Method," *ISPRS Int. J. Geo-Information*, vol. 12, no. 7, 2023, doi: 10.3390/ijgi12070285.
- [5] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.
- [6] O. Rahmawati and A. Fauzan, "Provincial Clustering Based on Education Indicators: K-Medoids Application and K-Medoids Outlier Handling," *Barekeng*, vol. 18, no. 2, pp. 1167–1178, 2024, doi: 10.30598/barekengvol18iss2pp1167-1178.
- [7] U. Malikussaleh, A. Utara, U. Islam, and K. Indonesia, "EVALUATING THE QUALITY OF K-MEDOIDS CLUSTERING ON CRIME DATA IN INDONESIA Retno, 1 Rozzi Kesuma Dinata, 2 Novia Hasdyna," vol. 8, no. 2, pp. 274–280, 2024.
- [8] A. Wibowo, M. Makruf, I. Virdyna, and F. C. Venna, "Penentuan Klaster Koridor

- [9] A. Haryanto, “Application of Clustering Method for Segmentation of Traffic Accident Profiling With K-Means And K-Medoids Case Study of Toll Bakauheni – Terbanggi Besar,” vol. 5, no. 8, pp. 9563–9581, 2025.
- [10] Q. Shang, Y. Yu, and T. Xie, “A Hybrid Method for Traffic State Classification Using K-Medoids Clustering and Self-Tuning Spectral Clustering,” *Sustain.*, vol. 14, no. 17, 2022, doi: 10.3390/su141711068.