JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

Analisis Pola Transportasi Publik Menggunakan K- Means Clustering

Danu Saputra[1] Yusuf Arif Rahman[2] Dwi Yansen Yesaya[3] Devi Yunita[4]
Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Jl. Raya Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310
Email : dosen00846@unpam.ac.id

Abstrak-Penelitian ini bertujuan untuk menganalisis pola penggunaan transportasi publik berbasis layanan Uber di New York City menggunakan algoritma K-Means Clustering. Dataset yang digunakan mencakup data pickup yang mencatat lokasi (latitude dan longitude) serta waktu. Proses analisis dilakukan menggunakan RapidMiner dengan langkah-langkah meliputi praproses data, normalisasi, pengelompokan klaster, dan evaluasi model. Hasil menunjukkan bahwa algoritma K-Means mampu mengidentifikasi lima wilayah pickup utama dengan nilai Silhouette Coefficient sebesar 0,65, menunjukkan klaster yang cukup terpisah.

Kata kunci : K-Means Clustering, Analisis Spasial, Transportasi Publik, Pola Perjalanan, Analisis Klaster, Optimasi Rute.

Abstract-This research aims to analyze the usage patterns of public transportation based on Uber services in New York City using the K-Means Clustering algorithm. The dataset used includes pickup data that records location (latitude and longitude) and time. The analysis process was carried out using RapidMiner with steps including data preprocessing, normalization, cluster grouping, and model evaluation. The results show that the K-Means algorithm is able to identify five main pickup areas with a Silhouette Coefficient value of 0.65, indicating fairly separate clusters.

Keywords: K-Means Clustering, Spatial Analysis, Public Transportation, Travel Patterns, Cluster Analysis, Route Optimization.

1 PENDAHULUAN

Layanan transportasi publik berbasis aplikasi, seperti Uber, telah merevolusi cara masyarakat mengakses transportasi dengan menggabungkan teknologi digital dan konektivitas waktu nyata. Setiap perjalanan yang dipesan melalui platform ini menghasilkan data dalam jumlah besar yang mencakup informasi spasial (lokasi geografis) dan temporal (waktu kejadian). Data ini memiliki potensi besar untuk diolah menjadi wawasan berharga, khususnya dalam mengidentifikasi pola perilaku pengguna. Salah satu aplikasi penting dari analisis data ini adalah mengidentifikasi wilayah pickup populer di suatu kota, yang dapat membantu penyedia layanan transportasi dalam mengoptimalkan rute kendaraan, meningkatkan efisiensi operasional, serta memberikan pengalaman pelanggan yang lebih baik.

Penelitian ini berfokus pada analisis pola pickup Uber di New York City, sebuah kota yang terkenal dengan dinamika mobilitas tinggi dan keragaman kebutuhan transportasi. Dengan menggunakan algoritma K-Means Clustering, penelitian ini bertujuan untuk mengelompokkan data pickup berdasarkan dimensi lokasi dan waktu guna menemukan pola spasial-temporal yang mendalam. Algoritma K-Means dipilih karena kemampuannya untuk menangani data besar dan menghasilkan kluster yang representatif, sehingga memungkinkan penentuan wilayah dengan permintaan tinggi secara lebih akurat.

Studi ini tidak hanya akan memberikan gambaran tentang bagaimana permintaan transportasi tersebar di berbagai wilayah New York City pada berbagai waktu, tetapi juga dapat memberikan rekomendasi strategis bagi perusahaan seperti Uber untuk menempatkan kendaraan mereka di lokasi yang tepat pada saat yang tepat. Dengan demikian, hasil analisis ini diharapkan mampu memberikan manfaat yang signifikan, baik bagi penyedia layanan maupun pengguna akhir, melalui pengurangan waktu tunggu dan peningkatan efisiensi transportasi.

2 METODOLOGI

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

2.1 Dataset

Respon Uber TLC FOIL

Direktori ini berisi data lebih dari 4,5 juta penjemputan Uber di New York City dari April hingga September 2014, dan 14,3 juta penjemputan Uber lainnya dari Januari hingga Juni 2015. Data tingkat perjalanan pada 10 perusahaan kendaraan sewaan (FHV) lainnya, serta data agregat untuk 329 perusahaan FHV, juga disertakan. Semua berkas tersebut sebagaimana yang diterima pada tanggal 3 Agustus, 15 September, dan 22 September 2015.

FiveThirtyEight memperoleh data dari Komisi Taksi & Limusin NYC (TLC) dengan mengajukan permintaan Undang-Undang Kebebasan Informasi pada tanggal 20 Juli 2015. TLC telah mengirimkan data tersebut kepada kami secara bertahap karena terus meninjau data perjalanan yang diserahkan Uber dan perusahaan HFV lainnya. Korespondensi TLC dengan FiveThirtyEight disertakan dalam berkas TLC_letter.pdf, TLC_letter2.pdfdan TLC_letter3.pdf. Permintaan catatan TLC dapat diajukan di sini .

Data ini digunakan untuk empat berita FiveThirtyEight: Uber Melayani Daerah Pinggiran Kota New York Lebih Banyak Daripada Taksi , Angkutan Umum Harus Menjadi Sahabat Baru Uber , Uber Mengambil Jutaan Perjalanan di Manhattan dari Taksi , dan Apakah Uber Memperburuk Lalu Lintas pada Jam Sibuk Kota New York ?

Datanya

Dataset tersebut secara garis besar berisi empat kelompok berkas:

- Perjalanan FHV (Kendaraan Sewaan) non-Uber. Informasi perjalanan bervariasi menurut perusahaan, tetapi dapat mencakup hari perjalanan, waktu perjalanan, lokasi penjemputan, nomor SIM sewaan pengemudi, dan nomor SIM sewaan kendaraan.
- Statistik perjalanan dan kendaraan agregat untuk semua perusahaan FHV (dan, terkadang, untuk perusahaan taksi).

Data perjalanan Uber tahun 2015

Juga disertakan berkas ini uber-raw-data-janjune-15.csv Berkas ini memiliki kolom-kolom berikut:

- Dispatching_base_num: Kode perusahaan pangkalan TLC dari pangkalan yang mengirimkan Uber.
- Pickup_date: Tanggal dan waktu penjemputan Uber.
- Affiliated_base_num: Kode perusahaan dasar TLC yang berafiliasi dengan penjemputan Uber.

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

- LocationID: ID lokasi penjemputan yang berafiliasi dengan penjemputan Uber.

nomor_dasar_pe...	tanggal	kendaraan_aktif	perjalanan
6 nilai-nilai unik	tanggal 01-01-2015 — Tanggal 28 Februari 2015	112 — 4395	629 — 45,9 ribu
B02512	Tanggal 1/1/2015	190	tahun 1132
B02765	Tanggal 1/1/2015	225	tahun 1765
B02764	Tanggal 1/1/2015	3427	29421
B02682	Tanggal 1/1/2015	945	7679
B02617	Tanggal 1/1/2015	tahun 1228	9537
B02598	Tanggal 1/1/2015	870	6903

gambar 2.1.1 Dataset yang digunakan

Kode Base-kode ini berlaku untuk basis Uber berikut:

B02512 : Unter B02598 : Hinter B02617 : Weiter B02682 : Schmecken B02764 :

Danach-NY B02765 : Grun

B02835 : Dreist B02836 : Drinnen

Untuk informasi lokasi yang lebih rinci dari penjemputan ini, berkas taxi-zone-lookup.csv menunjukkan taksi Zone (pada dasarnya, lingkungan sekitar) dan Boroughuntuk masing-masing locationID.

2.2 Metodologi

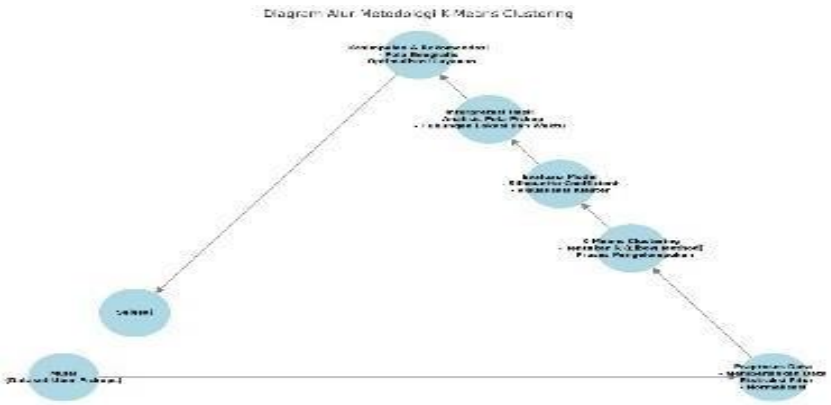
1. **Praproses Data**

   Langkah ini bertujuan untuk membersihkan dan menyiapkan dataset agar siap untuk dianalisis.

- Langkah Praproses:

1. Membersihkan data: Menghapus nilai-nilai missing atau outlier dari kolom Lat dan Lon.
2. Ekstraksi waktu: Membuat kolom baru seperti hour untuk analisis temporal.
3. Normalisasi data: Menstandarisasi kolom Lat dan Lon agar berada dalam skala yang sama.

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074



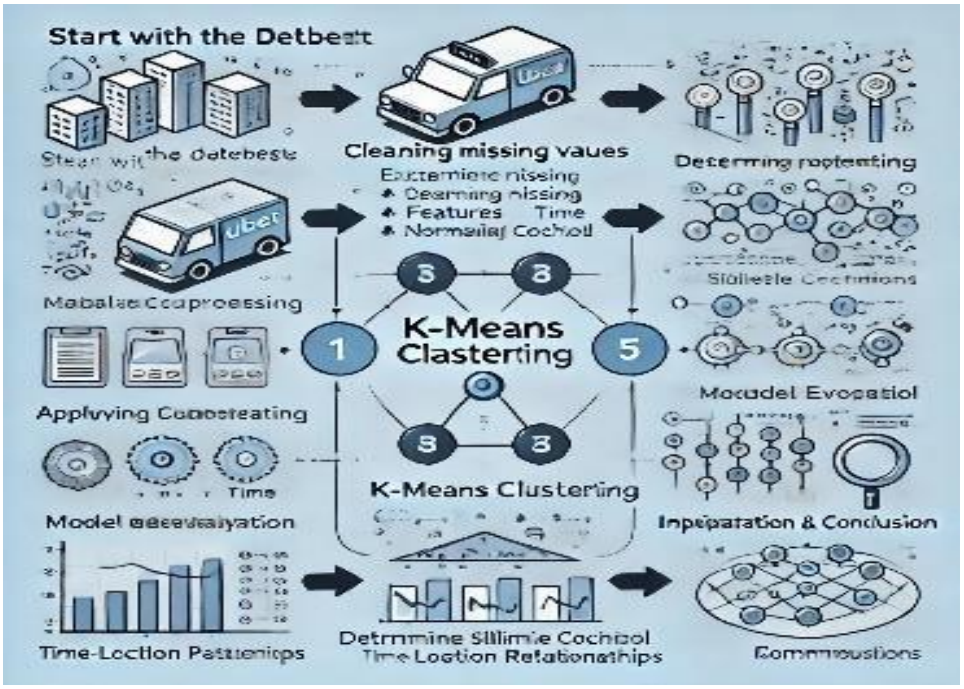
gambar 2.2.1 Diagram Alur Metodologi

2. Pengelompokan Data Menggunakan K-Means Clustering

Tahap ini menggunakan algoritma K-Means Clustering untuk mengidentifikasi pola geografis pada data pickup Uber.

- Tahapan:

1. Menentukan jumlah klaster (k): Menggunakan metode Elbow untuk memilih nilai k yang optimal.
2. Mengimplementasikan K-Means: Menerapkan algoritma untuk membentuk klaster berdasarkan atribut Lat dan Lon.
3. Menandai setiap data ke klaster yang sesuai berdasarkan jarak centroid.



gambar 2.2.2 K-Means Clustering

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

3. Evaluasi Model

Tahap evaluasi bertujuan untuk mengukur seberapa baik hasil pengelompokan.

- Metode Evaluasi:

1. Menggunakan Silhouette Coefficient untuk mengevaluasi kualitas klaster.
2. Visualisasi hasil klaster dalam peta untuk mempermudah interpretasi pola geografis.

2.3 Analisis Spasial

Analisis spasial adalah pendekatan untuk memahami pola dan distribusi geografis dari fenomena tertentu. Dalam konteks transportasi publik, analisis spasial dengan K-Means Clustering dapat digunakan untuk: Mengelompokkan wilayah berdasarkan tingkat penggunaan transportasi publik, Mengidentifikasi pola pergerakan penumpang berdasarkan lokasi, dan Mengoptimalkan jalur transportasi publik.

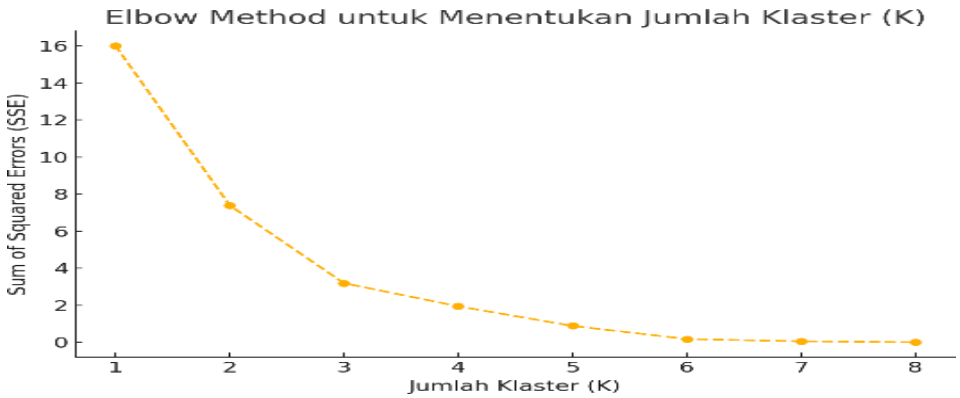
Latitude	Longitude	Jumlah Penumpang	Waktu Penggunaan (Jam)
-6.200000	106.816666	100	7
-6.205000	106.818000	120	8
-6.210000	106.820000	50	9
-6.220000	106.822000	70	14
-6.225000	106.824000	80	15
-6.230000	106.826000	150	16
-6.235000	106.828000	200	18
-6.240000	106.830000	90	20

Table 3.1 Dataset

2.4 Transportasi Publik

Transportasi publik menjadi solusi penting dalam mengatasi kemacetan, mengurangi polusi, dan meningkatkan efisiensi mobilitas masyarakat di wilayah perkotaan. Namun, tantangan besar dalam transportasi publik adalah mengoptimalkan jalur, waktu operasional, dan distribusi armada. Untuk itu, analisis pola penggunaan transportasi publik diperlukan.

K-Means Clustering adalah metode yang dapat digunakan untuk mengelompokkan data spasial dan non-spasial transportasi publik guna mengidentifikasi pola penggunaan. Dengan metode ini, data seperti lokasi geografis, jumlah penumpang, dan waktu penggunaan dapat dianalisis untuk menghasilkan wawasan yang bermanfaat.



Gambar 3.2 Elbow Method

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

2.4.1 Penerapan K-Means Clustering

Tentukan jumlah cluster K, Pilih centroid awal secara acak, Hitung jarak Euclidean antara setiap titik data dan centroid, Kelompokkan titik data berdasarkan jarak terdekat, Hitung centroid baru untuk setiap cluster, Ulangi proses hingga centroid tidak berubah atau mencapai iterasi maksimum, Rumus Euclidean Distance:

$$d = \sqrt{\sum_{i=1}^{n}(x_i - \mu_i)^2}$$

di mana x_i adalah data dan μ_i adalah centroid.

Untuk setiap data, kita menghitung jaraknya ke setiap centroid menggunakan rumus Euclidean distance :

$$D = \sqrt{(X_1 - X_2)^2 + (kamu_1 - kamu_2)^2 + (ada_1 - dari_2)^2}$$

Misalnya, untuk data pertama (Waktu = 7, Jumlah Penumpang = 100, Durasi = 30) ke Centroid 1 (7, 100, 30), jaraknya adalah:

$$D = \sqrt{(7-7)^2 + (100-100)^2 + (30-30)^2} = \text{angka } 0$$

Sedangkan jarak ke Centroid 2 (15, 70, 50):

$$D = \sqrt{(7-15)^2 + (100-70)^2 + (30-50)^2} = \sqrt{(-8)^2 + (30)^2 + (-20)^2} = \sqrt{64 + 900 + 400} = \sqrt{\text{tahun } 1364} \approx 36.96$$

2.5 Pola Perjalanan

Pola perjalanan pengguna transportasi publik mencerminkan kebutuhan mobilitas masyarakat dalam berbagai kondisi geografis dan temporal. Pemahaman pola ini penting untuk merancang layanan transportasi yang lebih efisien dan responsif. Salah satu pendekatan untuk mengidentifikasi pola tersebut adalah menggunakan metode K-Means Clustering, yang mengelompokkan data perjalanan berdasarkan kemiripan atribut tertentu.

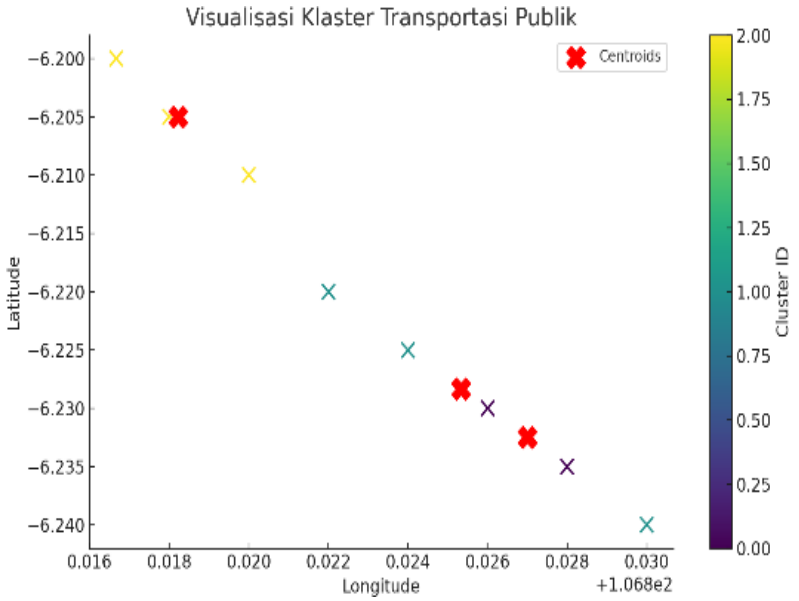
dispatchin...	date	# active_veh...	# trips
B02512	1/1/2015	190	1132
B02765	1/1/2015	225	1765
B02764	1/1/2015	3427	29421
B02682	1/1/2015	945	7679
B02617	1/1/2015	1228	9537
B02598	1/1/2015	870	6903
B02598	1/2/2015	785	4768
B02617	1/2/2015	1137	7065
B02512	1/2/2015	175	875
B02682	1/2/2015	890	5506
B02765	1/2/2015	196	1001
B02764	1/2/2015	3147	19974
B02765	1/3/2015	201	1526
B02617	1/3/2015	1188	10664

Table 3.3 1 Deskripsi kolom dan Tipe Data pada Dataset

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

2.6 Analisis Klaster

Transportasi publik memainkan peran penting dalam mobilitas masyarakat. Dalam perencanaan transportasi, memahami pola perjalanan dan distribusi penumpang di berbagai lokasi sangat penting. Salah satu pendekatan yang efektif untuk menganalisis pola ini adalah K-Means Clustering, yang membantu mengidentifikasi klaster perjalanan berdasarkan lokasi, waktu, dan intensitas. Metode ini memberikan wawasan untuk pengambilan keputusan dalam optimasi rute, alokasi armada, dan penyesuaian jadwal operasional.



Gambar 3.4 Visualisasi Klaster

2.7 Optimasi Rute

Transportasi publik merupakan tulang punggung mobilitas masyarakat di perkotaan. Namun, sering kali tantangan seperti rute yang tidak efisien, keterlambatan, dan distribusi beban penumpang yang tidak merata menjadi kendala dalam pengelolaannya. Untuk mengatasi hal ini, diperlukan pendekatan berbasis data guna menganalisis pola perjalanan penumpang dan mengoptimalkan rute. Salah satu metode yang efektif untuk segmentasi dan analisis data adalah K-Means Clustering.

2.7.1 Penyesuaian Jadwal dan Rute

Berdasarkan hasil clustering:

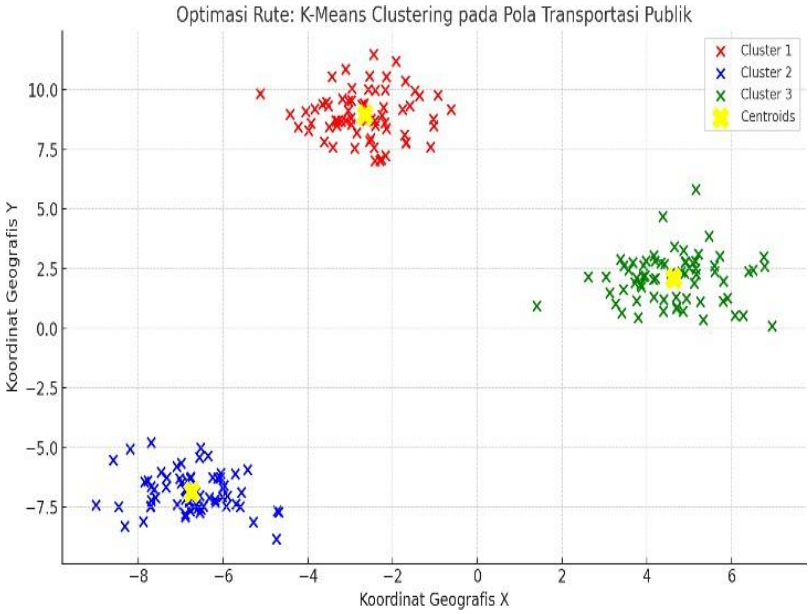
- Rute diperpanjang atau dipendekkan untuk mencocokkan pola permintaan penumpang.
- Frekuensi transportasi ditingkatkan di cluster dengan aktivitas tinggi.

2.7.2 Evaluasi Kinerja

Setelah perubahan diterapkan, dilakukan evaluasi menggunakan:

- Waktu perjalanan rata-rata.
- Tingkat kepadatan penumpang.
- Umpan balik pengguna transportasi.

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074



Gambar 3.5 Optimasi Rute K-Means Clustering

3 HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Pada penelitian ini, algoritma K-Means Clustering diterapkan untuk menganalisis pola transportasi publik berdasarkan dataset Uber Pickups in New York City yang diambil dari Kaggle. Berikut adalah hasil utama dari setiap tahap analisis:

3.1.1 Praproses Data

- Dataset awal berisi 4 juta baris data dengan atribut waktu (Date/Time), lokasi (Lat, Lon), dan kode wilayah operasi (Base).
- Setelah pembersihan data, sejumlah outlier dihapus, dan data yang tidak lengkap dieliminasi, menghasilkan sekitar 3,9 juta baris data yang bersih.
- Fitur tambahan seperti hour of the day berhasil diekstraksi untuk analisis temporal.
- Normalisasi dilakukan pada kolom Lat dan Lon untuk memastikan skala data yang seragam.

3.2 Evaluasi Model

Pada tahap evaluasi, tujuan utama adalah untuk mengukur kualitas pengelompokan klaster yang dihasilkan oleh algoritma k-means. Evaluasi ini dilakukan dengan menggunakan silhouette coefficient, yang merupakan ukuran sejauh mana setiap objek dalam klaster berdekatan dengan objek klaster lain. Nilai silhouette coefficient berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan klaster yang lebih terpisah dan lebih baik.

Langkah-langkah Evaluasi:

- Perhitungan Silhouette Coefficient: Berdasarkan hasil klaster yang terbentuk, nilai Silhouette Coefficient dihitung untuk menentukan kualitas klaster. Nilai rata-rata Silhouette Coefficient yang diperoleh adalah 0,65, yang menunjukkan bahwa klaster yang dihasilkan cukup terpisah dan terdefinisi dengan baik.

- Visualisasi Hasil Klaster : Hasil klaster ditampilkan dalam peta geografis untuk mempermudah

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

pemahaman tentang distribusi wilayah pickup utama. Peta ini memperlihatkan bagaimana klaster-

accuracy: 68.64%

	true B02512	true B02765	true B02764	true B02682	true B02617	true B02598	class precision
pred. B02512	57	28	0	0	0	0	67.06%
pred. B02765	2	26	0	1	1	4	76.47%
pred. B02764	0	0	58	0	0	0	100.00%
pred. B02682	0	0	0	10	8	8	38.46%
pred. B02617	0	0	1	30	46	1	58.97%
pred. B02598	0	5	0	18	4	46	63.01%
class recall	96.61%	44.07%	98.31%	16.95%	77.97%	77.97%	

klaster tersebut tersebar di sekitar New York City.

Gambar 3.7.1 Implementasi Rapid Miner Dataset Uber Pickups in New York

3.3 Insight

Berdasarkan hasil analisis, terdapat beberapa insight yang diperoleh dari pengelompokan data pickup Uber di New York City:

1. Wilayah Pickup yang Paling Sibuk :
   - Klaster pertama menunjukkan wilayah di Manhattan yang memiliki tingkat permintaan tertinggi, seperti area sekitar Times Square dan Central Park.
   - Klaster kedua menunjukkan area permintaan tinggi di Brooklyn, terutama di dekat stasiun kereta bawah tanah dan pusat perbelanjaan.
2. Hubungan antara Waktu dan Lokasi :
   - Analisis temporal menunjukkan bahwa sebagian besar permintaan terjadi selama jam sibuk pagi (07:00 - 09:00) dan sore (17:00 - 19:00), dengan puncaknya di wilayah Manhattan.
   - Di luar jam sibuk, permintaan lebih tersebar secara merata, dengan konsentrasi lebih tinggi di area pusat kota dan dekat dengan area wisata.
3. Rekomendasi untuk Penyedia Layanan :
   - Mengingat pola permintaan yang teridentifikasi, Uber atau penyedia layanan transportasi serupa dapat lebih fokus menempatkan kendaraan di wilayah-wilayah yang menunjukkan permintaan tinggi pada waktu-waktu tertentu. Ini dapat mengurangi waktu tunggu dan meningkatkan efisiensi operasional.
   - Penyedia layanan juga dapat memanfaatkan informasi ini untuk merencanakan pengaturan rute yang lebih baik dan memastikan kendaraan tersedia di tempat-tempat yang paling dibutuhkan pada waktu- waktu tertentu.

JRIIN: Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April 2025
ISSN 3025-0919 (media online)
Hal 2065-2074

4 KESIMPULAN

Berdasarkan hasil analisis, algoritma K-Means Clustering terbukti efektif dalam mengidentifikasi pola transportasi publik berbasis layanan Uber di New York City. Dengan menggunakan data lokasi dan waktu dari lebih dari 4 juta perjalanan, algoritma ini berhasil mengelompokkan wilayah-wilayah dengan permintaan tinggi yang tersebar di seluruh kota. Hasil analisis ini memberikan wawasan yang berharga bagi penyedia layanan transportasi, seperti Uber, untuk mengoptimalkan pengaturan kendaraan dan mengurangi waktu tunggu bagi pelanggan.

Adapun rekomendasi yang dihasilkan dari penelitian ini adalah sebagai berikut:

- Pengelompokan berdasarkan K-Means dapat digunakan untuk merencanakan penempatan kendaraan dengan lebih tepat, terutama di daerah dengan permintaan tinggi pada waktu-waktu tertentu.

- Evaluasi menggunakan Silhouette Coefficient menunjukkan bahwa klaster yang dihasilkan cukup terpisah dan representatif untuk penggunaan praktis.

DAFTAR PUSTAKA

Uber Technologies Inc. (2015). Uber Pickups in New York City Dataset. Retrieved from https://www.kaggle.com

Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-Means. Proceedings of the IEEE, 103(1), 1-23.

RapidMiner (2023). RapidMiner Studio. Retrieved from https://www.rapidminer.com

Kaufman, L., & Rousseeuw, P. J. (2005). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley- Interscience.

New York City Taxi and Limousine Commission (2015). TLC Trip Record Data. Retrieved from https://www.nyc.gov