

Penerapan Algoritma K-Means Dalam Penggunaan Transportasi Online

Aditya Surya Buana

Program Studi Teknik Informatika , Universitas Serang Raya

e-mail: kriessfeld@gmail.com

ABSTRAKSI

Perkembangan teknologi informasi dalam beberapa dekade terakhir telah merevolusi berbagai aspek kehidupan manusia, salah satunya adalah layanan transportasi. Keberadaan platform transportasi online memudahkan mobilitas sehari-hari—pengguna tak lagi perlu menempuh perjalanan mencari kendaraan atau menunggu lama di pinggir jalan, melainkan cukup melakukan pemesanan melalui aplikasi di ponsel pintar. Selain fungsi utama sebagai sarana berpindah tempat, layanan ini juga meluas ke pengantaran barang dan pemesanan makanan, sehingga menjadi ekosistem on-demand yang komprehensif bagi kebutuhan urban modern. Dengan semakin beragamnya cara masyarakat memanfaatkan transportasi online, penting untuk memahami pola pemakaian yang dominan agar penyedia layanan dapat menyesuaikan strategi operasional dan pemasaran. Salah satu pendekatan yang efektif untuk menggali pola tersebut adalah metode clustering, di mana data historis penggunaan dikelompokkan ke dalam segmen-segmen berdasarkan kemiripan karakteristik. Dalam penelitian ini, dua kluster utama dibentuk: C1 bagi pengguna yang kerap memanfaatkan jasa transportasi online, dan C2 bagi mereka yang penggunaannya relatif rendah atau sporadis. Pemisahan ini bertujuan untuk membedakan perilaku pengguna aktif dari pengguna pasif sehingga analisis dapat lebih tajam. Implementasi clustering pada dataset perjalanan online akan melibatkan tahapan pra-proses data—meliputi pembersihan entri duplikat, normalisasi nilai numerik (misalnya frekuensi pemesanan, total pengeluaran), serta encoding variabel kategorikal (jenis layanan, waktu pemesanan)—sebelum algoritma dijalankan. Hasil klasterisasi selanjutnya akan dianalisis untuk mengidentifikasi faktor pendorong di balik tingginya rendahnya intensitas penggunaan, seperti durasi rata-rata perjalanan, responsivitas driver, dan preferensi layanan tambahan (pengantaran makanan atau barang). Dengan memahami karakteristik kedua kluster ini, penelitian berupaya memberikan rekomendasi berbasis data kepada perusahaan transportasi online agar dapat menyusun paket promosi yang lebih tepat sasaran, mengoptimalkan rute dan alokasi armada, serta meningkatkan retensi pengguna. Secara keseluruhan, penelitian ini bertujuan untuk menjawab pertanyaan: “Apa saja kebutuhan dan kebiasaan masyarakat yang paling sering muncul dalam penggunaan transportasi online?”—sehingga layanan dapat terus disempurnakan sesuai preferensi pengguna, sekaligus mendorong pertumbuhan bisnis yang berkelanjutan.

Kata Kunci: clustering, k-means, transportasi online, segmentasi pengguna, analisis pola penggunaan

ABSTRACT

The rapid evolution of information technology over recent decades has revolutionized many aspects of human life, not least in transportation services. Online ride-hailing platforms have simplified daily mobility—users no longer need to scour the streets or wait at stands, but simply place an order via a smartphone app. Beyond point-to-point travel, these platforms have expanded into parcel delivery and food ordering, creating a comprehensive on-demand ecosystem for modern urban needs. As usage of online transportation diversifies, it becomes crucial for service providers to understand dominant usage patterns so they can tailor operational and marketing strategies effectively. One powerful approach to uncover these patterns is clustering, which groups historical usage data into segments based on similar characteristics. In this study, two primary clusters are defined: C1 for frequent users of online transport, and C2 for those who use it less often or sporadically. This segregation enables a clearer distinction between active and passive user behaviors, leading to more focused insights. Applying clustering to the online travel dataset involves several preprocessing steps—removing duplicate entries, normalizing numeric fields (such as booking frequency and total expenditure), and encoding categorical variables (like service type and booking time)—before running the algorithm. The resulting clusters are then analyzed to pinpoint the factors driving high or low usage intensity, such as average trip duration, driver response times, and preferences for additional services like food or parcel delivery. By understanding the profiles of these two clusters, the research offers data-driven recommendations to online transport companies: design more precisely targeted promotional packages, optimize route planning and fleet allocation, and boost user retention strategies. Ultimately, this study seeks to answer the question, “What needs and habits most frequently emerge in online transportation usage?”—enabling continuous refinement of services in line with user preferences and fostering sustainable business growth.

Penulis Korespondensi

Aditya Surya Buana

Tanggal Submit : 26/01/2024
Tanggal Diterima : 13/12/2024
Tanggal Terbit : 28/02/2025

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

I. PENDAHULUAN

Perkembangan teknologi informasi yang berlangsung sangat pesat dalam dua dekade terakhir telah mendorong transformasi hampir semua aspek kehidupan manusia menjadi lebih modern dan terintegrasi (Huda, 2019). Digitalisasi yang awalnya terbatas pada sektor komunikasi dan hiburan kini merambah ke ranah pelayanan publik, termasuk sistem transportasi. Transformasi ini tidak hanya mempercepat akses dan konektivitas, tetapi juga mengubah paradigma mobilitas masyarakat—dari ketergantungan pada kendaraan pribadi dan angkutan umum konvensional menjadi penggunaan moda transportasi yang lebih fleksibel dan on-demand.

Salah satu bentuk nyata dari revolusi transportasi ini adalah munculnya layanan transportasi online sebagai solusi mobilitas urban. Platform seperti Gojek, Grab, dan sejenisnya menyediakan akses jasa transportasi berbasis aplikasi mobile yang memungkinkan pengguna memesan ojek, taksi, hingga layanan pengantaran barang dan makanan dengan beberapa ketukan layar saja. Keberadaan layanan ini memanfaatkan kecanggihan GPS, sistem pembayaran digital, serta algoritma penjadwalan perjalanan real-time, sehingga menawarkan kenyamanan dan efisiensi waktu yang sulit ditandingi oleh moda transportasi tradisional.

Gojek, sebagai pionir layanan transportasi online di Indonesia, telah mengembangkan beragam fitur dalam aplikasinya untuk memenuhi kebutuhan pengguna yang semakin kompleks. Fitur-fitur seperti pemesanan instan, pemantauan lokasi mitra pengemudi secara live, perkiraan tarif otomatis, serta opsi pembayaran tunai maupun non-tunai menunjukkan bagaimana teknologi mampu meningkatkan kualitas layanan (Kuswanto et al., 2019). Selain itu, integrasi Gojek dengan ekosistem digital lainnya—misalnya GoFood, GoSend, dan GoPay—menjadikannya lebih dari sekadar alat transportasi, tetapi juga platform gaya hidup yang sangat berpengaruh di kota-kota besar.

Lebih lanjut, kehadiran Gojek ternyata tidak hanya memberikan kemudahan bagi konsumen, tetapi juga membuka lapangan pekerjaan baru bagi masyarakat yang membutuhkan. Menurut Kathan, Matzler, dan Veider (2016), individu yang kehilangan pekerjaan atau

berstatus pekerja informal kini dapat memperoleh pendapatan dengan memanfaatkan sepeda motor milik sendiri. Mekanisme pendaftaran dan verifikasi mitra pengemudi yang relatif sederhana memungkinkan partisipasi luas, sehingga layanan transportasi online menjadi instrumen inklusif dalam menyerap tenaga kerja di sektor informal.

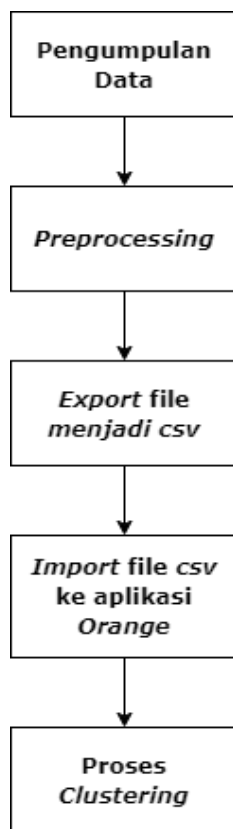
Meski menawarkan berbagai manfaat, kemunculan transportasi online juga memunculkan tantangan dan kontroversi, terutama terkait dampaknya terhadap pendapatan angkutan umum konvensional seperti bus dan taksi. Banyak pengemudi angkutan umum tradisional melaporkan penurunan pendapatan akibat berkompetisi dengan tarif yang lebih murah dari layanan online. Di sisi lain, strategi promosi agresif—seperti potongan tarif, program loyalitas, dan subsidi biaya perjalanan—menjadikan layanan online semakin diminati, terutama oleh kalangan pengguna yang sensitif terhadap harga dan mengutamakan kemudahan transaksi.

Dalam konteks tersebut, penting untuk memahami kebutuhan dan kebiasaan konsumen dalam menggunakan aplikasi transportasi online agar layanan yang disediakan dapat terus disesuaikan. Penelitian clustering dapat digunakan untuk mengelompokkan pola penggunaan berdasarkan frekuensi, jarak tempuh, tujuan perjalanan, dan preferensi layanan tambahan (misalnya pengantaran makanan atau barang). Dengan segmentasi data yang tepat, perusahaan dapat merancang paket promo, penjadwalan armada, dan personalisasi fitur yang lebih efektif untuk setiap segmen pengguna.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis penggunaan konsumen dalam aplikasi transportasi online berdasarkan keperluan mereka—apakah lebih sering untuk keperluan perjalanan harian, pengantaran barang, atau pemesanan makanan. Hasil analisis diharapkan memberikan wawasan berharga bagi penyedia layanan untuk mengoptimalkan pengalaman pengguna, meningkatkan kepuasan pelanggan, dan merancang strategi bisnis yang lebih responsif terhadap dinamika kebutuhan masyarakat urban.

II. METODE PENELITIAN

Metode yang digunakan pada penelitian ini Adalah Algoritma K-Means Clustering dan aplikasi Orange untuk membuktikan hasil perhitungannya. Clustering adalah salah satu teknik dari data mining yang bertujuan untuk mengelompokan suatu data berdasarkan suatu kesamaan-kesamaan yang ada pada suatu data (T. Syahputra, J. Halim, dan E. P. Sintho, 2018). K-Means merupakan suatu algoritma yang cukup mudah serta efektif untuk menemukan cluster yang ada dalam data (G. A. Marcoulides, 2005). Orange adalah suatu aplikasi open source yang dapat digunakan untuk melakukan analisis serta visualisasi data. Aplikasi ini disukai karena faktor kendalan serta inovatif. Aplikasi ini juga mempermudah penggunaannya karena menyajikan informasi dalam bentuk visual serta intuitif (R. A. Raffaidy Wiguna dan A. I. Rifai, 2021). Agar penelitian ini lebih terarah dan sistematis kerangka penelitian ini dapat dilihat pada Gambar. 1.



Gambar 1. Kerangka Penelitian

1. Pengumpulan Data

Teknik Pengumpulan Data Pengumpulan data adalah teknik atau cara yang dilakukan oleh peneliti untuk mengumpulkan data (Br Sembiring et al., 2022). Data yang didapat adalah data pengguna transportasi online Gojek yang berjumlah 218 data, data tersebut bersumber dari databoks.katadata.co.id yang di download pada tanggal 5 Januari 2024. Namun data tersebut tidak bisa langsung digunakan karena terdapat beberapa data yang tidak termasuk ke dalam tujuan

penelitian, jadi harus dilakukan pemilihan atau pembersihan data.

2. Preprocessing

Pada tahap ini, dilakukan pemilihan terhadap himpunan data yang sebelumnya telah diperoleh lalu mengubahnya menjadi format yang mungkin diproses dalam model yang telah didefinisikan lebih lanjut (Akhmad Budi, 2019). Dataset yang akan digunakan tidak dapat langsung diproses melainkan harus dibersihkan terlebih dahulu.

3. Export menjadi file csv

Setelah dibersihkan, copy dataset tersebut kedalam Microsoft Excel lalu save atau export dengan format csv, simpan file yang sudah di export tersebut ke dalam folder yang diinginkan.

4. Import file csv ke aplikasi Orange

Setelah dataset sudah siap digunakan, langkah selanjutnya adalah membuka aplikasi Orange lalu menyiapkan alat atau widget yang diperlukan untuk proses clustering, setelah itu import dataset yang sebelumnya sudah disimpan ke dalam aplikasi Orange.

5. Proses Clustering

Proses Clustering adalah metode pengelompokan sample ke dalam beberapa group atau subset atau cluster berdasarkan “kemiripannya” dengan instance lain. Pada prinsipnya clustering hampir sama dengan classification, hanya saja dataset yang digunakan tidak berpasangan atau tidak berlabel. Dataset semacam ini dapat dijumpai di sekitar kita dan jenisnya relatif cukup banyak. Sebagai contoh, perhatikan dataset jenis buku dan jumlah halamannya (Primartha R., 2021).

III. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan cara menghitung data yang tersedia lalu membandingkan hasil dari perhitungan tersebut menggunakan aplikasi Orange, tahapan-tahapan yang akan dilakukan dapat dijabarkan sebagai berikut:

1. Tahap Pengumpulan Data

Setelah data yang diperlukan telah terkumpul dan sudah melalui proses Preprocessing, maka data yang tersebut akan terlihat seperti pada Tabel 1.

Tabel 1. Dataset Bersih

No.	mode	jarak (km)	harga (rp)
1	BIKE	5,55	2500
2	FOOD	7,08	9000
3	SHOP	4,02	5000
4	FOOD	8,94	12000
....
....
218	CAR	7,13	13200

2. Tahap Perhitungan

Metode pada penelitian ini menggunakan algoritma K-Means, dimana algoritma K-Means melakukan clusterisasi pada data yang diuji. Tahap dalam pengelompokan menggunakan algoritma K-Means menurut Nur wakhidah (Wakhidah, 2010) adalah sebagai berikut:

- 1) Menentukan jumlah cluster dan centriod awal
- 2) 2. Menentukan jarak data dan pusat cluster menggunakan Euclidean Distance yang di rumuskan sebagai berikut:

$$D(i, j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2}$$

Dimana :

$D(i, j)$ = Jarak dari data ke i ke pusat cluster j

Xki = Data ke i pada atribut data ke k

Xkj = Titik pusat ke j pada atribut ke k

- 3) Kelompokan data kedalam cluster yang dengan jarak paling dekat
- 4) Hitung Pusat cluster yang baru dengan mencari nilai rata-rata pada kelompok data
- 5) Proses penentuan pusat cluster dan penempatan data dalam cluster di ulangi sampai nilai centroid tidak berubah lagi.

Dikarenakan data yang ada sangat banyak dan akan sangat sulit jika menghitungnya secara manual maka dilakukan perhitungan dengan menggunakan bantuan aplikasi Microsoft Excel 2021. Cluster yang akan digunakan adalah yang sering digunakan (cluster 1) dan jarang digunakan (cluster 2), dapat dilihat pada Tabel 2.

Tabel 2. Data Centroid

cluster	centroid (x)	centroid (y)
Sering (C1)	1	1000
Jarang (C2)	1	6000

Selanjutnya adalah menghitung jarak centroid dengan menggunakan rumus Euclidean Distance lalu mengelompokan data cluster dengan jarak paling dekat, sehingga diperoleh hasil sebagai berikut:

Tabel 3. Data Cluster Iterasi 1

mode	C1	C2	Cluster
Bike	0,2890	0,2109	C2
Car	0,1287	0,7876	C1
Food	0,6452	0,1452	C2
Shop	0,7142	0,2142	C2

Dari data diatas kita dapat melihat bahwa kelompok cluster 1 hanya ada 1 mode sedangkan cluster 2 terdapat 3 mode, supaya lebih mudah untuk melihat kelompok cluster dibuatlah tabel:

Tabel 4. Kelompok Cluster ke-1

Cluster	Anggota Cluster
C1	Car
C2	Bike,Food,Shop

Setelah itu dilakukan perhitungan ke-2 yang berguna untuk membandingkan hasil antara data dari perhitungan sebelumnya dan data dari perhitungan yang sedang dilakukan ini sama atau tidak, cluster dan centroid yang digunakan pada perhitungan ke-2 dapat dilihat pada tabel dibawah:

Tabel 5. Data Centroid Baru

cluster	centroid (x)	centroid (y)
Sering (C1)	8,0737	0,3890
Jarang (C2)	6,5946	0,9824

Kemudian langkah selanjutnya sama seperti sebelumnya yaitu menghitung jarak centroid dengan menggunakan rumus Euclidean Distance lalu mengelompokan data cluster dengan jarak paling dekat, sehingga diperoleh hasil sebagai berikut:

Tabel 6. Data Cluster Iterasi 2

mode	C1	C2	Cluster
Bike	0,2529	0,3404	C1
Car	0,7997	0,2063	C2
Food	0,3600	0,2333	C2
Shop	0,3776	0,2157	C2

Hasil dari perhitungan diatas kita kelompokkan kedalam tabel supaya dapat lebih mudah dilihat:

Tabel 7. Kelompok Cluster ke-2

Cluster	Anggota Cluster
C1	Bike
C2	Car,Food,Shop

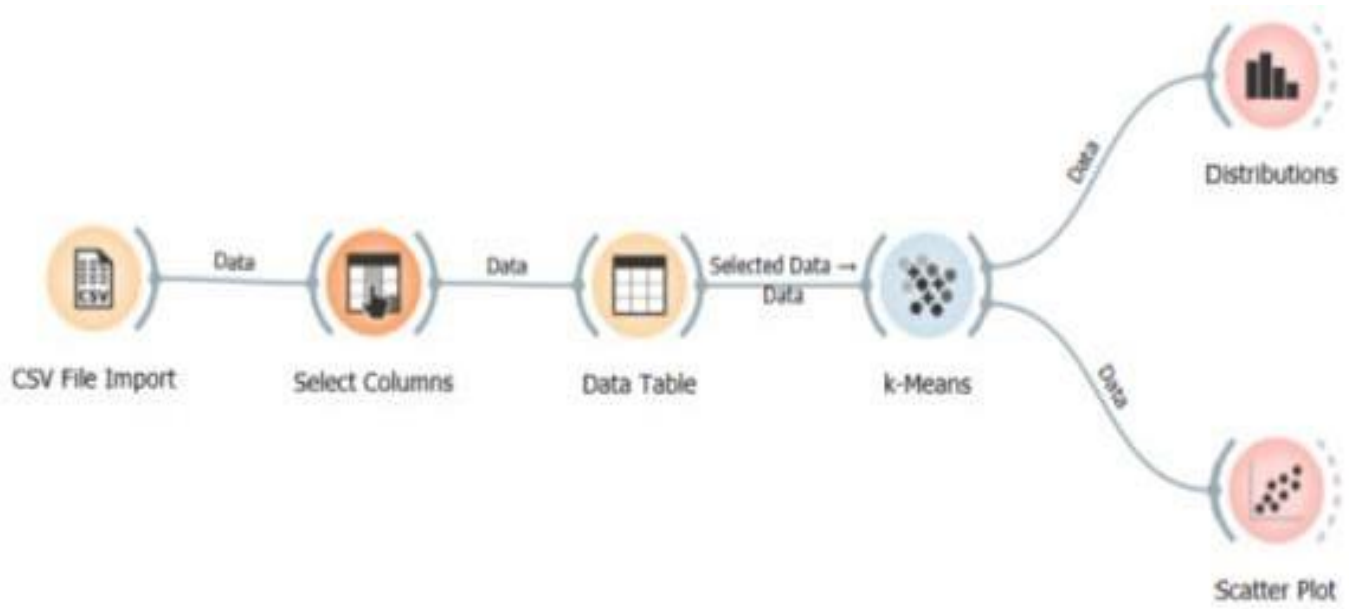
Dapat dilihat bahwa hasil dari perhitungan pertama dan kedua berbeda, maka akan dilakukan perhitungan berulang kali hingga hasil akhir yang didapat dari data yang diperoleh sama. Dalam penelitian ini dilakukan perhitungan berulang sebanyak 3x hingga diperoleh data perhitungan yang sama dengan data sebelumnya, hasil akhir dari perhitungan dapat dilihat pada tabel dibawah:

Tabel 8. Kelompok Cluster Akhir

Cluster	Anggota Cluster
C1	Bike
C2	Car,Food,Shop

3. Tahap Pembuktian

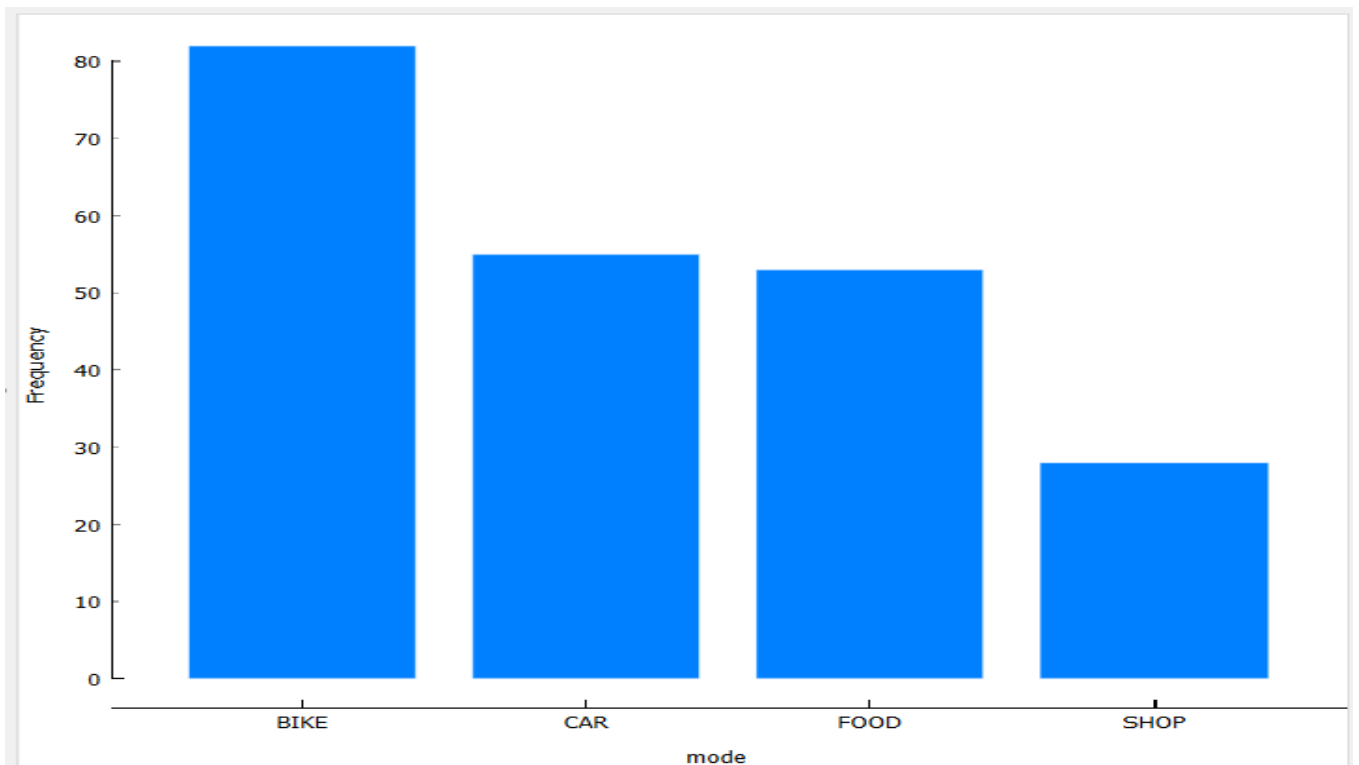
Setelah dilakukan perhitungan maka tahap selanjutnya yang dilakukan adalah pembuktian menggunakan aplikasi Orange untuk membuktikan apakah hasil perhitungan yang telah dilakukan itu benar atau tidak. Hal pertama yang harus dilakukan adalah menyiapkan widget yang akan digunakan lalu mengimport data file csv yang sebelumnya sudah kita simpan ke dalam aplikasi Orange (lihat Gambar 2).



Gambar 2. Persiapan Widget

Langkah selanjutnya adalah mengatur widget yang sudah ditambahkan tadi supaya hasil perhitungan yang dicari dapat terlihat, setelah selesai diatur kita dapat

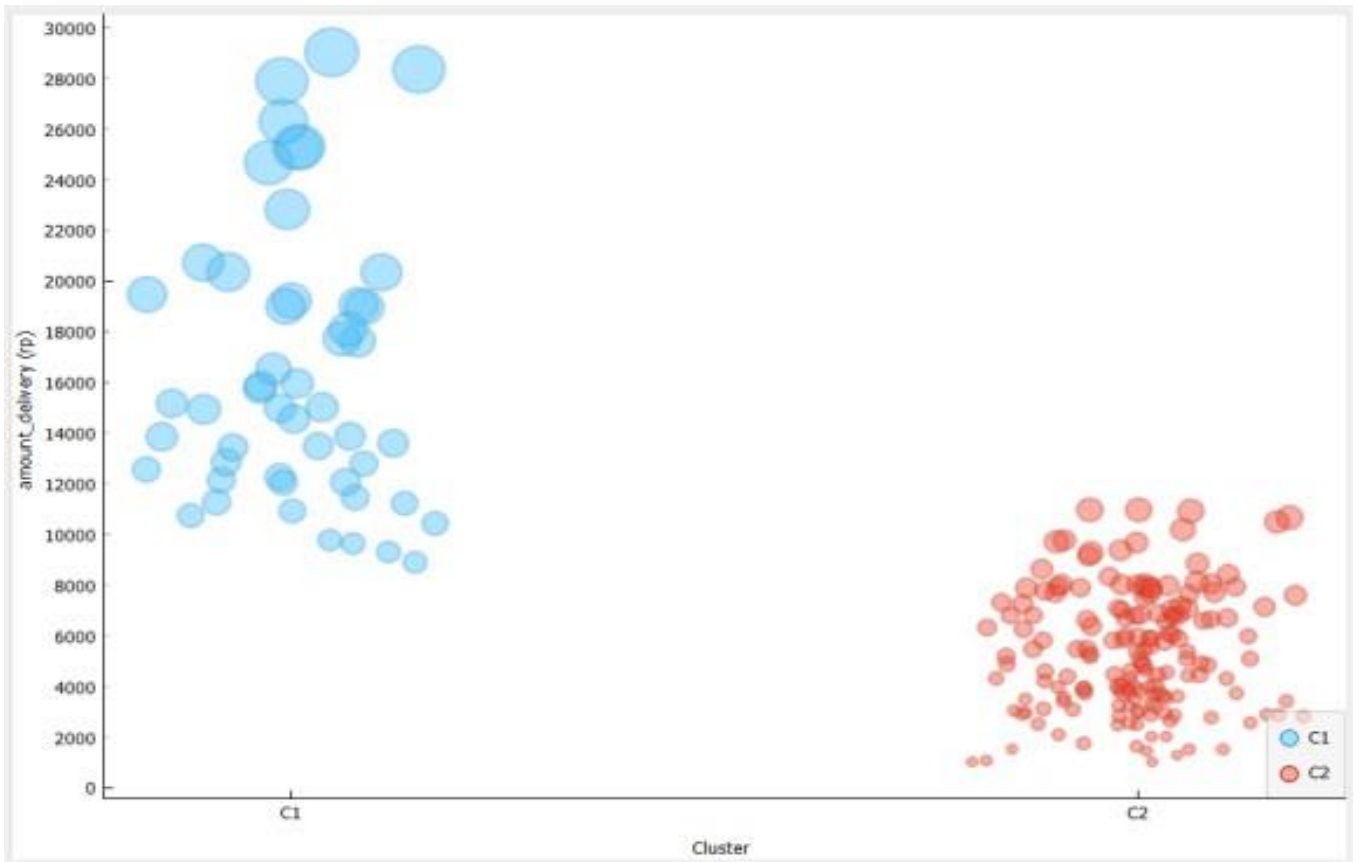
melihat pada bagian widget Distributions seperti pada Gambar 3



Gambar 3. Distributions

Dari gambar persentase distribution diatas dapat dilihat bahwa parameter untuk cluster 1 mode Bike lebih tinggi dari cluster 2 mode Car, Food dan Shop, hasil perhitungan yang ditampilkan pada gambar

tersebut sama dengan hasil perhitungan pada kelompok cluster akhir pada Tabel 8. Supaya lebih jelas kita dapat melihat hasil penyebaran cluster menggunakan Scatter Plot pada Gambar 4.

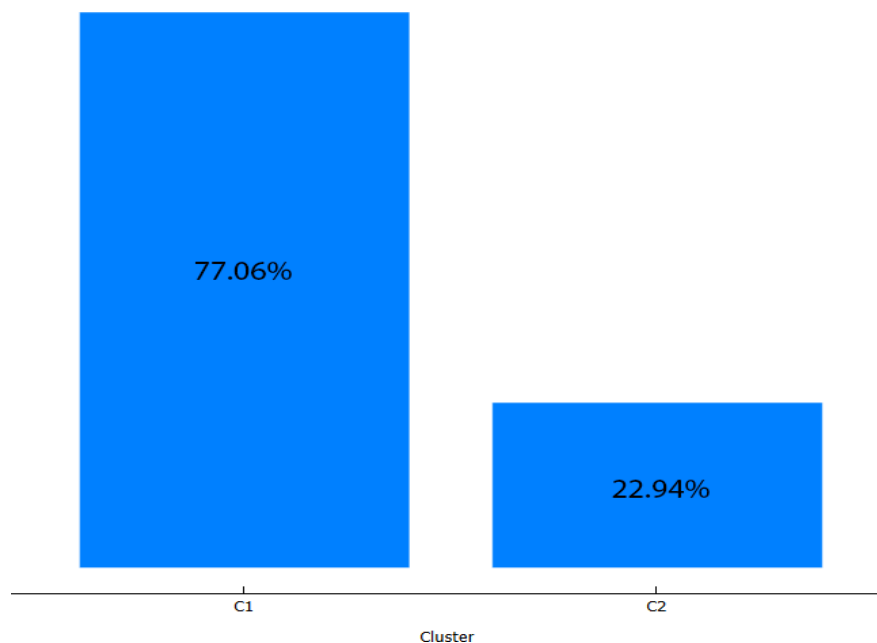


Gambar 4. Scatter Plot

4. 4. Hasil Perhitungan

Berdasarkan perhitungan yang telah dilakukan dapat dibuktikan bahwa hasil dari perhitungan tersebut benar karena hasil yang muncul sama dengan hasil perhitungan menggunakan aplikasi Orange. Dapat

disimpulkan bahwa nilai tertinggi persentase cluster C1 untuk Bike adalah 77.06% sedangkan persentase cluster C2 untuk Car, Food dan Shop adalah 22.94%. Persentase dari masing-masing cluster dapat dilihat pada Gambar 5.



Gambar 5. Hasil Perhitungan

V. KESIMPULAN

Penelitian ini mengungkap pola preferensi pengguna dalam memanfaatkan berbagai layanan Gojek—baik transportasi konvensional maupun pengantaran online—sehingga mampu menjadi landasan strategis bagi pengembang aplikasi untuk merancang fitur yang lebih relevan. Misalnya, dengan mengetahui segmen pengguna yang paling sering memilih mode Bike, tim produk dapat mengimplementasikan skema diskon dinamis atau program loyalitas khusus yang meningkatkan retensi dan kepuasan pelanggan. Selain itu, pemahaman mendalam tentang frekuensi pemesanan, jarak tempuh, dan jenis layanan yang paling diminati memungkinkan Gojek mengoptimalkan alokasi armada dan penyebaran mitra pengemudi secara real time, sehingga efisiensi operasional meningkat dan tingkat ketersediaan layanan di area padat permintaan dapat dijaga secara konsisten. Dengan demikian, hasil analisis data tidak hanya memperkaya wawasan internal, tetapi juga mendorong inovasi berkelanjutan dalam pengembangan fitur, mulai dari peningkatan antarmuka pengguna hingga penambahan opsi multimodal dan integrasi layanan lintas platform.

Lebih jauh, temuan penelitian ini menyediakan kerangka kerja bagi tim pemasaran Gojek untuk merancang kampanye promosi yang lebih tepat sasaran dan hemat biaya. Berdasarkan segmentasi perilaku konsumen, pemasaran dapat difokuskan pada kelompok yang responsif terhadap insentif tertentu—seperti penawaran bundle antar layanan GoRide dan GoSend atau program referral untuk pengguna baru—serta memilih kanal komunikasi yang paling efektif, baik melalui notifikasi in-app, email, maupun media sosial. Data penggunaan yang tersaji juga bisa digunakan untuk melakukan evaluasi keberhasilan promosi sebelumnya dan menyesuaikan taktik ke depan, misalnya menyesuaikan durasi promo atau besaran voucher berdasarkan tingkat aktivitas setiap klaster pengguna. Dengan memadukan hasil analisis kuantitatif dan feedback kualitatif, Gojek dapat memastikan strategi pemasaran dan pengembangan produk selaras dengan kebutuhan riil konsumen, sehingga menciptakan sinergi antara peningkatan pengalaman pengguna dan pertumbuhan bisnis yang berkelanjutan.

DAFTAR PUSTAKA

- [1]. N. Huda, “Analisis Kinerja Website Pt Pln (Persero) Menggunakan Metode Pieces,” *Sistemasi*, vol. 8, no. 1, pp. 78-89, 2019
- [2]. Kuswanto, A. et al. (2019) ‘The determinants of customer loyalty in the Indonesian ridesharing services: offline vs online’, *Innovation & Management Review, ahead-ofp (ahead-of-print)*.
- [3]. Kathan, W., Matzler, K. and Veider, V. (2016) ‘The sharing economy: Your business model’s friend or foe?’, *Business Horizons*. Kelley School of Business, Indiana University, 59(6), pp. 663–672.
- [4]. T. Syahputra, J. Halim, and E. P. Sintho, “Penerapan Data Mining Dalam Menentukan Pilihan Jurusan Bidang Studi SMA Menggunakan Metode,” Penerapan Data Min. dalam Menentukan Pilihan Jur. di Bid. Stud. SMA menggunakan Metod. Clust. Dengan Tek. Single Link. *JURTEKSI*, vol. IV, no. 2, pp. 1–4, 2018.
- [5]. G.A. Marcoulides, *Discovering Knowledge in Data: an Introduction to Data Mining*, vol. 100, no. 472. 2005.
- [6]. R. A. raffaidy Wiguna and A. I. Rifai, “Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining,” *J. Inf. Syst. Informatics*, vol. 3, no. 1, pp. 1–12, 2021,
- [7]. Br Sembiring, S. N., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 1(1), 31.
- [8]. Akhmad Budi, dan. (2019). Klasterisasi Indeks Pembangunan Manusia (Ipm) Per Kabupaten Di Indonesia Dengan Menggunakan Algoritma K-Means.
- [9]. Primartha, R. (2021). Algoritma Mechine Learning. Bandung : Penertbit Infromatika.
- [10]. Wakhidah, N. (2010). Clustering Menggunakan K-Means Algorithm. *Jurnal Transformatika*, 8(1), 33.