



Implementasi Algoritma *K-Means* untuk Pengelompokan Pengguna QRIS Tap pada Moda Transportasi Umum di Wilayah Jabodetabek

Chintya Eka Paramitha*, Arief Wibowo

Universitas Budi Luhur

DOI:

<https://doi.org/10.53697/emak.v7i1.3685>

*Correspondence:

Email:

2531600027@student.budiluhur.ac.id

Received: 22-11-2025

Accepted: 22-12-2025

Published: 22-01-2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: This study aims to segment QRIS Tap users in public transportation payments through the Bayarind e-wallet application. The K-Means clustering method is employed using factual, non-perceptual data, including users' age, duration of application usage, frequency of QRIS Tap usage, and average transaction value. The dataset consists of 63 QRIS Tap users in the Jabodetabek area, which has been transformed into numerical form. Cluster evaluation is conducted using the Davies–Bouldin Index (DBI) and Within-Cluster Sum of Squares (WCSS) to determine the optimal number of clusters. The scientific contribution of this study lies in the application of user segmentation based on actual behavioral data (non-perceptual) within the specific context of QRIS Tap usage for public transportation payments, a topic that remains limited in prior studies. Furthermore, this study integrates DBI and WCSS as complementary evaluation metrics to ensure a more objective and robust cluster configuration. The results indicate that a four-cluster configuration ($K = 4$) provides the most informative segmentation. These clusters represent new users with low activity, loyal users with moderate transaction levels, experienced users with diverse transaction patterns, and premium users with high transaction frequency and value. This segmentation offers empirical insights into QRIS Tap user characteristics and serves as a strategic foundation for decision-making in the development of digital payment systems and public transportation services.

Keywords: E-Wallet, K-Means Clustering, Public Transportation, QRIS Tap

Pendahuluan

Perkembangan teknologi digital telah mendorong transformasi signifikan dalam sistem keuangan, khususnya pada layanan pembayaran. Masyarakat secara bertahap beralih dari transaksi tunai (cash-based) menuju sistem pembayaran digital (cashless society) yang dinilai lebih cepat, efisien, dan aman (Akram et al, 2024) (Ananda & Sulisty, 2025). Di Indonesia, salah satu inovasi nasional dalam sistem pembayaran digital adalah QRIS (Quick Response Code Indonesian Standard) yang dikembangkan oleh Bank Indonesia untuk menyatukan berbagai standar pembayaran berbasis QR code ke dalam satu sistem nasional yang interoperable Bank Indonesia, (2020). Kehadiran QRIS bertujuan untuk meningkatkan efisiensi transaksi sekaligus memperluas inklusi keuangan di berbagai lapisan masyarakat (Davies & Bouldin, 1979) (RapidMiner, 2023).

Sebagai pengembangan dari QRIS, Bank Indonesia memperkenalkan QRIS Tap yang memanfaatkan teknologi Near Field Communication (NFC), sehingga pengguna dapat melakukan transaksi hanya dengan menempelkan (tap) perangkat ke reader merchant tanpa perlu melakukan pemindaian kode secara manual (Indonesia, 2023) (Kim & Mirusmonov, 2020). Teknologi ini dinilai mampu mempercepat waktu transaksi dan meningkatkan kenyamanan pengguna, terutama pada layanan berintensitas tinggi. Salah satu platform yang telah mengimplementasikan teknologi ini adalah aplikasi e-wallet Bayarind, yang digunakan dalam berbagai transaksi masyarakat, termasuk pembayaran transportasi umum seperti Transjakarta, KRL (Commuter Line), LRT, dan MRT. Integrasi QRIS Tap pada layanan tersebut memungkinkan transaksi dilakukan secara cepat, minim kontak fisik, serta mendukung prinsip higienitas pascapandemi COVID-19.

Meskipun demikian, pemanfaatan QRIS Tap pada pembayaran transportasi umum menunjukkan tingkat adopsi dan pola penggunaan yang beragam antar pengguna. Variasi ini tercermin dari perbedaan karakteristik demografis, lama penggunaan aplikasi, frekuensi transaksi, serta nilai transaksi yang dilakukan. Oleh karena itu, diperlukan pendekatan analitis berbasis data yang mampu mengelompokkan pengguna berdasarkan pola penggunaan aktual (non-persepsi), bukan hanya berdasarkan persepsi subjektif pengguna. Salah satu pendekatan yang relevan untuk tujuan tersebut adalah metode clustering. Di antara berbagai teknik klasterisasi yang tersedia, seperti K-Medoids dan hierarchical clustering, algoritma K-Means Clustering dipilih karena kemampuannya dalam mengelompokkan data numerik berdasarkan tingkat kemiripan secara objektif dan efisien (Han et al, 2012) (Jain, 2010).

Metode data mining berbasis unsupervised learning seperti K-Means Clustering banyak digunakan dalam penelitian segmentasi pengguna karena kemampuannya mengelompokkan data numerik dalam jumlah besar secara efisien. Beberapa penelitian terkini menunjukkan bahwa K-Means efektif diterapkan pada analisis perilaku pengguna sistem pembayaran digital dan layanan transportasi berbasis teknologi (Rahman & Putri, 2022) (Wang et al, 2021). Selain itu, algoritma ini dinilai mampu menghasilkan klaster yang interpretatif apabila variabel yang digunakan bersifat faktual dan terukur Zhang et al, (2023).

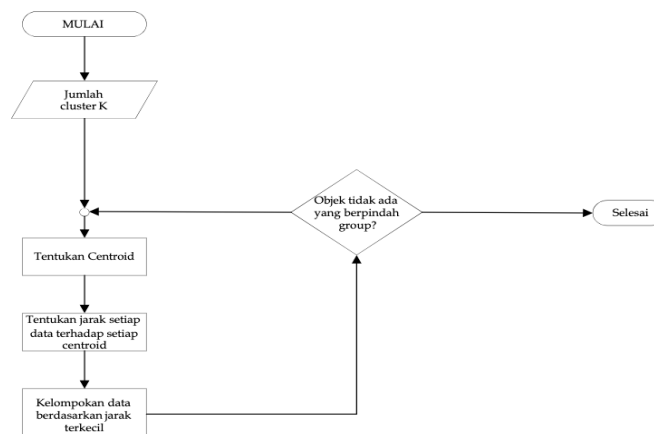
Sejumlah penelitian terdahulu telah memanfaatkan algoritma K-Means untuk pengelompokan data pada berbagai bidang. Akram et al, (2024) menerapkan K-Means untuk mengelompokkan prestasi belajar siswa berdasarkan skor akademik dan menghasilkan tiga klaster dengan tingkat kemampuan yang berbeda. Yusuf & Munandar, (2025) menggunakan K-Means untuk mengelompokkan wilayah berdasarkan indikator kesejahteraan publik dan menemukan lima klaster dengan karakteristik sosial-ekonomi yang berbeda. Hidayat & Kusniyanti, (2022) menerapkan K-Means dalam pengelompokan data penjualan guna memahami pola transaksi, sementara Irawati et al, (2025) memanfaatkan K-Means yang divalidasi dengan Davies–Bouldin Index (DBI) dan Elbow Method untuk memetakan kinerja ekspor kopi Indonesia berdasarkan negara tujuan. Meskipun demikian, sebagian besar penelitian tersebut belum secara spesifik mengkaji

segmentasi pengguna sistem pembayaran digital berbasis perilaku aktual dalam konteks transportasi umum.

Berdasarkan celah penelitian tersebut, novelty penelitian ini terletak pada penerapan K-Means Clustering untuk melakukan segmentasi pengguna QRIS Tap pada pembayaran transportasi umum melalui aplikasi e-wallet Bayarind dengan menggunakan data faktual non-persepsi. Proses klusterisasi dilakukan dengan menguji beberapa skenario jumlah kluster dan dievaluasi menggunakan kombinasi Davies–Bouldin Index (DBI) dan Within-Cluster Sum of Squares (WCSS) guna memperoleh konfigurasi kluster yang paling optimal dan informatif.

Metodologi

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining berbasis unsupervised learning. Algoritma yang digunakan adalah K-Means Clustering, dengan tujuan untuk mengelompokkan pengguna QRIS Tap pada pembayaran transportasi umum melalui aplikasi e-wallet Bayarind berdasarkan pola penggunaan aktual. Data yang digunakan merupakan data primer yang diperoleh melalui kuesioner terhadap pengguna QRIS Tap pada layanan transportasi umum di wilayah Jabodetabek. Dataset terdiri dari 63 responden, di mana setiap responden direpresentasikan sebagai satu objek data. Variabel yang digunakan bersifat faktual dan non-persepsi, meliputi usia pengguna (X1), lama penggunaan aplikasi e-wallet (X2), frekuensi penggunaan QRIS Tap (X3), dan rata-rata nilai transaksi (X4).



Gambar 1. Flowchart Algoritma K-Means Klusterisasi

Dalam proses evaluasi kualitas kluster, penelitian ini menggunakan Davies–Bouldin Index (DBI) sebagai ukuran validitas internal kluster. DBI banyak digunakan dalam penelitian terkini karena mampu menilai tingkat kohesi intra-kluster dan separasi antar-kluster secara objektif, khususnya pada data pengguna layanan digital dan finansial (Kumar & Singh, 2021) (Prasetyo et al, 2024). Nilai DBI yang semakin kecil menunjukkan struktur kluster yang semakin optimal. Proses pengelompokan data menggunakan algoritma K-Means diawali dengan penentuan jumlah kluster (K) yang dievaluasi menggunakan metode Davies–Bouldin Index (DBI) untuk memperoleh jumlah kluster optimal. Selanjutnya,

centroid awal untuk setiap kluster ditentukan secara acak atau berdasarkan nilai yang merepresentasikan karakteristik data. Jarak antara setiap data dan centroid dihitung menggunakan metode Euclidean Distance, kemudian data dikelompokkan ke dalam kluster dengan jarak terdekat. Posisi centroid diperbarui dengan menghitung nilai rata-rata seluruh data dalam masing-masing kluster, dan proses ini dilakukan secara iteratif hingga tidak terjadi lagi perubahan posisi centroid maupun perpindahan anggota kluster, yang menandakan bahwa proses klusterisasi telah mencapai kondisi konvergen. Jarak antara data dan centroid dihitung menggunakan Euclidean Distance sebagai berikut:

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

dengan keterangan:

$D(x,y)$ adalah jarak antara data ke- x dan centroid ke- y ,

x_i adalah nilai atribut ke- i pada data,

y_i adalah nilai atribut ke- i pada centroid, dan

n adalah jumlah atribut yang digunakan dalam proses klusterisasi.

Hasil dan Pembahasan

Hasil segmentasi dengan empat kluster ($K = 4$) sejalan dengan penelitian sebelumnya pada konteks e wallet dan pembayaran digital, yang menunjukkan bahwa jumlah kluster antara tiga hingga lima mampu merepresentasikan heterogenitas perilaku pengguna secara optimal (Setiyawan & Rizki, 2023) (Wijaya & Hartono, 2024). Segmentasi berbasis frekuensi dan nilai transaksi juga terbukti efektif dalam mengidentifikasi kelompok pengguna dengan tingkat adopsi yang berbeda Li & Huang, (2022). Perbedaan utama penelitian ini dibandingkan studi sebelumnya terletak pada penggunaan data perilaku aktual (non persepsi) dalam konteks QRIS Tap pada transportasi umum. Jika sebagian besar studi terdahulu berfokus pada niat penggunaan atau kepuasan, penelitian ini mengelompokkan pengguna berdasarkan perilaku transaksi riil, sehingga memberikan kontribusi empiris yang lebih operasional bagi pengambil kebijakan (Permatasari & Cahyono, 2022) (Valencia-Arias & others, 2025).

Skema Encoding

Data responden pada penelitian ini berbentuk data kategorikal sehingga perlu ditransformasikan ke dalam bentuk numerik agar dapat diproses menggunakan algoritma K-Means. Proses encoding dilakukan secara ordinal, di mana kode numerik yang lebih besar merepresentasikan tingkat atau intensitas kategori yang lebih tinggi. Skema encoding diterapkan pada empat variabel penelitian, yaitu usia (X1), lama penggunaan aplikasi e-wallet (X2), frekuensi penggunaan QRIS Tap (X3), dan nilai transaksi (X4). Variabel usia dikodekan berdasarkan kelompok rentang usia, sedangkan lama penggunaan aplikasi dikategorikan berdasarkan durasi penggunaan dari kurang dari 3 bulan hingga lebih dari 1 tahun. Frekuensi penggunaan QRIS Tap diklasifikasikan berdasarkan intensitas penggunaan mingguan, dan nilai transaksi dikategorikan berdasarkan rentang besaran rata-rata transaksi.

Seluruh kategori pada masing-masing variabel diberikan kode numerik secara berurutan untuk menjaga keterurutan data dan memungkinkan proses klasterisasi dilakukan secara objektif. Dataset hasil encoding selanjutnya digunakan sebagai input dalam proses klasterisasi K-Means. Data yang digunakan bersifat faktual dan non-persepsi, diperoleh dari kuesioner pengguna QRIS Tap pada transportasi umum melalui aplikasi e-wallet Bayarind, dengan setiap responden direpresentasikan sebagai satu objek data.

Variabel	Kategori	Kode
Usia (X1)	< 20	1
	20–25	2
	26–35	3
	36–45	4
	> 45	5

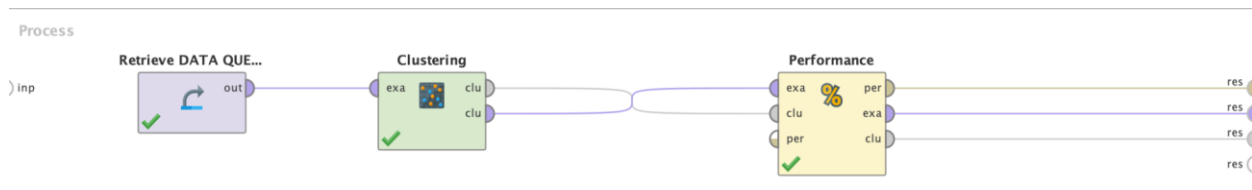
Gambar 2. Contoh Skema Encoding Variabel Usia

Dataset

Dataset terdiri dari 63 responden, dengan masing-masing responden memiliki empat variabel penelitian, yaitu usia (X1), lama penggunaan aplikasi e-wallet (X2), frekuensi penggunaan QRIS Tap (X3), dan nilai transaksi (X4). Seluruh variabel tersebut telah melalui proses *encoding* sehingga direpresentasikan dalam bentuk numerik sesuai dengan skema yang telah dijelaskan pada subbab sebelumnya. Tabel data responden disajikan dalam dua bagian untuk memudahkan penyajian dan keterbacaan, mengingat jumlah data yang cukup besar. Setiap baris data menunjukkan satu responden yang diberi kode R1 hingga R63, sedangkan setiap kolom menunjukkan nilai variabel X1 hingga X4 yang digunakan dalam proses klasterisasi. Dataset hasil *encoding* ini selanjutnya digunakan sebagai input pada proses klasterisasi menggunakan algoritma K-Means. Penggunaan data numerik yang konsisten dan terstruktur memungkinkan proses pengelompokan dilakukan secara objektif dan sesuai dengan asumsi metode clustering yang digunakan dalam penelitian ini.

Implementasi Klasterisasi Menggunakan RapidMiner

Implementasi proses klasterisasi dilakukan dengan menggunakan perangkat lunak RapidMiner sebagai alat bantu analisis data. RapidMiner digunakan untuk memastikan proses pengolahan data berjalan secara konsisten, efisien, dan dapat direplikasi. Tahapan implementasi dimulai dengan pemanggilan dataset yang telah melalui proses pra-pemrosesan. Dataset kemudian diproses menggunakan operator K-Means untuk membentuk klaster berdasarkan kemiripan atribut numerik. Selanjutnya, hasil klasterisasi dievaluasi menggunakan operator Performance (Clustering) untuk memperoleh nilai Davies–Bouldin Index (DBI) dan Within-Cluster Sum of Squares (WCSS). Nilai evaluasi tersebut digunakan sebagai dasar dalam menentukan jumlah klaster yang optimal serta menilai kualitas segmentasi yang dihasilkan.

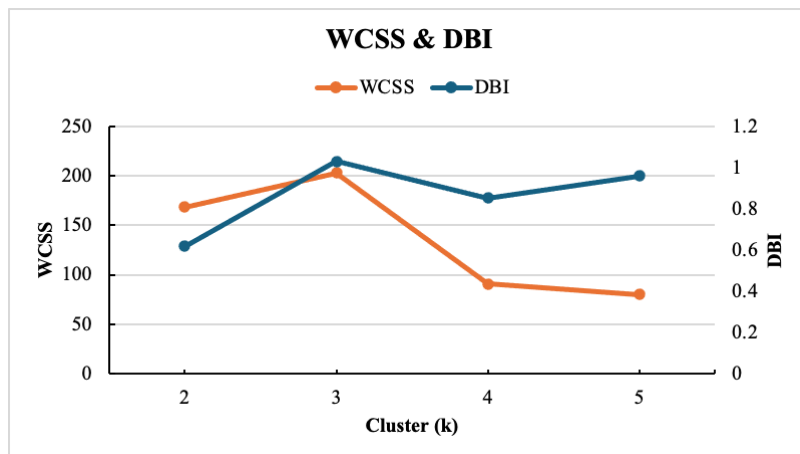


Gambar 3. Diagram proses klusterisasi K-Means menggunakan RapidMiner yang terdiri dari tahapan pengambilan data (*Retrieve*), proses klusterisasi (*K-Means*), dan evaluasi performa kluster menggunakan *Performance (Clustering)*.

Gambar 2 menunjukkan alur proses klusterisasi menggunakan perangkat lunak RapidMiner. Tahap pertama adalah Retrieve Dataset, yaitu proses pengambilan data penelitian dari repositori RapidMiner yang telah melalui tahap pra-pemrosesan dan siap digunakan sebagai input analisis. Data yang diambil berisi variabel numerik hasil encoding yang merepresentasikan karakteristik pengguna QRIS Tap. Tahap selanjutnya adalah Clustering, di mana algoritma K-Means diterapkan untuk mengelompokkan data berdasarkan tingkat kemiripan menggunakan jarak Euclidean, sehingga setiap data dialokasikan ke dalam kluster dengan centroid terdekat. Tahap terakhir adalah Performance (Clustering), yang berfungsi untuk mengevaluasi kualitas hasil klusterisasi dengan mengukur tingkat kekompakan intra-kluster dan pemisahan antar-kluster. Hasil evaluasi ini digunakan sebagai dasar dalam menentukan konfigurasi kluster yang paling optimal serta mendukung interpretasi hasil segmentasi pengguna.

Evaluasi Kluster Menggunakan Davies–Bouldin Index (DBI) dan WCSS

Evaluasi kualitas kluster pada penelitian ini dilakukan menggunakan Davies–Bouldin Index (DBI) dan Within-Cluster Sum of Squares (WCSS). Davies–Bouldin Index digunakan untuk mengukur rasio antara jarak intra-kluster dan jarak antar-kluster, di mana nilai DBI yang lebih kecil menunjukkan hasil klusterisasi yang lebih optimal karena mencerminkan kluster yang kompak dan terpisah dengan baik [8]. Sementara itu, WCSS digunakan untuk melihat seberapa besar variasi data di dalam setiap kluster. Nilai WCSS yang rendah menunjukkan bahwa anggota kluster memiliki jarak yang dekat terhadap centroid-nya, sehingga menghasilkan kluster yang lebih homogen. Dalam praktik evaluasi K-Means, tren WCSS yang terus menurun seiring bertambahnya jumlah kluster akan mencapai titik pelandaian (*elbow*), yang mencerminkan jumlah kluster optimal secara keseimbangan antara kompleksitas model dan kualitas segmentasi. Pada penelitian ini dilakukan pengujian terhadap beberapa konfigurasi jumlah kluster. Hasil DBI menunjukkan bahwa meskipun nilai DBI terkecil diperoleh pada $K = 2$, konfigurasi tersebut menghasilkan segmentasi yang terlalu luas dan kurang informatif untuk interpretasi analisis. Sebaliknya, konfigurasi $K = 4$ menunjukkan pemisahan kluster yang lebih seimbang dengan nilai DBI yang tetap berada dalam rentang rendah, serta mendukung interpretasi perilaku pengguna secara lebih tajam. Dengan demikian, konfigurasi $K = 4$ dipilih sebagai model akhir dalam penelitian ini dan digunakan sebagai dasar interpretasi hasil klusterisasi pada tahap selanjutnya (Ananda & Sulisty, 2025) (Jain, 2010).



Gambar 4 Grafik WCSS & DBI

Jumlah Kluster (K)	DBI	WCSS
2	0.618688	168.6715
3	1.030673	203.1809
4	0.853912	90.59016
5	0.959928	80.17204

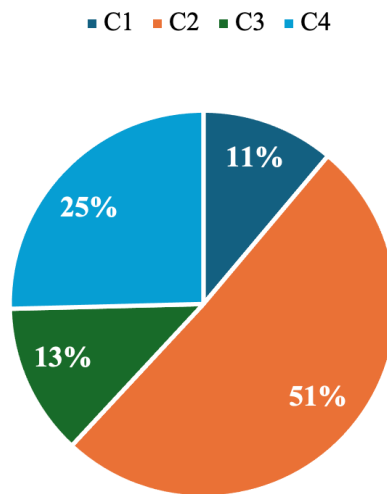
Gambar 5. Hasil Nilai DBI & WCSS

Grafik *WCSS & DBI* pada gambar menunjukkan dinamika kualitas kluster terhadap perubahan jumlah kluster ($k = 2$ sampai $k = 5$) dengan menggunakan dua metrik evaluasi yaitu *Within-Cluster Sum of Squares (WCSS)* dan *Davies–Bouldin Index (DBI)*. Secara umum, kedua metrik ini digunakan untuk mengevaluasi keseimbangan antara homogenitas intra-kluster dan separasi antar-kluster.

Dari sisi *WCSS*, terlihat bahwa nilai *WCSS* cenderung menurun seiring bertambahnya jumlah kluster. Pada $k = 2$, *WCSS* berada pada kisaran tinggi (± 165), kemudian meningkat pada $k = 3$ (± 200), sebelum menurun secara signifikan pada $k = 4$ (± 90) dan mendekati titik pelandaian pada $k = 5$ (± 80). Penurunan tajam dari $k = 3$ menuju $k = 4$ mengindikasikan peningkatan homogenitas kluster, sementara penurunan yang semakin kecil dari $k = 4$ ke $k = 5$ menunjukkan bahwa penambahan kluster setelah titik tersebut mulai menghasilkan *diminishing returns* dan tidak memberikan perbaikan kualitas yang substansial. Dengan demikian, $k = 4$ muncul sebagai kandidat titik siku (*elbow*) pada metode *Elbow*, yang menunjukkan jumlah kluster paling efisien tanpa meningkatkan kompleksitas berlebihan.

Distribusi Anggota Klaster

Distribusi setiap Cluster



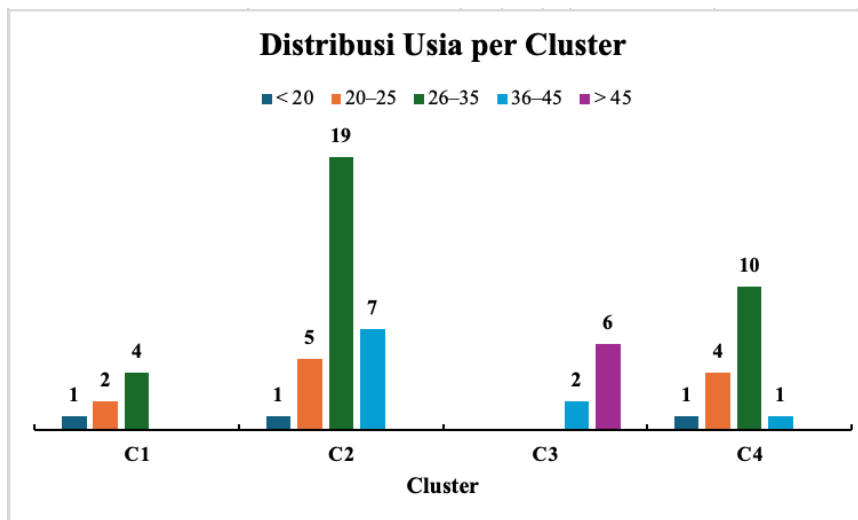
Gambar 6. *Distribusi Anggota Klaster Pengguna Qris Tap (K=4)*

Diagram menunjukkan bahwa Klaster 2 (C2) memiliki proporsi terbesar sebesar 51%, diikuti oleh Klaster 4 (C4) sebesar 25%, Klaster 3 (C3) sebesar 13%, dan Klaster 1 (C1) sebesar 11%. Hal ini menunjukkan bahwa sebagian besar data terkonsentrasi pada C2, sementara klaster lainnya memiliki jumlah anggota yang lebih kecil.

Distribusi Usia Per Klaster

Klaster	< 3 bulan	3–6 bulan	7–12 bulan	> 1 tahun	Total
C1	–	3	4	–	7
C2	–	–	2	30	32
C3	–	–	–	8	8
C4	–	–	–	16	16
Total	0	3	4	54	63

Gambar 7. *Distribusi Usia Responden pada Setiap Klaster*



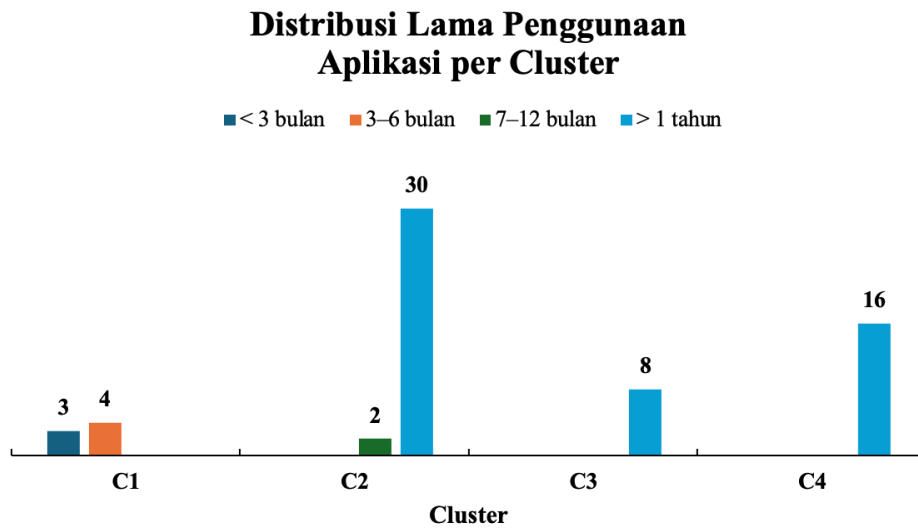
Gambar 8. Distribusi Usia per Klaster

Grafik menunjukkan perbedaan distribusi usia pada setiap klaster. Klaster 1 (C1), Klaster 2 (C2), dan Klaster 4 (C4) didominasi oleh kelompok usia 26–35 tahun. Klaster 3 (C3) memiliki pola berbeda dengan dominasi usia di atas 45 tahun. Secara umum, kelompok usia 26–35 tahun mendominasi sebagian besar klaster, menunjukkan bahwa mayoritas pengguna berada pada usia produktif.

Distribusi Lama Penggunaan Aplikasi

Klaster	< 3 bulan	3–6 bulan	7–12 bulan	> 1 tahun	Total
C1	0	3	4	0	7
C2	0	0	2	30	32
C3	0	0	0	8	8
C4	0	0	0	16	16
Total	0	3	6	54	63

Gambar 9. Distribusi Lama Penggunaan Aplikasi pada Setiap Klaster



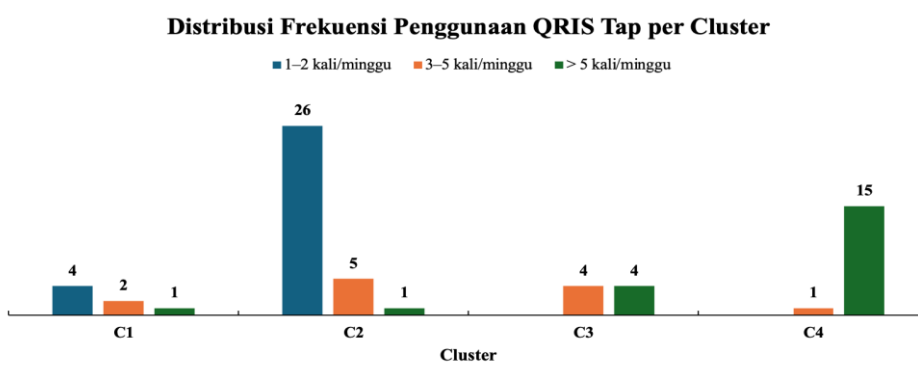
Gambar 10. Distribusi Lama Penggunaan Aplikasi per Klaster

Grafik menunjukkan perbedaan lama penggunaan aplikasi pada setiap klaster. Klaster 1 (C1) didominasi pengguna baru dengan masa penggunaan kurang dari 6 bulan. Klaster 2 (C2), Klaster 3 (C3), dan Klaster 4 (C4) mayoritas terdiri dari pengguna yang telah menggunakan aplikasi lebih dari 1 tahun. Hal ini menunjukkan tingkat loyalitas pengguna yang tinggi, dengan C1 merepresentasikan segmen pengguna baru.

Distribusi Frekuensi Penggunaan Qris Tap per Cluster

Klaster	1-2 kali/minggu	3-5 kali/minggu	> 5 kali/minggu	Total
C1	4	2	1	7
C2	26	5	1	32
C3	0	4	4	8
C4	0	1	15	16
Total	30	12	21	63

Gambar 11 Distribusi frekuensi Penggunaan Qris Tap pada Setiap Klaster



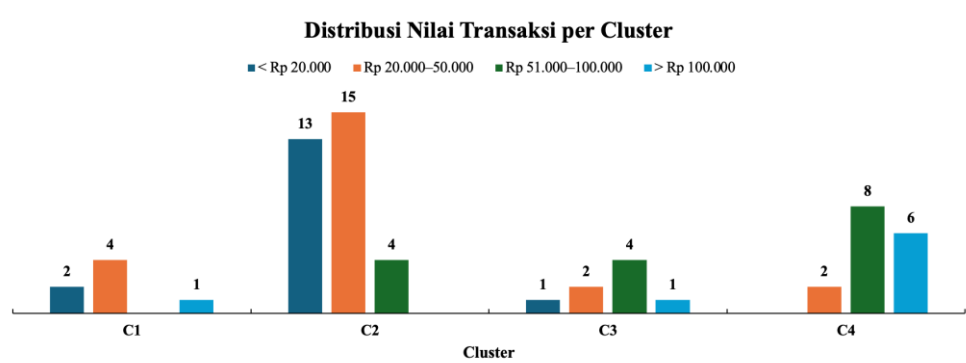
Gambar 12. Distribusi Frekuensi Penggunaan Qris Tap per Klaster

Grafik menunjukkan perbedaan frekuensi penggunaan QRIS Tap pada setiap klaster. Klaster 1 (C1) dan Klaster 2 (C2) didominasi oleh pengguna dengan frekuensi rendah (1–2 kali per minggu). Klaster 3 (C3) menunjukkan pola yang lebih seimbang antara frekuensi sedang dan tinggi. Klaster 4 (C4) didominasi oleh pengguna dengan frekuensi tinggi (>5 kali per minggu), yang mencerminkan tingkat engagement tertinggi..

Distribusi Nilai Transaksi per Klaster

Klaster	< Rp 20.000	Rp 20.000–50.000	Rp 51.000–100.000	> Rp 100.000	Total
C1	2	4	0	1	7
C2	13	15	4	0	32
C3	1	2	4	1	8
C4	0	2	8	6	16
Total	16	23	16	8	63

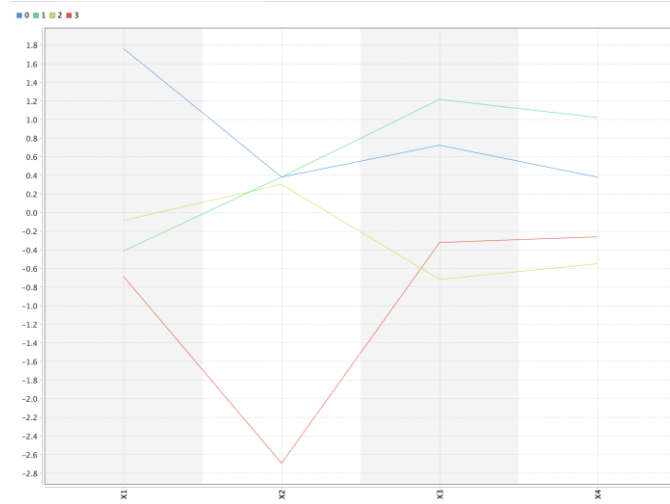
Gambar 13. Distribusi Nilai Transaksi per Klaster



Gambar 14. Distribusi Nilai Transaksi per Klaster

Grafik menunjukkan perbedaan nilai transaksi pada setiap klaster. Klaster 1 (C1) dan Klaster 2 (C2) didominasi oleh transaksi bernilai rendah hingga menengah. Klaster 3 (C3) memiliki distribusi nilai transaksi yang lebih merata. Klaster 4 (C4) didominasi oleh transaksi bernilai tinggi, menunjukkan segmen pengguna dengan daya beli tertinggi Mumtaha & Khoiri, (2019).

Interpretasi Karakteristik Setiap Kluster



Gambar 15. Diagram Line Plot K-Means 4 Cluster

Gambar 9 menunjukkan perbedaan pola centroid pada masing-masing kluster berdasarkan variabel X1–X4. Nilai centroid yang digunakan merupakan nilai rata-rata terstandarisasi, sehingga mencerminkan kecenderungan relatif setiap variabel dalam kluster. Perbedaan pola tersebut menunjukkan bahwa setiap kluster memiliki karakteristik pengguna yang unik. Berdasarkan hasil klusterisasi K-Means dengan jumlah kluster optimal sebanyak empat ($K = 4$), serta didukung oleh analisis distribusi usia, lama penggunaan aplikasi, frekuensi penggunaan, dan nilai transaksi, karakteristik masing-masing kluster diinterpretasikan sebagai berikut.

1. Kluster 1 (C1) Pengguna Baru dengan Aktivitas Transaksi Rendah

Kluster 1 (C1) merupakan kluster dengan jumlah anggota paling sedikit. Kluster ini didominasi oleh pengguna berusia 26–35 tahun dengan lama penggunaan aplikasi kurang dari 6 bulan. Dari sisi perilaku transaksi, anggota C1 cenderung memiliki frekuensi dan nilai transaksi yang rendah hingga menengah. Karakteristik tersebut menunjukkan bahwa C1 merepresentasikan segmen pengguna baru yang masih berada pada tahap awal adopsi dan eksplorasi fitur aplikasi.

2. Kluster 2 (C2) Pengguna Produktif dengan Loyalitas Tinggi dan Transaksi Menengah

Kluster 2 (C2) merupakan kluster dengan jumlah anggota terbesar. Mayoritas anggotanya berada pada rentang usia produktif, khususnya 26–35 tahun, serta didominasi oleh pengguna dengan lama penggunaan aplikasi lebih dari 1 tahun. Meskipun demikian, frekuensi penggunaan dan nilai transaksi pada kluster ini cenderung rendah hingga menengah. Karakteristik ini menunjukkan bahwa pengguna C2 telah memiliki tingkat loyalitas yang tinggi terhadap aplikasi, namun belum memaksimalkan intensitas dan nilai transaksi secara optimal.

3. Klaster 3 (C3) Pengguna Berpengalaman dengan Pola Transaksi Beragam

Klaster 3 (C3) didominasi oleh pengguna berusia di atas 45 tahun dengan lama penggunaan aplikasi lebih dari 1 tahun. Frekuensi penggunaan dan nilai transaksi pada klaster ini menunjukkan distribusi yang relatif seimbang antara kategori menengah dan tinggi. Karakteristik ini mencerminkan pengguna yang telah berpengalaman dan memiliki pola transaksi yang stabil.

4. Klaster 4 (C4) Pengguna Premium dengan Daya Beli Tinggi

Klaster 4 (C4) didominasi oleh pengguna berusia 26–35 tahun dengan lama penggunaan aplikasi lebih dari 1 tahun. Klaster ini memiliki frekuensi penggunaan tertinggi, dengan mayoritas anggota melakukan transaksi lebih dari 5 kali per minggu. Selain itu, nilai transaksi pada klaster ini cenderung tinggi, menunjukkan bahwa C4 merepresentasikan segmen pengguna dengan tingkat engagement dan daya beli tertinggi. Secara keseluruhan, hasil interpretasi menunjukkan bahwa klasterisasi berhasil memisahkan pengguna ke dalam segmen yang berbeda berdasarkan karakteristik demografis dan perilaku penggunaan, yaitu:

- 1) C1: Pengguna baru dengan aktivitas transaksi rendah
- 2) C2: Pengguna inti yang loyal dengan transaksi rutin bernilai menengah
- 3) C3: Pengguna dewasa/senior berpengalaman dengan pola transaksi beragam
- 4) C4: Pengguna premium dengan daya beli dan nilai transaksi tinggi

Segmentasi ini memberikan gambaran yang jelas mengenai perbedaan perilaku pengguna pada setiap klaster dan dapat menjadi dasar bagi pengambilan keputusan strategis maupun pengembangan layanan yang lebih terarah. Temuan penelitian ini sejalan dengan studi terkini yang menyatakan bahwa frekuensi penggunaan dan nilai transaksi merupakan indikator utama dalam segmentasi pengguna pembayaran berbasis QR dan e-wallet. Penelitian oleh Sari et al, (2025) menunjukkan bahwa pengguna dengan intensitas transaksi tinggi cenderung membentuk klaster tersendiri dengan karakteristik loyalitas yang kuat terhadap layanan digital Santoso & Wicaksono, (2024).

Simpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa segmentasi pengguna QRIS Tap pada moda transportasi umum di wilayah Jabodetabek melalui aplikasi e-wallet Bayarind berhasil dilakukan menggunakan algoritma K-Means dengan empat variabel faktual, yaitu usia, lama penggunaan aplikasi, frekuensi transaksi, dan nilai transaksi. Evaluasi menggunakan Davies–Bouldin Index dan Within-Cluster Sum of Squares menunjukkan bahwa konfigurasi empat klaster ($K = 4$) merupakan model yang paling optimal dan stabil. Klasterisasi ini menghasilkan tipologi perilaku pengguna yang jelas, mulai dari pengguna pemula dengan aktivitas rendah hingga pengguna aktif dengan frekuensi dan nilai transaksi tinggi, sehingga menegaskan efektivitas metode K-Means dalam memetakan karakteristik pengguna QRIS Tap pada layanan transportasi umum. Hasil penelitian menegaskan bahwa metode K Means efektif digunakan untuk segmentasi pengguna pembayaran digital berbasis data faktual, sejalan dengan temuan penelitian

sebelumnya di bidang digital payment dan financial technology. Namun, penelitian ini masih memiliki keterbatasan dari sisi jumlah responden dan cakupan wilayah, sehingga penelitian selanjutnya disarankan untuk menggunakan data yang lebih luas serta membandingkan metode clustering lain seperti DBSCAN atau hierarchical clustering.

Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan untuk menggunakan jumlah dan cakupan data yang lebih luas, mencakup wilayah serta karakteristik pengguna yang beragam. Penelitian mendatang juga dapat mengombinasikan data perilaku faktual dengan variabel persepsi pengguna guna memperoleh pemahaman yang lebih komprehensif terhadap pola adopsi QRIS Tap. Selain itu, penerapan dan perbandingan metode clustering lain, seperti hierarchical clustering atau density-based clustering, dapat dilakukan untuk mengevaluasi konsistensi dan kualitas segmentasi yang dihasilkan.

Daftar Pustaka

- Akram, A. et al. (2024). Penerapan algoritma K-Means dalam pengelompokan prestasi belajar siswa SMK Negeri 1 Gowa. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(2), 145–152.
- Ananda, T., & Sulistyono, S. (2025). Machine learning in financial user segmentation: A systematic review. *Journal of AI in Finance*, 7(1), 1–15.
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Han, J. et al. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Hidayat, R., & Kusniyanti, H. (2022). Penerapan algoritma K-Means untuk pengelompokan data penjualan pada Cafe 47°Coffee. *Jurnal Informatika Dan Sistem Informasi*, 6(3), 210–218.
- Indonesia, B. (2020). *QRIS: Quick Response Code Indonesian Standard*. Bank Indonesia.
- Indonesia, B. (2023). *Pengembangan QRIS Tap berbasis Near Field Communication (NFC)*. Bank Indonesia.
- Irawati, F. et al. (2025). Pemetaan kinerja ekspor kopi Indonesia berdasarkan negara tujuan menggunakan metode K-Means clustering. *Jurnal Ekonomi Dan Bisnis Digital*, 4(1), 55–66.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-Means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Kim, C., & Mirusmonov, M. (2020). Consumers' adoption of contactless payment: NFC and mobile wallet. *Journal of Retailing and Consumer Services*, 57, 102216.
- Kumar, A., & Singh, R. (2021). Cluster validation techniques in data mining: A comparative study. *Journal of Big Data Analytics*, 6(2), 45–58.
- Li, Y., & Huang, W. (2022). Cashless payments and consumer transaction patterns. *Electronic Commerce Research*, 22(4), 1011–1030.

- Mumtaha, H. A., & Khoiri, H. A. (2019). Analisis Dampak Perkembangan Revolusi Industri 4.0 dan Society 5.0 Pada Perilaku Masyarakat Ekonomi (E-Commerce). *JURNAL PILAR TEKNOLOGI: Jurnal Ilmiah Ilmu Ilmu Teknik*, 4(2), 55–60. <https://doi.org/10.33319/piltek.v4i2.39>
- Permatasari, R., & Cahyono, E. (2022). Analisis penggunaan QRIS pada UMKM di Jawa Tengah. *Jurnal Sistem Informasi*, 18(2), 85–96.
- Prasetyo, E. et al. (2024). Evaluasi kualitas kluster menggunakan Davies–Bouldin Index pada data transaksi digital. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 33–42.
- Rahman, A., & Putri, D. A. (2022). Analisis segmentasi pengguna e-wallet menggunakan metode K-Means clustering. *Jurnal Sistem Informasi*, 18(3), 210–221.
- RapidMiner. (2023). *RapidMiner Studio documentation*. <https://rapidminer.com>
- Santoso, P., & Wicaksono, A. (2024). User behaviour on QRIS transactions in micro payments. *Journal of Indonesian Payment Systems*, 6(1), 44–58.
- Sari, R. et al. (2025). Clustering analysis of QR-based payment users in public transportation. *Journal of Digital Business and Economics*, 4(1), 55–68.
- Setiyawan, B., & Rizki, A. (2023). Segmentation of digital payment users using K-Means. *International Journal of Data Mining*, 9(2), 77–89.
- Valencia-Arias, A., & others. (2025). Research trends in mobile payment adoption: A systematic review. *Technology in Society*, 72, 102421.
- Wang, Y. et al. (2021). User behavior clustering in digital payment systems using K-Means algorithm. *IEEE Access*, 9, 118234–118245.
- Wijaya, F., & Hartono, D. (2024). Cluster analysis of payment behavior among e-wallet users. *Journal of Digital Economy*, 10(3), 201–215.
- Yusuf, M. D., & Munandar, T. A. (2025). Analisis kluster wilayah berdasarkan indikator kesejahteraan publik menggunakan metode K-Means. *Jurnal Sistem Informasi Dan Data Mining*, 9(1), 22–31.
- Zhang, L. et al. (2023). Unsupervised learning for customer segmentation in fintech services. *Expert Systems with Applications*, 213, 118902.