Contents lists available at JESA

Journal of Engineering and Science Application

journal homepage: https://jesa.aks.or.id/index.php/jesa/index

Implementasi K-Means Clustering Dengan Pendekatan RFM Untuk Pengelompokan Pola Pembayaran Iuran Bulanan di Komplek Kepu

Muhamad Rafi Akmalludin[a*], Niki Ratama[b]

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

Article Info

Keywords:

K-Means Clustering
RFM
Pola Pembayaran
Iuran Bulanan
Data Mining

ABSTRACT

Pengelolaan iuran bulanan di lingkungan perumahan menghadapi dua permasalahan: penerimaan iuran yang belum optimal dan ketiadaan dasar terukur bagi pengurus untuk membedakan warga berdasarkan pola pembayarannya. Di Komplek Kepu, iuran yang tercatat terbayar pada tahun 2025 baru mencapai 63,2% dari potensi penerimaan. Penelitian ini bertujuan mengidentifikasi pola pembayaran iuran 214 warga periode Januari–Desember 2025 berdasarkan atribut RFM (Recency, Frequency, Monetary) menggunakan K-Means clustering, menganalisis karakteristik tiap kelompok, dan menyusun usulan pendekatan penagihan. Metode yang digunakan adalah kuantitatif dengan teknik sensus dan pendekatan data mining, meliputi pembersihan data, transformasi RFM, normalisasi Min-Max, penentuan K melalui Metode Elbow, serta evaluasi dengan Silhouette Coefficient, Davies-Bouldin Index, dan Calinski-Harabasz Index, dengan uji Kruskal-Wallis, Eta Squared, dan Mann-Whitney U sebagai konfirmasi internal. Hasil menunjukkan terbentuk tiga cluster (K=3; WCSS=8,2868): Warga Disiplin (40 warga/18,69%), Warga Tidak Konsisten (89 warga/41,59%), dan Warga Berisiko Tinggi (85 warga/39,72%), dengan Silhouette=0,5398, DBI=0,7130, dan CHI=314,03. Ketiga cluster berbeda secara statistik pada atribut RFM ($p<0,0001$), terutama pada Frequency dan Monetary ($\eta^2=0,8808$). Hasil ini dapat menjadi dasar empiris untuk menentukan prioritas penagihan. Rekomendasi merupakan usulan berbasis data historis yang belum diuji di lapangan, dan penelitian ini tidak menganalisis penyebab ketidakteraturan pembayaran.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Muhamad Rafi Akmalludin
Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Pamulang
Tangerang Selatan, Indonesia
Email: rafiakmal587@gmail.com

1. LATAR BELAKANG

Pengelolaan keuangan di tingkat komunitas perumahan merupakan aspek kritis yang menentukan keberlanjutan fasilitas dan layanan bersama. Iuran bulanan warga berfungsi sebagai sumber pendapatan utama untuk operasional kebersihan, keamanan, perawatan infrastruktur, dan kegiatan sosial kemasyarakatan. Data [1] mencatat pengeluaran per kapita untuk perumahan dan fasilitas rumah tangga mencapai Rp391.751

per bulan atau 26,11% dari total pengeluaran bukan makanan nasional, mencerminkan besarnya signifikansi ekonomi sektor perumahan di Indonesia. Namun, berbeda dengan pengelolaan keuangan organisasi formal yang sudah didukung sistem informasi terintegrasi, pengelolaan iuran di komunitas perumahan masih sangat bergantung pada kapasitas dan inisiatif pengurus yang bersifat sukarela dan non-profesional.

Kondisi ini menimbulkan berbagai permasalahan operasional yang terdokumentasi secara konsisten dalam literatur. [2] menemukan bahwa pencatatan manual pada buku besar kertas menyebabkan keterlambatan rekonsiliasi dan rentan terhadap kehilangan data. [3]) mendokumentasikan bahwa sistem penagihan door-to-door tanpa penjadwalan sistematis menghasilkan inkonsistensi dalam frekuensi dan waktu penagihan. [4] mengidentifikasi bahwa ketiadaan laporan keuangan periodik menyulitkan perencanaan anggaran yang akurat. Ketiga studi tersebut menunjukkan bahwa pengelolaan iuran di tingkat komunitas masih menghadapi kendala pencatatan dan penagihan yang belum terstruktur.

Permasalahan tersebut tampak pada pengelolaan iuran di Komplek Kepu. Dari 214 warga yang wajib membayar iuran Rp50.000 per bulan, potensi penerimaan selama Januari–Desember 2025 adalah Rp128.400.000, sedangkan iuran yang tercatat terbayar baru mencapai 63,2% dari potensi tersebut, sehingga berisiko memengaruhi ketersediaan kas operasional. Berdasarkan pengamatan langsung penulis, penagihan dilakukan oleh pengurus kepada seluruh warga pada hari yang sama, sementara catatan pembayaran hanya berfungsi sebagai arsip administratif dan belum diolah menjadi informasi mengenai pola pembayaran warga.

Pendekatan data mining menawarkan solusi yang relevan untuk mengatasi keterbatasan tersebut. K-Means clustering, sebagai salah satu algoritma unsupervised learning yang paling banyak digunakan, memungkinkan pengelompokan objek data ke dalam cluster-cluster berdasarkan kemiripan atribut tanpa memerlukan label kelas sebelumnya [5] ;[6] . Integrasi K-Means dengan model RFM (Recency, Frequency, Monetary) yang secara tradisional digunakan dalam pemasaran untuk segmentasi pelanggan memungkinkan representasi perilaku pembayaran dalam tiga dimensi yang saling melengkapi. Kombinasi keduanya telah terbukti menghasilkan segmentasi yang interpretatif dan actionable dalam berbagai konteks pembayaran berbasis kewajiban.

Sejumlah penelitian terdahulu telah menerapkan K-Means dalam konteks analisis perilaku pembayaran dengan hasil yang menjanjikan. [7] berhasil mengelompokkan keterlambatan pembayaran SPP di SMK menggunakan K-Means dan menghasilkan dua klaster yang memudahkan sekolah merancang sistem penagihan berbasis risiko. [8] mengaplikasikan K-Means untuk memetakan kontribusi metode pembayaran digital ke dalam empat kategori berdasarkan nilai dan frekuensi transaksi pada payment aggregator. [9] menggunakan K-Means untuk mengelompokkan nasabah kredit berdasarkan risiko pembayaran pada lembaga keuangan, menghasilkan tiga klaster yang menjadi dasar manajemen risiko portofolio. [10] mengintegrasikan model RFM dan K-Means untuk segmentasi donor organisasi non-profit, menunjukkan efektivitas adaptasi RFM di luar konteks pemasaran komersial. [11] menerapkan K-Means untuk klasifikasi pengunjung mal berdasarkan perilaku pembelian, sementara [12] menggunakannya untuk memprediksi kelulusan mata kuliah mahasiswa dengan mengidentifikasi pola partisipasi akademik.

Meskipun demikian, terdapat kesenjangan penelitian yang signifikan: belum ada studi yang secara spesifik menerapkan K-Means berbasis RFM dalam konteks iuran bulanan lingkungan perumahan yang bersifat wajib, kolektif, dan dikelola secara komunal di Indonesia. Konteks ini memiliki karakteristik unik yang membedakannya dari domain-domain yang telah diteliti: (1) kewajiban bersifat komunal dengan mekanisme sanksi non-formal berbasis tekanan sosial; (2) ketergantungan langsung pada kesadaran kolektif warga tanpa mekanisme enforcement yang kuat; (3) pengelolaan oleh pengurus non-profesional dengan sumber daya terbatas; serta (4) dampak langsung terhadap kualitas lingkungan tempat tinggal seluruh warga. Karakteristik unik ini menjadikan konteks iuran perumahan sebagai domain yang layak dan penting untuk dieksplorasi secara akademis. Berdasarkan kesenjangan tersebut, penelitian ini bertujuan: (1) mengidentifikasi pola pembayaran iuran bulanan 214 warga Komplek Kepu periode Januari–Desember 2025 berdasarkan atribut RFM dan mengelompokkan warga menggunakan K-Means clustering; (2) menganalisis karakteristik setiap kelompok; dan (3) menyusun usulan pendekatan penagihan untuk setiap kelompok berdasarkan data historis pembayaran.

2. METODE PENELITIAN

2.1. Jenis Penelitian dan Sumber Data

Penelitian ini menggunakan metode kuantitatif dengan pendekatan studi kasus eksploratif berbasis data mining. Data primer berupa catatan historis transaksi pembayaran iuran bulanan seluruh warga Komplek Kepu periode Januari–Desember 2025 (12 bulan), diperoleh dari dokumentasi administrasi pengurus komplek. Iuran bersifat tetap sebesar Rp50.000 per kepala keluarga per bulan. Data mencakup informasi identitas kepala keluarga (yang kemudian dianonimkan), bulan iuran, tanggal pembayaran, dan status pembayaran (lunas/belum). Populasi penelitian berjumlah 214 kepala keluarga yang tercatat sebagai wajib iuran selama periode pengamatan, dan seluruhnya dianalisis tanpa pengecualian menggunakan teknik sensus.

Keputusan sensus didasarkan pada: (1) jumlah populasi yang terjangkau secara komputasi; (2) kebutuhan gambaran lengkap kondisi riil komplek; dan (3) eliminasi bias sampling untuk mendukung keakuratan rekomendasi manajerial [13].

2.2. Definisi Operasional Variabel RFM

Perilaku pembayaran warga direpresentasikan melalui tiga atribut yang diadaptasi dari model RFM yang dipopulerkan [14] dan dioperasionalisasikan lebih lanjut oleh [15] : (1) Recency (R) selisih bulan antara pembayaran terakhir warga dengan akhir periode pengamatan (Desember 2025); nilai lebih kecil berarti pembayaran lebih baru dilakukan; (2) Frequency (F) jumlah total bulan di mana warga melakukan pembayaran iuran selama 12 bulan periode pengamatan, dengan nilai minimum 3 dan maksimum 12; dan (3) Monetary (M) total nominal iuran yang berhasil dibayarkan selama 12 bulan, berkisar antara Rp150.000 hingga Rp600.000. Karena iuran bersifat tetap, Monetary merupakan fungsi linear dari Frequency (M = F × Rp50.000). Uji sensitivitas yang dilakukan menjalankan K-Means hanya pada dimensi Recency dan Frequency menghasilkan clustering identik 100%, mengkonfirmasi bahwa redundansi Monetary tidak mengubah hasil pengelompokan. Ketiga atribut dipertahankan untuk konsistensi dengan kerangka RFM standar dalam literatur.

2.3. Landasan Teori Algoritma K-Means

K-Means merupakan algoritma clustering berbasis partisi yang pertama kali diperkenalkan [16] . Prinsip dasarnya adalah mempartisi dataset berisi n objek ke dalam K cluster sedemikian rupa sehingga total jarak kuadrat objek ke centroid cluster-nya (WCSS) diminimalkan. Secara formal, algoritma ini memecahkan permasalahan optimasi: $WCSS = \Sigma_{i=1}^{K} \Sigma_{x \in C_i} ||x - \mu_i||^2$, di mana C_i adalah himpunan objek dalam cluster ke-i dan μ_i adalah centroid cluster ke-i [17].

Jarak Euclidean digunakan sebagai fungsi jarak antar titik data dan centroid: $d(x, y) = \sqrt{[\Sigma_{j=1}^{n} (x_j - y_j)^2]}$, di mana x dan y adalah dua titik data dalam ruang n-dimensi. Pemilihan jarak Euclidean didasarkan pada kesesuaiannya untuk data numerik kontinu yang telah dinormalisasi, sebagaimana direkomendasikan [6] untuk aplikasi clustering pada data perilaku.

Inisialisasi K-Means++ digunakan untuk mempercepat konvergensi dan meningkatkan kualitas clustering. Dibandingkan inisialisasi acak standar, K-Means++ memilih centroid awal yang saling berjauhan secara proporsional terhadap kuadrat jarak, sehingga mengurangi risiko konvergensi pada solusi lokal suboptimal [18]. Parameter yang digunakan: n_clusters=3, init='k-means++', max_iter=100, n_init=10, random_state=42.

Metode Elbow digunakan untuk menentukan jumlah cluster optimal. Metode ini memplot nilai WCSS untuk berbagai nilai K dan mengidentifikasi titik 'siku' di mana penurunan WCSS mulai melandai secara signifikan. Titik siku menandakan nilai K yang memberikan trade-off optimal antara kompleksitas model dan kualitas clustering.

2.4. Tahapan Analisis

Pra-pemrosesan: (a) Pembersihan data pengecekan missing values, duplikasi, dan inkonsistensi format; (b) Transformasi RFM agregasi data longitudinal menjadi satu rekod per warga; (c) Normalisasi Min-Max: Xnorm = (X − Xmin) / (Xmax − Xmin), memastikan setiap variabel berbobot seimbang dalam perhitungan jarak Euclidean.

Evaluasi Clustering: Tiga metrik komplementer digunakan. (1) Silhouette Coefficient: mengukur seberapa mirip objek dengan cluster-nya sendiri dibanding cluster terdekat lainnya, rentang −1 hingga +1 dan semakin mendekati 1 semakin baik [19]. (2) Davies-Bouldin Index (DBI): mengukur rasio jarak intra-cluster terhadap jarak antar-cluster; semakin kecil semakin baik, dengan nilai ideal di bawah 1,0 [20] . (3) Calinski-Harabasz Index (CHI): mengukur rasio dispersi antar-cluster terhadap dispersi dalam-cluster; semakin besar nilainya semakin baik [21].

Validasi Inferensial: Sebagai alat konfirmasi internal, dilakukan uji statistik bertahap: uji normalitas Shapiro-Wilk → jika tidak normal, gunakan Kruskal-Wallis sebagai alternatif non-parametrik ANOVA → Eta Squared (η^2) untuk mengukur effect size → Mann-Whitney U sebagai uji post-hoc antar pasangan cluster. Implementasi menggunakan Python 3.11.5, library Scipy 1.11.1, Scikit-learn 1.3.0, Pandas 2.0.3, Matplotlib 3.7.2, dan Seaborn 0.12.2 di lingkungan Jupyter Notebook 6.5.4.

3. HASIL DAN PEMBAHASAN

3.1. Gambaran Dataset dan Pra-Pemrosesan

Dataset terdiri dari catatan pembayaran 214 kepala keluarga selama 12 bulan. Setelah proses pembersihan data, tidak ditemukan missing values maupun duplikasi pada seluruh variabel. Hasil transformasi RFM menghasilkan tiga variabel turunan per warga: Recency berkisar 0–5 bulan (rata-rata 0,248; std 0,810), Frequency 3–12 bulan (rata-rata 7,584; std 2,281), dan Monetary Rp150.000–Rp600.000

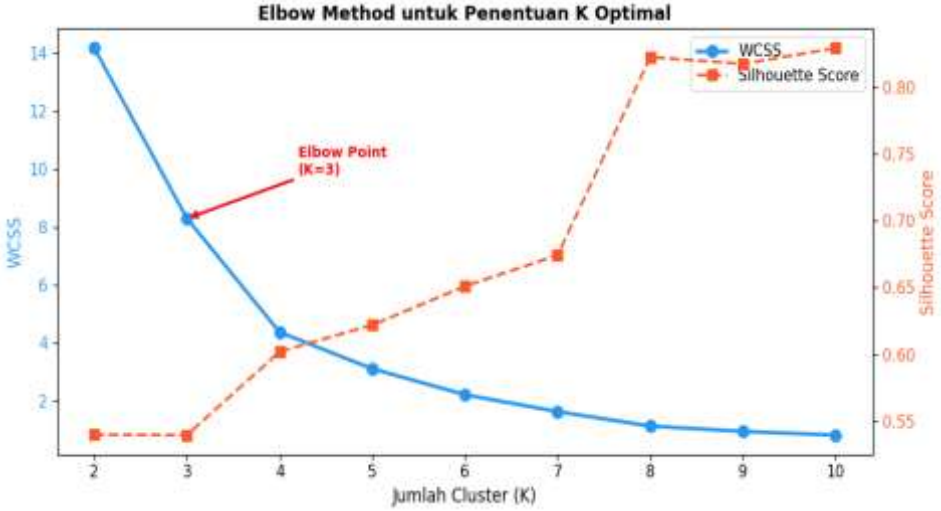
(rata-rata Rp379.206; std Rp114.037). Distribusi Recency yang condong ke nilai 0 (median=0) mengindikasikan mayoritas warga masih aktif membayar mendekati akhir periode. Standar deviasi Monetary sebesar Rp114.037 setara dengan deviasi 2,28 bulan pembayaran mengindikasikan variasi perilaku kontribusi yang signifikan dan heterogen antar warga, mengindikasikan variasi perilaku kontribusi yang cukup besar antar warga, sehingga pengelompokan relevan untuk dilakukan. Nilai Frequency terendah adalah 3 bulan, yang berarti bahkan warga yang paling tidak aktif pun pernah membayar setidaknya tiga kali dalam setahun, menunjukkan tidak ada warga yang sepenuhnya absen selama seluruh periode pengamatan.

3.2. Penentuan Jumlah Cluster Optimal

Tabel 1. Hasil Perhitungan WCSS dan Silhouette Score (K2-10)

K	WCSS	Silhouette Score
2	14,1670	0,5401
3 *	8,2868	0,5398
4	4,3589	0,6021
5	3,0971	0,6221
6	2,4183	0,6798
7	1,7350	0,7512
8	1,1163	0,8224
9	0,9332	0,8173
10	0,8074	0,8291

**Elbow point terpilih*



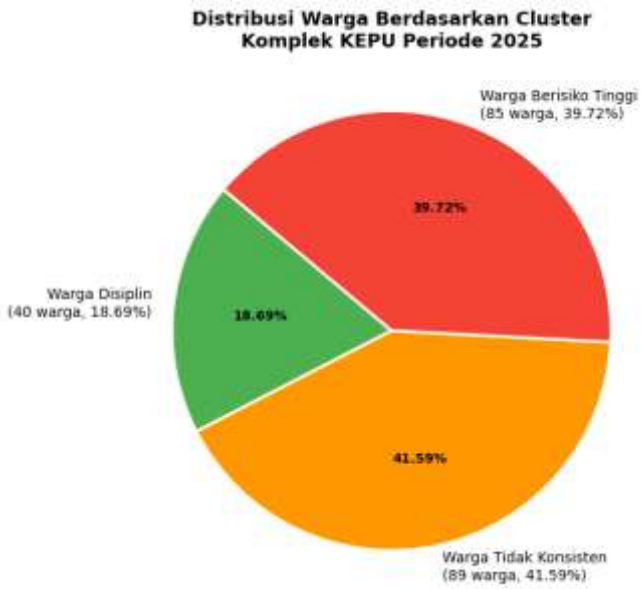
Gambar 1. Grafik Elbow Method untuk Penentu Nilai K Optimal

Grafik Elbow Method (Gambar 1) menampilkan kurva WCSS yang menurun dari 14,1670 pada K=2 hingga 0,8074 pada K=10. Penurunan WCSS terbesar secara absolut terjadi dari K=2 ke K=3 (5,8802 poin; 41,5%), kemudian melandai pada K≥4, meskipun penurunan relatif K=3 ke K=4 (47,4%) masih cukup besar sehingga siku tidak sepenuhnya tegas. Nilai Silhouette K=2 (0,5401) dan K=3 (0,5398) hampir sama. Meskipun Silhouette Score tertinggi terdapat pada K=8 (0,8224) dan K=10 (0,8291), segmentasi lebih dari lima cluster terlalu granular untuk ditindaklanjuti secara operasional oleh pengurus komplek dengan keterbatasan waktu dan sumber daya. Silhouette K=3 sebesar 0,5398 berada di atas ambang batas 0,5 yang mengindikasikan struktur cluster yang memadai [19]. K=3 dipilih berdasarkan pertimbangan gabungan: penurunan WCSS yang melandai setelah K=3, Silhouette di atas 0,5, serta tiga kelompok yang lebih informatif daripada dua kelompok untuk menentukan prioritas penagihan.

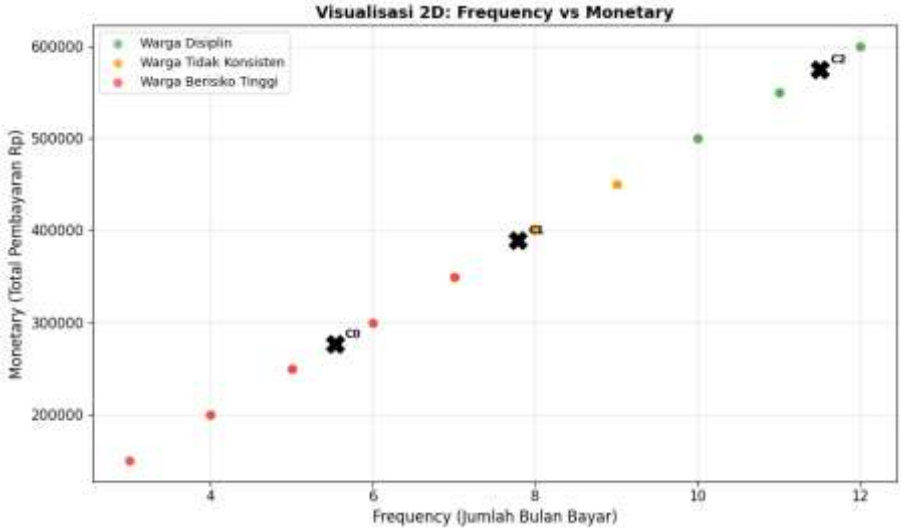
3.3. Hasil Implementasi K-Means

Tabel 2. Centroid Akhir dan Distribusi Cluster K-Means (K=3)

Label Cluster	R (Bln)	F (Bln)	M (Rp)	Warga	%
Cluster 2 – Warga Disiplin	0,000	11,500	575.000	40	18,69%
Cluster 1 – Tdk Konsisten	0,056	7,787	389.326	89	41,59%
Cluster 0 – Berisiko Tinggi	0,565	5,529	276.471	85	39,72%



Gambar 2. Distribusi Keanggotaan Cluster Hasil K-Means (K=3)



Gambar 3. Visualisasi Scatter Plot Hasil Clustering K-Means pada Atribut Frequency dan Monetary

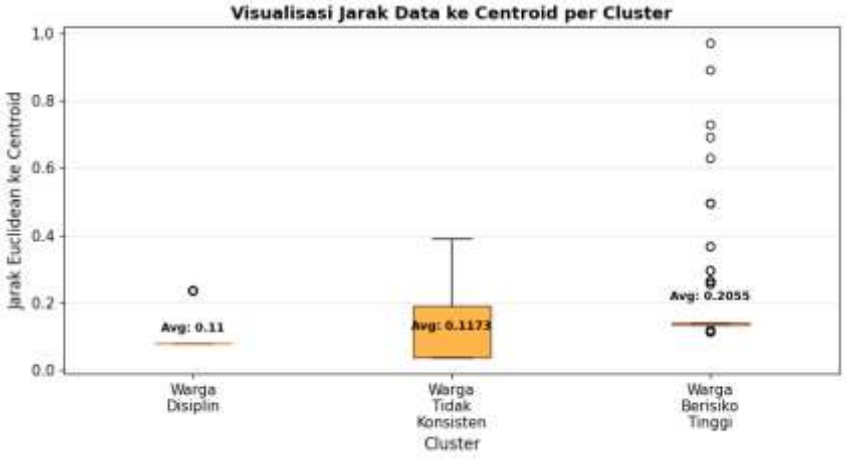
Distribusi keanggotaan cluster secara proporsional disajikan pada Gambar 2. Algoritma K-Means mencapai konvergensi dalam 3 iterasi dengan WCSS final 8,2868. Pemisahan geometris ketiga cluster pada ruang fitur Frequency–Monetary dapat diamati pada Gambar 3, di mana titik-titik dari masing-masing cluster membentuk kelompok yang terdistingsi dengan jelas.

Sebanyak 81,31% warga (174 dari 214) berada pada kategori tidak konsisten atau berisiko tinggi, yang menunjukkan perlunya penagihan yang lebih terarah. Apabila selisih penerimaan terhadap potensi diuraikan menurut kelompok, Cluster 0 menyumbang sekitar 58,2%, Cluster 1 sekitar 39,7%, dan Cluster 2 sekitar 2,1%. Cluster 2 (Warga Disiplin, 40 warga) memiliki Recency=0, artinya seluruh anggotanya tercatat membayar pada bulan Desember 2025. Frequency rata-rata 11,5 bulan dengan standar deviasi kecil (0,816) menandakan seluruh anggota konsisten membayar 11–12 bulan dari 12 bulan, membentuk kelompok yang homogen dan dapat diandalkan. Cluster 1 (Warga Tidak Konsisten, 89 warga) merupakan kluster terbesar dengan Frequency rata-rata 7,787 bulan artinya warga melewatkan pembayaran rata-rata 4,2 bulan per tahun. Recency yang mendekati 0 (0,056 bulan) menunjukkan sebagian besar masih aktif mendekati akhir periode, menandakan potensi pembinaan yang masih terbuka lebar. Cluster 0 (Warga Berisiko Tinggi, 85 warga) memiliki kombinasi Recency tertinggi (0,565 bulan) dan Frequency terendah (5,529 bulan), mengindikasikan warga yang hanya membayar sekitar 5–6 bulan dan sebagian sudah tidak aktif sejak beberapa bulan terakhir. Standar deviasi Recency yang paling besar (1,190) mencerminkan heterogenitas yang tinggi beberapa warga masih sesekali membayar sementara yang lain sudah lama absen.

3.4. Evaluasi Kuantitatif Clustering

Tabel 3. Evaluasi Kuantitatif Hasil Clustering K=3

Metrik	Nilai	Threshold	Interpretasi
Silhouette Coefficient	0,5398	> 0,5 = Baik	Struktur cluster kuat: pemisah memadai
Davies-Bouldin Index	0,7130	< 1,0 = Baik	Separasi antar cluster baik
Calinski-Harabasz Index	314,03	Semakin tinggi = Baik	Dispersi antar cluster sangat dominan
WCSS Final	8,2868	Semakin kecil = Baik	Kekompakan optimal di elbow K=3



Gambar 4. Visualisasi Jarak Data ke Centroid

Visualisasi scatter plot pada Gambar 3 secara konsisten dengan ketiga metrik evaluasi menunjukkan bahwa meskipun batas antar cluster tidak sepenuhnya rigid, pola pemisahan cukup jelas untuk mendukung pengambilan keputusan manajerial. Silhouette Coefficient 0,5398 mengkonfirmasi bahwa setiap objek data secara rata-rata jauh lebih dekat ke centroid cluster-nya sendiri dibanding centroid cluster tetangga. DBI 0,7130 yang jauh di bawah ambang 1,0 menandakan cluster yang kompak secara internal dan terpisah dengan jelas satu sama lain. CHI 314,03 yang tinggi mengindikasikan bahwa dispersi antar cluster jauh mendominasi dispersi di dalam cluster, sebuah karakteristik clustering berkualitas tinggi [21]. Secara kolektif, ketiga metrik berada dalam rentang yang diterima sebagai clustering berkualitas baik menurut standar yang ditetapkan [22].

3.5. Validasi Statistik Inferensial

Tabel 4. Hasil Uji Kruskal-Wallis dan Effect Size (Eta Squared)

Variabel	H-statistik	p-value	η^2 (Eta Squared)	Interpretasi η^2
Recency	24,2168	0,000006	0,1020	Efek Besar
Frequency	186,4697	< 0,0001	0,8808	Efek Sangat Besar
Monetary	186,4697	< 0,0001	0,8808	Efek Sangat Besar

Tabel 5. Hasil Uji Post-Hoc Mann-Whitney U Antar Pasangan Cluster

Pasangan Cluster	U-stat (R)	U-stat (F)	U-stat (M)	p-value (F&M)	Signifikan?
Cluster 2 vs Cluster 1	1700,0	0,0	0,0	< 0,0001	Ya (F,M); Tidak (R)
Cluster 2 vs Cluster 0	84,0	0,0	0,0	< 0,0001	Ya (R, F, M)
Cluster 1 vs Cluster 0	2405,0	0,0	0,0	< 0,0001	Ya (R, F, M)

Uji normalitas Shapiro-Wilk menunjukkan hampir seluruh kelompok pada ketiga atribut tidak berdistribusi normal ($p<0,0001$), sehingga uji Kruskal-Wallis digunakan sebagai alternatif non-parametrik yang valid menggantikan ANOVA one-way. Ketiga atribut menunjukkan perbedaan yang signifikan secara statistik antar cluster (seluruh $p<0,05$). Nilai H yang jauh lebih tinggi pada Frequency dan Monetary ($H=186,47$) dibandingkan Recency ($H=24,22$) mengkonfirmasi bahwa frekuensi dan total kontribusi pembayaran adalah dimensi pembeda yang paling dominan. Eta Squared $\eta^2=0,8808$ pada Frequency dan Monetary mengindikasikan bahwa 88,08% variansi total kedua variabel tersebut dapat dijelaskan oleh keanggotaan cluster sebuah effect size yang dikategorikan sangat besar menurut [23], yang menunjukkan bahwa perbedaan antar cluster nyata secara statistik.

Uji post-hoc Mann-Whitney U (Tabel 5) terhadap seluruh 9 kombinasi pasangan (3 pasangan × 3 variabel) menghasilkan 8 perbandingan yang signifikan. Satu pengecualian yang bermakna secara analitis adalah Recency antara Cluster 2 dan Cluster 1 ($U=1700,0$; $p=0,1776$; tidak signifikan), yang mengungkapkan bahwa kedua kelompok tidak dapat dibedakan berdasarkan kapan terakhir mereka membayar keduanya

sama-sama aktif mendekati akhir periode. Pembeda utama antara Warga Disiplin dan Warga Tidak Konsisten adalah Frequency dan Monetary semata. Temuan ini menghasilkan insight spesifik yang tidak akan terungkap dari nilai centroid saja: dalam konteks iuran bulanan yang bersifat kumulatif, kelengkapan pembayaran sepanjang tahun jauh lebih informatif sebagai indikator perilaku dibandingkan kebaruan pembayaran terakhir.

3.6. Pembahasan

Hasil penelitian ini sejalan dengan penelitian terdahulu yang menerapkan K-Means pada data pembayaran. [7] menerapkan K-Means pada data pembayaran SPP, dan penelitian ini menunjukkan bahwa pendekatan serupa dapat diterapkan pada data iuran komunitas perumahan yang memiliki dinamika sosial berbeda. Pemisahan tiga kelompok pada ruang fitur Frequency–Monetary (Gambar 3) juga sejalan dengan pemisahan cluster pada penelitian [9] di konteks kredit, dan adaptasi RFM ke konteks non-komersial sebagaimana dilakukan [10] pada donor non-profit tampak dapat diterapkan pula pada iuran komunitas.

Penelitian ini menambahkan uji statistik inferensial (Kruskal-Wallis dan Mann-Whitney U) di samping metrik internal (Silhouette, DBI, CHI) sebagai konfirmasi bahwa perbedaan antar cluster nyata secara statistik. Perlu dicatat bahwa karena cluster dibentuk dari variabel yang sama dengan yang diuji, perbedaan yang signifikan memang diharapkan, sehingga uji ini menegaskan keterpisahan cluster, bukan membuktikan kebermaknaan segmentasi di lapangan. Effect size η^2=0,8808 pada Frequency dan Monetary tergolong sangat besar menurut kriteria [23]. Recency tidak membedakan Cluster 2 dan Cluster 1 (p=0,1776), yang menunjukkan bahwa pada iuran bulanan akumulatif, kelengkapan pembayaran sepanjang tahun lebih informatif daripada kebaruan pembayaran terakhir.

Keterbatasan penelitian perlu dicatat secara transparan. Pertama, cakupan data satu tahun (2025) membatasi kemampuan mendeteksi tren perilaku jangka panjang dan menguji stabilitas cluster lintas waktu. Kedua, redundansi antara Monetary dan Frequency akibat iuran tetap menjadi keterbatasan struktural model RFM dalam konteks ini, meskipun uji sensitivitas membuktikan hal tersebut tidak mengubah hasil pengelompokan. Ketiga, penelitian tidak mempertimbangkan faktor-faktor eksternal yang potensial mempengaruhi perilaku pembayaran, seperti kondisi sosial-ekonomi, durasi tinggal, kepemilikan hunian (kontrak vs milik), maupun faktor kepuasan terhadap layanan pengurus. Keempat, hasil penelitian bersifat kontekstual dan tidak dapat digeneralisasikan secara langsung ke komunitas perumahan lain tanpa validasi ulang yang mempertimbangkan perbedaan demografis, ekonomi, dan karakteristik komunitas. Kelima, penelitian ini hanya menggunakan data pemasukan iuran dan tidak menganalisis penyebab ketidakteraturan pembayaran, termasuk transparansi pengelolaan dana dan kepercayaan warga terhadap pengurus; rekomendasi yang dihasilkan juga belum diuji penerapannya di lapangan.

4. KESIMPULAN

Pembayaran iuran warga Komplek Kepu pada tahun 2025 belum merata: iuran yang tercatat terbayar baru mencapai 63,2% dari potensi penerimaan. Berdasarkan atribut RFM, 214 warga terkelompok menjadi tiga cluster melalui K-Means (K=3; Silhouette=0,5398; DBI=0,7130; CHI=314,03), yaitu Warga Disiplin (40 warga/18,69%), Warga Tidak Konsisten (89 warga/41,59%), dan Warga Berisiko Tinggi (85 warga/39,72%). Ketiga cluster berbeda secara statistik pada atribut RFM, dengan Frequency dan Monetary sebagai pembeda utama, sedangkan Recency tidak membedakan Warga Disiplin dan Tidak Konsisten (p=0,1776). Berdasarkan karakteristik tersebut, diusulkan pendekatan penagihan terdiferensiasi: apresiasi bagi Warga Disiplin, pengingat terjadwal bagi Warga Tidak Konsisten, dan pendekatan bertahap melalui koordinator wilayah atau blok bagi Warga Berisiko Tinggi. Rekomendasi ini merupakan usulan berbasis data historis dan belum diuji di lapangan. Penelitian selanjutnya disarankan untuk: (1) menguji penerapan rekomendasi dan mengukur dampaknya terhadap tingkat pembayaran; (2) mengkaji faktor penyebab ketidakteraturan pembayaran, termasuk persepsi transparansi pengelolaan dana, kepercayaan terhadap pengurus, dan kondisi ekonomi warga; (3) memperluas periode data menjadi ≥2 tahun; (4) mengeksplorasi K=4–5; (5) membandingkan K-Means dengan K-Medoids, DBSCAN, atau Hierarchical Clustering; serta (6) menambahkan variabel kontekstual di luar RFM seperti durasi tinggal atau status kepemilikan hunian.

REFERENCES

[1] BPS, “Badan Pusat Statistik (BPS). (2024). Indikator Kesejahteraan Rakyat 2024.,” 2024. Accessed: Jun. 01, 2026. [Online]. Available: https://www.bps.go.id/id/publication/2024/11/06/3ef10d3d82ed93f616ba9113/indikator-kesejahteraan-rakyat-2024.html

[2] M. Y. Yustrinita and S. O. Aprilia, “Sistem Pengelolaan Data Warga, Informasi Kegiatan dan Informasi Penggunaan Iuran pada Lingkungan Rukun Tetangga (RT) Berbasis Web,” *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 1, p. 77, Mar. 2021, doi: 10.32493/informatika.v6i1.8283.

[3] Maulida and Fariz, “Sistem Informasi Pembayaran Iuran Bulanan Berbasis Website Studi Kasus: Rw 004 Ragunan Pasar Minggu Jakarta Selatan Dki Jakarta,” Jakarta, 2022. Accessed: Jun. 01, 2026. [Online]. Available: https://elibrary.unikom.ac.id/id/eprint/7537/

[4] R. Prameswari and Miharyono, "RANCANG BANGUN SISTEM INFORMASI PENCATATAN KEUANGAN IURAN WARGA (STUDI KASUS: PERUMAHAN DUTA GRAHA RT 04 TANGERANG BANTEN)," 2022. Accessed: Jun. 01, 2026. [Online]. Available: http://repository.unsada.ac.id/4932/
[5] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
[6] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," Aug. 01, 2020, *MDPI AG*. doi: 10.3390/electronics9081295.
[7] E. Susilowati, U. Al Ghazali Cilacap Jalan Kemerdekaan Barat No, K. Kidul, and K. Kesugihan, "PENERAPAN METODE K-MEANS CLUSTER UNTUK MENGKLASIFIKASI PEMBAYARAN SPP SISWA (STUDI KASUS SMK DIPONEGORO 2 RAWALO TAHUN 2022-2023) APPLICATION OF K-MEANS CLUSTER METHOD TO CLASSIFY STUDENTS' SPP PAYMENTS (CASE STUDY OF SMK DIPONEGORO 2 RAWALO ACADEMIC YEAR 2022-2023)," 2023, doi: 10.36987/jpms.v10i2.7131.
[8] D. F. Suntoro, N. Fitriani, and A. Wibowo, "SEGMENTASI PENDAPATAN DARI PAYMENT AGGREGATOR MENGGUNAKAN METODE KLASTERISASI K-MEANS," *Jurnal Komputer dan Informatika*, vol. 13, no. 1, pp. 18–25, Mar. 2025, doi: 10.35508/jicon.v13i1.12691.
[9] S. Nur Illah, N. Suarna, I. Ali, and D. Solihudin, "Journal of Artificial Intelligence and Engineering Applications K-Means Clustering Method to Make Credit Payment Groupinhg Efficient," 2025. [Online]. Available: https://ioinformatic.org/
[10] N. Trezandy Lapatta, R. Ardiansyah, and D. Shinta Angreni, "Donor Segmentation Analysis Using the RFM Model and K-Means Clustering to Optimize Fundraising Strategies," 2024. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC
[11] Irvansah Satria Pamungkas, Dhimas Maulana Hadi, Gilang Aditya, Rosyidah Astari Nur, Aries Saifudin, and Teti Desyani, "Klasterisasi Pengunjung Mall untuk Menentukan Target Pasar Ponsel Terbaru Menggunakan Algoritma K-Means Clustering," *Jurnal Informatika Universitas Pamulang*, vol. 6, Sep. 2021, Accessed: Jun. 03, 2026. [Online]. Available: https://media.neliti.com/media/publications/467299-none-aefb5239.pdf
[12] Ahzril Pria Adistya, Novara Luthfyani, Perahim Tara, Richky Adriyan, Rifaldi, and Perani Rosyani, "Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa," *OKTAL: Jurnal Ilmu Komputer dan Sains*, vol. 2, no. 8, pp. 2301–2306, 2023, Accessed: Jun. 03, 2026. [Online]. Available: https://journal.mediapublikasi.id/index.php/oktal/article/view/1610/1857
[13] E. Weyant, "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th Edition," *Journal of Electronic Resources in Medical Libraries*, vol. 19, no. 1–2, pp. 54–55, Apr. 2022, doi: 10.1080/15424065.2022.2046231.
[14] A. M. Hughes, *Strategic database marketing*. McGraw-Hill Pub. Co., 2005.
[15] R. F. Devi, F. H. Siswanto, N. Azkia, and J. Heikal, "Customer Segmentation Based on RFM Analysis as the Basis of Marketing Strategy Case Study of PT Pertiwi Agung Pharmaceutical Industry (LANDSON)," *BUDGETING : Journal of Business, Management and Accounting*, vol. 5, no. 2, pp. 766–774, Apr. 2024, doi: 10.31539/budgeting.v5i2.8994.
[16] J. Macqueen, "SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS," 1967.
[17] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (N. Y).*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/J.INS.2022.11.139.
[18] D. Arthur and S. Vassilvitskii, "k-means++: The Advantages of Careful Seeding," 2007.
[19] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," 1987.
[20] D. L. Davies and D. W. Bouldin, "A cluster separation measure," 1979.
[21] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974, doi: 10.1080/03610927408827101.
[22] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.
[23] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences Second Edition," 1988.