

Penerapan Algoritma K-Means Clustering Untuk Sistem Segmentasi Pelanggan Berdasarkan Model Rfm (Studi Kasus: PT Sehati Bangunan Abadi)

Lusi Damayanti^{1*}, Ahmad Faisol², Nurlaily Vendyansyah³
Fakultas Teknologi Industri, Program Studi Teknik Informatika
Institut Teknologi Nasional Malang
2218135@scholar.itn.ac.id

Abstrak

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan berdasarkan data transaksi menggunakan algoritma K-Means Clustering pada PT Sehati Bangunan Abadi (SBA) tahun 2024. Data yang digunakan mencakup tiga variabel utama, yaitu Recency, Frekuensi, dan Total Nominal, yang sebelumnya telah melalui proses pembersihan serta normalisasi agar setiap variabel memiliki skala yang seimbang. Proses pengelompokan dilakukan dengan menghitung jarak menggunakan rumus *Euclidean Distance* untuk menentukan kesamaan antar data. Hasil penelitian menunjukkan bahwa pelanggan dapat dikelompokkan menjadi tiga kategori, yaitu Prioritas, VVIP, dan VIP, dengan nilai Silhouette Score masing-masing sebesar 0.5884 untuk KAI, 0.5101 untuk End User, dan 0.6274 untuk Singres Member. Nilai tersebut menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik.

Kata kunci: K-Means Clustering, Segmentasi Pelanggan, Data Mining

Abstract

This study aims to perform customer segmentation based on transaction data using the K-Means Clustering algorithm at PT Sehati Bangunan Abadi (SBA) in 2024. The data used includes three main variables, namely Recency, Frequency, and Total Amount, which have previously undergone a cleaning and normalization process to ensure that each variable has a balanced scale. The clustering process was carried out by calculating the distance using Euclidean Distance to determine the similarity between data points. The results show that customers can be grouped into three categories: Priority, VVIP, and VIP, with Silhouette Score values of 0.5864 for KAI, 0.51017 for End User, and 0.6274 for Singres Member. These values indicate that the clustering results have a fairly good quality.

Keywords: K-Means Clustering, Customer Segmentation, Data Mining

PENDAHULUAN

PT Sehati Bangunan Abadi (SBA) merupakan perusahaan yang berfokus pada distribusi serta penjualan material bangunan dan berkedudukan di Tangerang. Entitas usaha tersebut dikenal dengan konsep “supermarket bangunan”, yang menyediakan berbagai kebutuhan konstruksi dan interior secara lengkap, mulai dari material dasar hingga produk finishing. Dalam menjalankan

operasionalnya, SBA memiliki beragam pelanggan yang berasal dari berbagai kategori. Seiring dengan peningkatan jumlah pelanggan setiap tahunnya, perusahaan menghadapi tantangan dalam mengidentifikasi pelanggan yang memberikan kontribusi terbesar terhadap pendapatan.

Selama ini, proses penilaian pelanggan masih dilakukan secara manual, sehingga membutuhkan waktu yang cukup

lama dan berpotensi menimbulkan ketidaktepatan dalam menentukan pelanggan yang paling loyal. Akibatnya, strategi pemasaran dan program loyalitas yang diterapkan belum sepenuhnya efektif dalam meningkatkan retensi pelanggan maupun penjualan.

Untuk menjawab permasalahan tersebut, diperlukan sebuah sistem yang mampu mengelompokkan pelanggan secara lebih objektif dan efisien berdasarkan riwayat transaksinya. Salah satu pendekatan yang sesuai yaitu algoritma K-Means, yaitu yang memisahkan data ke dalam sejumlah kelompok (cluster) berdasarkan tingkat kemiripan karakteristik tertentu (Dinatri, Lestanti, & Budiman, 2023). Melalui pendekatan ini, pelanggan dapat dikelompokkan menjadi beberapa segmen seperti Prioritas, VVIP, dan VIP.

Hasil segmentasi yang dihasilkan dari sistem ini berperan dalam mendukung perusahaan merancang strategi pemasaran yang lebih terarah, misalnya dengan memberikan penawaran khusus kepada pelanggan tertentu atau menyusun program loyalitas yang sesuai dengan profil pelanggan. Selain itu, sistem segmentasi pelanggan berbasis website ini juga mendukung manajemen dalam memantau perilaku pelanggan secara langsung, sehingga proses proses penentuan keputusan dapat berlangsung dengan lebih efisien dan tepat.

Meskipun metode K-Means telah banyak digunakan dalam berbagai penelitian untuk keperluan clustering, sebagian besar penelitian sebelumnya masih berfokus pada domain akademik, layanan, atau penjualan produk secara umum, dan belum secara spesifik mengimplementasikannya pada segmentasi pelanggan perusahaan material

bangunan dengan memanfaatkan data transaksi sebagai dasar pengelompokan. Oleh karena itu, penelitian ini memiliki kebaruan pada penerapan K-Means yang diintegrasikan ke dalam sistem berbasis website untuk melakukan segmentasi pelanggan PT Sehati Bangunan Abadi berdasarkan perilaku transaksi, sehingga diharapkan mampu memberikan kontribusi praktis dalam mendukung pengambilan keputusan bisnis perusahaan.

Rumusan Masalah

Berikut Rumusan masalah pada penelitian saya:

1. Bagaimana merancang sistem segmentasi pelanggan berbasis website yang dapat mengelompokkan pelanggan PT Sehati Bangunan Abadi berdasarkan data transaksi yang dimiliki?
2. Bagaimana mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan pelanggan KAI, Singres Member, End user menjadi tiga kategori, yaitu Prioritas, VIP, dan VVIP?

Tujuan Penelitian

Penelitian ini disusun untuk mencapai beberapa tujuan antara lain:

1. Untuk merancang sistem segmentasi pelanggan yang mampu mengelompokkan pelanggan PT Sehati Bangunan Abadi berdasarkan data transaksi.
2. Untuk mengimplementasikan algoritma K-Means Clustering dalam proses pengelompokan pelanggan KAI, Singres Member, End user menjadi tiga kategori, yaitu Prioritas, VIP, dan VVIP.

Manfaat Penelitian

Terdapat beberapa manfaat sebagai berikut:

1. Penelitian ini hanya berfokus pada proses segmentasi pelanggan pelanggan KAI , Singres Member , End user pada PT Sehati Bangunan Abadi berdasarkan data transaksi.
2. Mempermudah identifikasi pelanggan prioritas sehingga dapat meningkatkan loyalitas dan potensi penjualan.
3. Menjadi referensi penerapan algoritma K-Means Clustering pada segmentasi pelanggan.
4. Mendapatkan pelayanan yang lebih sesuai dengan kebutuhan dan preferensi mereka

Tinjauan Pustaka

Menurut Yudono Prayoga dan rekan penelitiannya dengan judul “Penerapan Metode K-Means Pada Sistem Informasi Akademik Untuk Pengelompokan Siswa Berprestasi di Upt Sma Negeri 3 Kota Pasuruan Berbasis Web” sistem tersebut dirancang untuk mengklasifikasikan siswa menjadi tiga kelompok, yakni siswa dengan prestasi tinggi, siswa dengan capaian menengah, serta siswa yang berada pada kategori rendah. Sistem diuji fungsionalitasnya, dan hasil pengelompokan mencapai tingkat kesesuaian 88% . (Prayoga, Mahmudi, & Zahro, 2021).

Menurut Andy Febrianto, dan rekan penelitiannya dengan judul “Penerapan Metode K-Means Untuk Clustering Pengunjung Perpustakaan Itn Malang” yang bertujuan untuk mengelompokkan data pengunjung dan peminjaman di perpustakaan ITN Malang. Aplikasi berbasis website ini memudahkan petugas perpustakaan dalam menganalisis pola kunjungan dan peminjaman. Hasil

pengujian menunjukkan aplikasi bekerja sangat baik (Febrianto, Achmadi, & Sasmito, 2021).

Menurut Annisa Chintia, dan rekan penelitiannya dengan judul “Penerapan Metode K-Means Clustering Untuk Pengelompokan Data Barang Penjualan Berbasis Web Pada Koperasi Pt. X” yang bertujuan untuk mengelompokkan data barang penjualan pada koperasi berbasis web. Produk dikelompokkan ke dalam kategori "paling laku", "laku", dan "tidak laku", dan hasil clustering sistem selaras 100% dengan perhitungan manual di Excel (ChintiaDevi, Zahro', & Vendyansyah, 2020).

Berdasarkan penelitian-penelitian terdahulu tersebut, dapat disimpulkan bahwa metode K-Means terbukti efektif digunakan untuk proses pengelompokan data pada berbagai bidang, seperti bidang akademik, layanan perpustakaan, dan penjualan produk. Meskipun menggunakan algoritma yang sama, setiap penelitian memiliki perbedaan pada objek penelitian, variabel yang digunakan, serta tujuan pengelompokan. Penelitian Prayoga et al. berfokus pada pengelompokan prestasi siswa, penelitian Febrianto et al. menekankan pada analisis pola aktivitas pengunjung perpustakaan, sedangkan penelitian ChintiaDevi et al. berfokus pada pengelompokan tingkat kelakuan produk.

Namun, penelitian-penelitian tersebut belum secara spesifik membahas segmentasi pelanggan berdasarkan perilaku transaksi dengan pendekatan Recency, Frequency, dan Monetary (RFM) sebagai dasar dalam mendukung pengambilan keputusan bisnis. Oleh karena itu, penelitian ini diposisikan untuk melengkapi penelitian sebelumnya dengan menerapkan metode K-Means dalam mengelompokkan pelanggan berdasarkan

data transaksi, sehingga hasil clustering yang dihasilkan dapat dimanfaatkan sebagai sistem pendukung keputusan bagi perusahaan.

LANDASAN TEORI

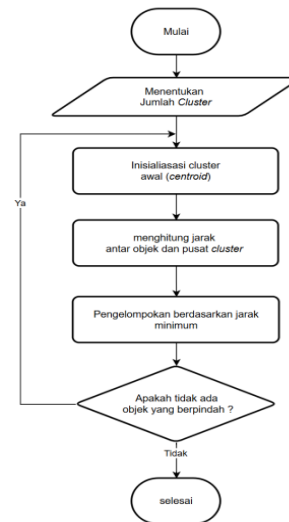
Segmentasi Pelanggan

Segmentasi pelanggan adalah proses pengelompokan konsumen ke dalam beberapa kelompok atau segmen berdasarkan kesamaan karakteristik, kebutuhan, maupun pola perilakunya dalam melakukan transaksi. Melalui proses ini, perusahaan dapat mengenali pola perilaku pelanggan secara lebih mendalam sehingga strategi pemasaran yang diterapkan menjadi lebih efektif, efisien, dan sesuai dengan kebutuhan masing-masing segmen (Andrew Lomaksan Manuel Tampubolon, 2024). Menurut Wahyuni dkk. (2023) dalam *Jurnal SimanteC*, segmentasi pelanggan adalah metode pengelompokan pelanggan ke dalam beberapa *cluster* (kelompok) di mana setiap kelompok memiliki anggota dengan karakteristik yang mirip satu sama lain. Tujuan utama dari segmentasi ini adalah agar perusahaan dapat memahami perbedaan kebutuhan antar pelanggan dan menyesuaikan penawaran produk atau layanan yang paling relevan bagi tiap segmen.

Algoritma K-Means

Algoritma *K-Means* digunakan untuk mengelompokkan data yang sederhana dan banyak dipakai. Popularitasnya disebabkan oleh kemudahan penerapan serta kecepatannya dalam memproses data berukuran besar. Pada dasarnya, *K-Means* bekerja dengan konsep titik pusat (*centroid*) sebagai dasar pembentukan kluster (Dinanti, Lestanti, & Budiman,

2023). *Flowchart* Algoritma *K-Means Clustering* sebagai berikut:



Gambar 1. Algoritma K-Means

Berdasarkan Gambar 1. tahapan pada algoritma *K-Means Clustering* dapat dijelaskan sebagai berikut:

- Menentukan jumlah *cluster* (*K*) yang diinginkan
- Menginisialisasi vektor centroid secara acak.
- Menghitung jarak antara setiap data dengan centroid menggunakan metode *Euclidean Distance* sesuai Persamaan (1).

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Keterangan :

X_i : Data kriteria

μ_j : centroid pada cluster ke-*j*

- Menghitung kembali posisi centroid setiap cluster menggunakan Persamaan (2)

$$M_j = \frac{1}{n_j} (\sum Z_p) \quad (2)$$

Keterangan :

(M_j) : nilai centroid baru untuk cluster ke-*j*

N_j : jumlah data dalam cluster ke-*j*.

$\sum(Z_p)$: jumlah semua data dalam cluster ke-*j*

- Mengulangi langkah ke-3 dan ke-hingga posisi centroid tidak berubah signifikan atau jumlah iterasi yang

ditentukan tercapai. Proses berhenti ketika hasil pengelompokan telah konvergen.

Normalisasi Data

Pada tahap pra-pemrosesan data, dilakukan proses normalisasi untuk menghindari perbedaan skala antar variabel yang terlalu besar. Proses ini menggunakan metode Min-Max Normalization dengan rentang nilai antara 0 hingga 1. Rentang tersebut dipilih agar seluruh data berada dalam skala yang sama, di mana nilai 0 mewakili nilai terendah (minimum) dan nilai 1 mewakili nilai tertinggi (maksimum) (Perdana, S. A., Florentin, S. F., & Santoso, A. 2022).

Proses normalisasi dalam penelitian ini menggunakan Persamaan (3):

$$X_n = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Keterangan :

- X_n : Nilai hasil normalisasi
- x_0 : Nilai data awal
- x_{min} : Nilai minimum data
- x_{max} : Nilai maximum data

Silhouette Score

Silhouette Score merupakan metode evaluasi yang digunakan untuk menilai sejauh mana algoritma *clustering* berhasil membedakan data ke dalam kelompok yang berbeda. Teknik ini menghitung nilai *silhouette* untuk setiap data, dimana menunjukkan seberapa tepat suatu data berada di dalam klusternya sendiri dibandingkan dengan kedekatannya terhadap kluster lain (Hendrastuty, 2024).

$$s(x) = \frac{b(x) - a(x)}{\max(a(x), b(x))} \quad (4)$$

Keterangan:

$S(x)$: nilai silhouette

$b(x)$: rata-rata jarak dari cluster yang sama.

$a(x)$: rata-rata jarak dari cluster yang berbeda.

RFM

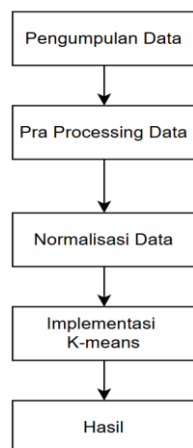
Model Recency, Frequency, dan Monetary (RFM) merupakan pendekatan segmentasi pelanggan yang didasarkan pada perilaku transaksi historis (Sembiring & Simatupang, 2023). Recency menunjukkan jarak waktu sejak transaksi terakhir pelanggan, Frequency menggambarkan intensitas transaksi dalam periode tertentu, sedangkan Monetary merepresentasikan total nilai transaksi pelanggan (Prasetyo & Ramdhani, 2022).

Kombinasi ketiga variabel tersebut digunakan untuk mengidentifikasi tingkat aktivitas, loyalitas, dan nilai ekonomi pelanggan. Dalam penelitian ini, variabel RFM dimanfaatkan sebagai atribut utama yang selanjutnya diproses menggunakan metode K-Means Clustering untuk membentuk kelompok pelanggan dengan karakteristik perilaku transaksi yang serupa.

METODE PENELITIAN

Penelitian ini menggunakan metode K-Means *Clustering* untuk mengelompokkan pelanggan berdasarkan kemiripan karakteristik perilaku transaksinya. Proses pengolahan data dan penerapan algoritma dilakukan melalui aplikasi berbasis web yang dibangun menggunakan *framework Laravel* (PHP) sebagai *software* pendukung utama, dengan basis data MySQL untuk pengelolaan data. Langkah pertama dimulai dari proses pengumpulan data transaksi, kemudian data tersebut melalui tahap pembersihan dan penyesuaian agar tidak ada nilai yang hilang atau tidak sesuai. Setelah itu, dilakukan proses normalisasi supaya setiap atribut memiliki skala yang seragam dan hasil perhitungan menjadi lebih akurat.

Tahapan berikutnya adalah penerapan algoritma K-Means, di mana sistem akan menentukan titik pusat awal (*centroid*) secara acak. Selanjutnya, setiap data ditentukan jaraknya ke setiap centroid dengan memakai perhitungan *Euclidean Distance* yang dapat mengetahui kedekatannya dengan suatu kelompok. Nilai centroid akan diperbarui secara berulang sampai posisinya stabil atau tidak berubah secara signifikan. Proses ini berhenti ketika hasil pengelompokan dianggap konvergen, yang berarti data sudah terbagi secara optimal ke dalam beberapa kluster pelanggan yang memiliki kesamaan pola tertentu.



Gambar 2. Alur Metode Penelitian

ANALISIS DAN PERANCANGAN

1. Pengumpulan data

Pengumpulan Data Pada diambil dari data transaksi pelanggan PT Sehati Bangunan Abadi (SBA) untuk periode tahun 2024. Data yang diperoleh mencakup berbagai atribut pelanggan dan transaksi yang nantinya menjadi dasar dalam proses analisis.

2. Pra-Processing Data

Setelah data terkumpul, dilakukan pembersihan data (*data cleaning*) untuk menghilangkan duplikasi, memperbaiki kesalahan input, serta menangani data

yang tidak lengkap. Tahap ini bertujuan agar data yang digunakan benar-benar valid dan siap untuk dianalisis lebih lanjut.

3. Normalisasi Data

Proses normalisasi dilakukan agar setiap atribut data berada dalam skala yang seragam. Hal ini penting karena perbedaan satuan atau rentang nilai dapat memengaruhi hasil perhitungan jarak dalam metode *K-Means*.

4. Implementasi K-Means

Pada fase ini, teknik K-Means diterapkan untuk mengelompokkan data sesuai tingkat kemiripannya. Prosesnya dimulai dengan menentukan jumlah cluster yang akan dibuat serta menetapkan centroid awal secara acak. Setelah itu, setiap jarak tiap data ke centroid dihitung dengan memanfaatkan formula *Euclidean Distance*. Nilai-nilai jarak tersebut kemudian digunakan guna menyesuaikan posisi centroid yang baru sampai proses pengelompokan mencapai kondisi stabil atau tidak mengalami perubahan lagi.

5. Hasil

Tahap akhir adalah interpretasi hasil klusterisasi. Data yang telah terbagi dalam beberapa kelompok pelanggan kemudian dianalisis untuk memahami pola dan karakteristik tiap kluster.

IMPLEMENTASI DAN PEMBAHASAN

1. Pengumpulan Data

Penelitian ini menggunakan dataset hasil pengolahan data transaksi pelanggan PT Sehati Bangunan Abadi (SBA) pada periode tahun 2024. Data diperoleh dari sistem penjualan perusahaan yang telah melalui proses pembersihan dan validasi agar informasi yang digunakan akurat dan konsisten. Dataset mencakup data Transaksi sebanyak 1334 dengan data pelanggan sebanyak 165, yang digunakan

sebagai dasar dalam proses analisis menggunakan metode *K-Means Clustering*.

Tabel 1. Data Transaksi

N0	ID Transaksi	Tanggal	Nama Customer	Nama Barang	Kategori Barang	Total Transaksi
1	SBA01DTSI24010020	09/01/2024	Mitra Sembada	SINGRES (DC) WT HAKO BABY BLUE (G) 10X10CM (90LBR=0.9 0M2)	1	Rp 3.790.800
2	SBA01DTSI24010041	12/01/2024	Mitra Sembada	SINGRES WT TAKE HONEY (G) 7.5X30CM (52LBR=1.1 7M2)	1	Rp 21.375.974
3	SBA01DTSI24010021	12/01/2024	Mitra Sembada	SINGRES MOSAIC SQUARE CARRARA (M) 30.2X30.2CM (20LBR=1.8 2M2)	1	Rp 113.965.766
4	SBA01DTSI24010020	09/01/2024	Mitra Sembada	SINGRES (DC) WT HAKO OLIVE GREEN (G) 10X10CM (90LBR=0.9 0M2)	1	Rp 3.790.800
5	SBA01DTSI24010072	17/01/2024	Xenix Interior	SINGRES MOSAIC FISH SCALE ROSE (G) 29.2X33CM (13LBR=1.2 5M2)	1	Rp 13.994.360
1344	SBA01DTSI24010011	08/01/2024	Mitra Sembada	SINGRES MOSAIC KIT-KAT PISTACHIO (G) 28X30.5CM (15LBR=1.2 8M2)	1	Rp 2.034.483

Tabel 2. Data Perhitungan

kode	Recency	Frekuensi	Total Nominal
EUC000001	643	1	1430000
EUC000005	636	1	1106280
KAI000012	607	2	10540530
KAI000030	649	1	1361880
KAI000041	650	1	4868820
.....
KAI000049	639	2	20681986

2. Normalisasi Data

Normalisasi data dilakukan menggunakan teknik *Min-Max Scaling*, yaitu metode yang mengonversi nilai setiap atribut ke dalam skala 0 sampai 1. Tujuannya adalah agar semua variabel berada pada rentang yang sama sehingga setiap atribut memiliki pengaruh yang seimbang dalam perhitungan jarak pada metode K-Means meliputi atribut inti yaitu Recency, Frekuensi, serta Total Nominal Transaksi. Hasil Normalisasi dilihat pada table 3.

Tabel 3. Hasil Normalisasi Data

Recency	Frekuensi	Total Nominal
0,867256637	0	0,030096122

0,805309735	0	0,016035339
0,548672566	1	0,425812126
0,920353982	0	0,027137328
0,92920354	0	0,179461319
.....
0,831858407	1	0,866306417

3. Penerapan K-Means Clustering

a. Menentukan Titik Pusat *cluster* (*centroid*)

Pada tahap ini, ditentukan jumlah $k = 3$, sehingga titik pusat cluster yang diambil juga sebanyak 3 data secara acak dari tabel normalisasi, yaitu pada tiga baris awal tabel tersebut. Hasil menentukan titik pusat dilihat pada tabel 4.

Tabel 4. Centroid Awal

Centroid	Recency	Frekuensi	Total Nominal
C1	0,2941	0	0,0049
C2	0,431	0	0
C3	1	0,5	0,1441

b. Perhitungan Jarak Antar Cluster

Selanjutnya dilakukan perhitungan jarak tiap data terhadap centroid dihitung menggunakan formula *Euclidean Distance* untuk menentukan tingkat kedekatan antar data. Rekapitulasi hasil perhitungan jarak antar cluster ditampilkan pada Tabel 5.

Tabel 5. Perhitungan data setiap cluster

Kode	C1	C2	C3	Jarak Terdekat	cluster
EUC000001	1430000,114	1430000,128	1429999,719	1429999,719	3
EUC000005	1106280,152	1106280,166	1106279,757	1106279,757	3
KAI000012	10540529,99	10540530	10540529,59	10540529,59	3
KAI000030	1361880,124	1361880,138	1361879,729	1361879,729	3
KAI000041	4868820,013	4868820,027	4868819,618	4868819,618	3
.....
KAI000049	20681985,98	20681985,99	20681985,58	20681985,58	3

c. Proses Pembaharuan Centroid

Setelah tahap perhitungan jarak selesai, dilakukan pembaruan nilai centroid pada setiap iterasi dengan cara menghitung nilai rata-rata seluruh data yang berada pada tiap cluster. Pembaruan nilai centroid dapat diamati pada Tabel 6.

Tabel 6. Proses Pembaruan Centroid

Cluster	Recency	Frekuensi	Total Nominal
---------	---------	-----------	---------------

C1	0	0	0
C2	0	0	0
C3	5733,9	12,2	55822761,6

Proses iterasi akan berhenti ketika posisi centroid tidak mengalami perubahan lagi atau komposisi cluster tetap sama seperti iterasi sebelumnya. Jika kondisi ini telah tercapai, maka proses perhitungan pada metode K-Means dinyatakan telah mencapai titik konvergen dan dinyatakan selesai. Ringkasan hasil akhir pengelompokan data disajikan pada Tabel 7.

Tabel 7. Hasil K-Means Clustering

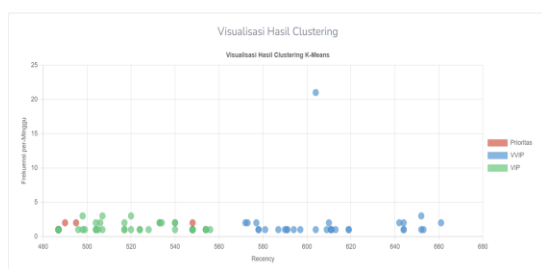
Kode	C1	C2	C3	Jarak Terdekat	cluster
EUC000001	53885461,64	692900,0069	1430000,145	692900,0069	2
EUC000005	54209181,64	369180,0112	1106280,183	369180,0112	2
KAI000012	44774931,68	9803430	10540530,02	9803430	2
KAI000030	53953581,64	624780,0087	1361880,155	624780,0087	2
KAI000041	50446641,65	4131720,001	4868820,043	4131720,001	2
*****	*****	*****	*****	*****	*****
KAI000049	34633475,74	19944886	20681986,01	19944886	2

Hasil *clustering* kemudian divalidasi menggunakan *silhouette score*. Nilai Silhouette Score sebesar 0.588 untuk kategori KAI, 0.6274 untuk kategori Singres Member, dan 0.51017 untuk kategori End User menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik. Hasil *silhouette score* disajikan pada Table 8.

Tabel 8. *silhouette score*

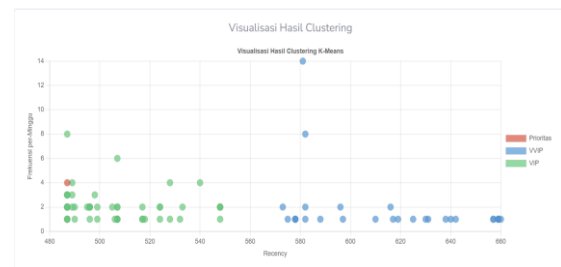
Kategori Pelanggan	<i>silhouette score</i>
KAI	0,5884
Singres Member	0,6274
End user	0,5101

4. Tampilan Visualisasi Clustering



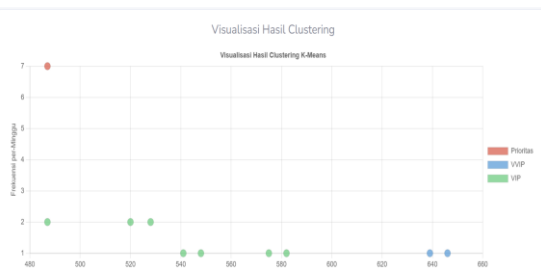
Gambar 3. Visualisasi Hasil Cluster kategori pelanggan KAI

Berdasarkan Gambar 3, hasil visualisasi clustering untuk kategori pelanggan KAI menunjukkan bahwa metode K-Means berhasil mengelompokkan pelanggan menjadi tiga kelompok utama, meliputi Cluster 1 (Prioritas) 3 pelanggan, Cluster 2 (VVIP) 30 pelanggan, dan Cluster 3 (VIP) 42 pelanggan. Pelanggan yang memiliki nilai *recency* rendah dan frekuensi tinggi termasuk dalam cluster Prioritas, yang menggambarkan tingkat aktivitas serta loyalitas yang tinggi. Sementara itu, pelanggan dengan *recency* tinggi dan frekuensi rendah cenderung tergabung dalam cluster VIP dan VVIP.



Gambar 4. Visualisasi Hasil Cluster kategori pelanggan Singres Member

Berdasarkan Gambar 4, hasil visualisasi *clustering* untuk kategori pelanggan Singres Member memperlihatkan bahwa metode K-Means berhasil mengelompokkan pelanggan ke dalam tiga cluster, yaitu Cluster 1 (Prioritas) sebanyak 1 pelanggan, Cluster 2 (VVIP) sebanyak 28 pelanggan dan Cluster 3 (VIP) sebanyak 48 pelanggan. Pelanggan yang memiliki nilai *recency* rendah dan *frekuensi* tinggi termasuk ke dalam cluster Prioritas, sedangkan pelanggan dengan *recency* tinggi dan *frekuensi* rendah tergolong dalam cluster VIP dan VVIP.



Gambar 5. Visualisasi Hasil Cluster kategori pelanggan End User

Berdasarkan Gambar 5, hasil visualisasi clustering pada kategori pelanggan End User memperlihatkan bahwa metode K-Means juga berhasil mengelompokkan pelanggan menjadi tiga cluster, yaitu Cluster 1 (Prioritas) sebanyak 1 pelanggan, Cluster 2 (VVIP) sebanyak 2 pelanggan, dan Cluster 3 (VIP) sebanyak 7 pelanggan. Pelanggan dengan nilai recency yang rendah dan frekuensi tinggi berada dalam cluster Prioritas, sementara pelanggan dengan recency tinggi serta frekuensi rendah tergolong pada cluster VIP dan VVIP.

KESIMPULAN

Penelitian ini dapat disimpulkan seperti berikut:

1. Penerapan metode K-Means Clustering berjalan dengan lancar sehingga mampu membedakan pelanggan ke dalam tiga kategori, yaitu Prioritas, VVIP, dan VIP.
2. Hasil evaluasi menggunakan *Silhouette Score* menunjukkan nilai sebesar 0.5884 untuk kategori KAI, 0.5101 untuk End User, dan 0.6274 untuk Singres Member, yang menandakan bahwa kualitas pengelompokan berada pada tingkat cukup baik.

Saran

Terdapat beberapa saran dari peneliti, seperti berikut:

1. Pada penelitian selanjutnya, disarankan menggunakan data dengan periode yang lebih panjang serta jumlah data yang lebih besar. Hal ini bertujuan agar hasil segmentasi pelanggan dapat mewakili seluruh populasi secara lebih akurat dan tidak terbatas pada sebagian kecil data saja.
2. Diharapkan penelitian berikutnya dapat mengombinasikan metode K-Means dengan algoritma lain.

DAFTAR PUSTAKA

- Dinanti, N., Lestanti, S., & Budiman, S. N. (2023). *Penerapan Algoritma K-Means Untuk Pengelompokan Tindak Kejahatan Di Wilayah Hukum Polres Blitar Kota*. Jati (Jurnal Mahasiswa Teknik Informatika), 7(5), 3770-3777.
- Prayoga, Y., Mahmudi, A., & Zahro, H. Z. (2021). *Penerapan Metode K Means Pada Sistem Informasi Akademik Untuk Pengelompokan Siswa Berprestasi Di Upt Sma Negeri 3 Kota Pasuruan Berbasis Web*. Jati (Jurnal Mahasiswa Teknik Informatika), 5(2), 822-828.
- Febrianto, A., Achmadi, S., & Sasmito, A. P. (2021). *Implementing K-Means Method For Visitor Clustering Library Itn Malang*. Jati (Jurnal Mahasiswa Teknik Informatika), 5(1), 61-70.
- Prasetya, Y. Y., Faisol, A., & Vendyansah, N. (2021). *Sistem Informasi Geografis Hasil Produksi Padi Di Kabupaten Malang Menggunakan Metode K-Means Clustering*. Jati (Jurnal Mahasiswa Teknik Informatika), 5(2), 806-814.
- Andrew Lomaksan Manuel Tampubolon, T. M. (2024). *Segmentasi Pelanggan Majalah Pada Situs Web E-Commerce*. Teknologi Informasi Dan Ilmu Komputer (Jtiik), 11, 1243-1252.

- Perdana, S. A., Florentin, S. F., & Santoso, A. (2022). *Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift*. Sebatik, 26(2), 446-457.
- Hendrastuty, N. (2024). *Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa*. Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom), 3(1), 46-56.
- Sembiring, E. S., & Simatupang, T. A. (2023). *Segmentasi Pelanggan Berdasarkan Recency, Frequency, Monetary (Rfm) Untuk Meningkatkan Loyalitas Pelanggan*. Jurnal Sistem Informasi Bisnis, 10(1), 45–53.
- Prasetyo, E., & Ramdhani, M. A. (2022). *Analisis segmentasi pelanggan menggunakan model RFM (Recency, Frequency, Monetary) pada e-commerce*. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), 9(3), 120–129.