Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM (Studi Kasus: PT. AJM Global Pratama)

1st Filiana Flora Jenonica Pabutungan
Departemen Sistem Informasi
Telkom University
Surabaya, Indonesia
filianaflora@student.telkomuniversity.ac.id

2nd Sri Hidayati
Departemen Sistem Informasi
Telkom University
Surabaya, Indonesia
srihidayatisi@telkomuniversity.ac.id

3rd Tita Ayu Rospricilia
Departemen Sistem Informasi
Telkom University
Surabaya, Indonesia
titaayu@telkomuniversity.ac.id

***Abstrak* — Dalam era bisnis modern yang berkembang pesat, data telah menjadi elemen krusial bagi perusahaan untuk meningkatkan efisiensi dan mendukung pengambilan keputusan yang akurat. PT. AJM Global Pratama masih menghadapi tantangan dalam pengelolaan data transaksi pelanggan, karena sebagian besar proses masih dilakukan secara manual, sehingga menyulitkan perusahaan dalam menganalisis serta memahami pola pembelian pelanggan secara komprehensif. Penelitian ini bertujuan untuk menganalisis dan melakukan segmentasi pelanggan berdasarkan perilaku pembelian menggunakan metode K-Means *Clustering* dan model Recency, Frequency, dan Monetary (RFM). Data transaksi pelanggan diolah dengan mengacu pada metodologi CRISP-DM, yang meliputi tahapan pemahaman data, persiapan data, pemodelan, dan evaluasi. Hasil penelitian menunjukkan bahwa proses segmentasi berhasil mengidentifikasi tiga kelompok utama pelanggan, yaitu *High Value Customers*, Regular Customers, dan *Low Value Customers*, yang masing-masing memiliki karakteristik pola pembelian yang berbeda. Penerapan metode ini memungkinkan perusahaan memperoleh pemahaman yang lebih mendalam mengenai perilaku pelanggan serta mendukung pengelolaan hubungan pelanggan yang lebih efektif dan berbasis data di PT. AJM Global Pratama.**

***Kata kunci*— K-Means *Clustering*, RFM, CRISP-DM, Segmentasi Pelanggan, Streamlit.**

I. PENDAHULUAN

Dalam era bisnis modern yang berkembang dengan pesat, data telah menjadi aspek krusial bagi perusahaan. Semakin banyak organisasi, baik skala kecil maupun besar, menyadari bahwa pemanfaatan data merupakan alat penting untuk mendukung pengambilan keputusan yang lebih akurat, khususnya data transaksi pelanggan[1]. Oleh karena itu, pengolahan data yang tepat sangat diperlukan agar informasi yang tersedia dapat dimanfaatkan secara optimal. Kemampuan untuk mengelola dan memahami data yang besar dan beragam memberikan potensi manfaat yang signifikan[2]. Namun, pada praktiknya, masih banyak perusahaan yang mengandalkan metode manual dalam pencatatan dan pemahaman data transaksi pelanggan. Metode ini memiliki berbagai kelemahan, antara lain membutuhkan waktu dan tenaga yang besar, sehingga menjadi kurang efisien dan kurang akurat[3]. Kondisi tersebut menyebabkan kesulitan dalam memahami pola transaksi pelanggan secara menyeluruh.

PT. AJM Global Pratama merupakan perusahaan yang bergerak di bidang pengadaan barang industri, khususnya PVC Curtain. Perusahaan ini sebelumnya bernama CV. Agung Jaya Mandiri yang didirikan pada tahun 2012, kemudian pada tahun 2023 mengubah status badan hukum menjadi perseroan terbatas (PT) dengan tujuan memperluas jangkauan klien serta meningkatkan kredibilitas di mata mitra bisnis. Dalam kegiatan operasionalnya, perusahaan mengelola berbagai data bisnis, termasuk data transaksi pelanggan, yang hingga saat ini masih dicatat secara manual dan menggunakan *Microsoft Excel*. Metode pencatatan ganda ini diterapkan sebagai langkah antisipasi terhadap kejadian kehilangan data di masa lalu yang menyulitkan proses pemulihan data.

Berdasarkan hasil wawancara dengan staf *Accounting* dan *Finance*, Ibu Fida Laila menyatakan bahwa meskipun penggunaan Excel memungkinkan pencarian data yang lebih cepat dibandingkan metode manual, perusahaan masih menghadapi berbagai kendala dalam pengelolaan data. Data yang tersimpan dalam *Excel* hanya berupa ringkasan total pembelian per perusahaan tanpa rincian transaksi yang lengkap, sehingga proses analisis secara otomatis menjadi sulit dan memerlukan waktu yang cukup lama. Oleh karena itu, PT. AJM Global Pratama berencana mengembangkan sistem yang lebih modern guna meningkatkan daya saing serta mengikuti perkembangan industri dalam pengelolaan data.

Penelitian ini menggunakan metode *Clustering* dengan *model Recency, Frequency, and Monetary* (RFM). *Clustering* merupakan salah satu teknik dalam data mining yang telah terbukti efektif dalam menyelesaikan permasalahan kompleks di bidang ilmu komputer dan statistika[3]. Algoritma *K-Means Clustering* digunakan untuk mengelompokkan data ke dalam satu atau beberapa klaster berdasarkan atribut tertentu, di mana objek data dalam satu klaster memiliki tingkat kemiripan yang tinggi, sedangkan objek data antar klaster memiliki tingkat kemiripan yang rendah[4]. Sementara itu, model RFM merupakan pendekatan analisis yang digunakan untuk melakukan segmentasi pelanggan berdasarkan perilaku transaksi. Model

ini terdiri dari tiga komponen utama, yaitu *Recency (R*) yang menunjukkan seberapa baru pelanggan melakukan transaksi terakhir, *Frequency (F)* yang mengukur seberapa sering pelanggan melakukan transaksi dalam periode tertentu, dan *Monetary (M)* yang merepresentasikan total nilai transaksi yang dilakukan pelanggan. Efektivitas model ini sangat bergantung pada kualitas dan kelengkapan data transaksi yang tersedia[3].

Berdasarkan penelitian terdahulu mengenai penerapan *data mining* dengan metode *Clustering* untuk analisis segmentasi pelanggan, pendekatan ini terbukti mampu menyelesaikan permasalahan pada usaha kecil dan menengah (UKM). Permasalahan tersebut umumnya muncul karena program pendampingan UKM sering kali diterapkan tanpa mempertimbangkan karakteristik masing-masing usaha, sehingga hasil yang diperoleh kurang optimal. Oleh karena itu, pendekatan berbasis data seperti segmentasi pelanggan menggunakan model RFM dan metode *K-Means Clustering* diperlukan untuk membantu lembaga pendukung dalam merancang strategi pendampingan yang lebih tepat sasaran dan relevan[5]. Selain itu, penelitian yang dilakukan pada PT. SID, sebuah perusahaan B2B, menunjukkan adanya penurunan pendapatan yang mendorong perlunya pengelompokan pelanggan. Permasalahan tersebut berhasil diatasi melalui penerapan metode *K-Means Clustering* dan model RFM[6].

Berdasarkan penelitian terdahulu serta permasalahan yang dihadapi oleh PT. AJM Global Pratama, penelitian ini menerapkan segmentasi pelanggan menggunakan metode *K-Means Clustering* dengan model *Recency, Frequency, and Monetary* (RFM) untuk membantu perusahaan dalam mengelompokkan pelanggan berdasarkan pola transaksi. Selain itu, dikembangkan pula sebuah dashboard visualisasi menggunakan *Python Streamlit* untuk mempermudah interpretasi data. Dengan demikian, penelitian ini bertujuan untuk melakukan proses *clustering* dan menganalisis hasil segmentasi pelanggan guna mendukung PT. AJM Global Pratama dalam pengelolaan data transaksi pelanggan, serta memberikan rekomendasi strategi berdasarkan visualisasi data dalam bentuk *dashboard streamlit*.

II. KAJIAN TEORI

A. *Cross Industry Standard for Data Mining (CRISP-DM)*

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan salah satu pendekatan yang paling banyak digunakan oleh para profesional di bidang pemodelan data. Tujuan utama dari metode ini adalah untuk menemukan pola-pola yang bermakna dan bernilai dari data yang dianalisis[7]. CRISP-DM menyediakan kerangka kerja yang sistematis serta urutan tahapan yang jelas, sehingga memudahkan peneliti dalam mengikuti alur kerja yang terstruktur pada aktivitas data mining[4]. Metode ini terdiri atas enam fase utama. Adapun tahapan yang diterapkan dalam penelitian ini dijelaskan sebagai berikut. Fase pertama adalah Business Understanding, yang bertujuan untuk memahami permasalahan yang dihadapi oleh PT. AJM Global Pratama, yaitu pengelolaan data pelanggan dan strategi pemasaran yang belum optimal. Penelitian ini berfokus pada penerapan *Customer Relationship Management* (CRM) menggunakan model *Recency, Frequency, and Monetary (RFM)* dan metode *K-Means Clustering* untuk menghasilkan segmentasi pelanggan yang mendukung pengambilan keputusan berbasis data. Fase Data Understanding dilakukan untuk memahami struktur dan karakteristik data transaksi pelanggan yang diperoleh dari PT. AJM Global Pratama. Dataset yang digunakan mencakup informasi seperti nomor faktur, tanggal transaksi, nama perusahaan, produk, jumlah pembelian, dan total nilai transaksi.

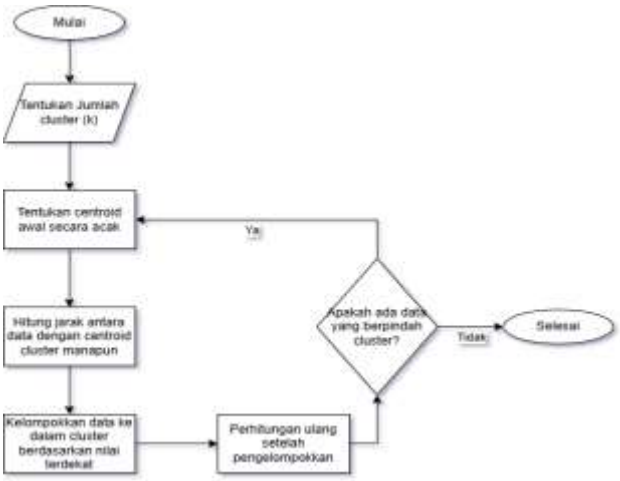
TABEL 1
(VARIABEL DATASET)

Variable	Description	Data Type
Nomor Faktur	Nomor unik transaksi pelanggan	Kategorik
Tanggal Transaksi	Waktu terjadinya transasksi	Numerik
Nama Perusahaan	Nama perusahaan pelanggan	Kategorik
Barang	Jenis produk yang dibeli	Kategorik
Kuantitas	Jumlah produk yang dibeli	Numerik
Jumlah Transaksi	Total nilai uang yang dikeluarkan pelanggan	Numerik

Tahap Data Preparation mencakup proses pembersihan data, penghapusan data duplikat, penanganan *outlier*, serta perhitungan nilai RFM *(Recency, Frequency, dan Monetary)* untuk setiap pelanggan berdasarkan data transaksi yang tersedia. Pada tahap *Modeling*, pelanggan dikelompokkan berdasarkan nilai RFM menggunakan metode *K-Means Clustering*. Penentuan jumlah *cluster* optimal dilakukan dengan menggunakan *Elbow Method* melalui identifikasi titik perubahan signifikan pada nilai *Sum of Squared Error (SSE).* Tahap Evaluation bertujuan untuk menilai kualitas hasil pengelompokan yang diperoleh. Evaluasi dilakukan menggunakan *Silhouette Coefficient*, yang digunakan untuk mengukur tingkat kohesi dalam *cluster* dan tingkat separasi antar*cluster*, di mana nilai yang semakin mendekati 1 menunjukkan kinerja *clustering* yang semakin baik. Tahap akhir adalah *Deployment*, yaitu implementasi hasil segmentasi pelanggan dalam bentuk dashboard interaktif berbasis *Streamlit*. Dashboard ini menyajikan hasil *clustering*, ringkasan statistik pada setiap *cluster*, serta visualisasi grafis yang membantu PT. AJM Global Pratama dalam memahami karakteristik pelanggan dan mendukung pengambilan keputusan bisnis yang lebih efektif.

B. *K-Means Clustering*

K-Means Clustering merupakan metode pengelompokan data non-hierarkis yang bertujuan untuk membagi data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan karakteristik antarobjek. Setiap data akan dialokasikan ke dalam *cluster* dengan centroid atau titik pusat terdekat, sehingga objek-objek yang berada dalam *cluster* yang sama memiliki tingkat kemiripan yang tinggi dan perbedaan yang jelas dibandingkan dengan objek pada *cluster* lainnya[8]. Gambar 1 menunjukkan tahapan-tahapan dalam algoritma *K-Means Clustering*.



GAMBAR 1
(ALGORITMA K-MEANS)

C. *Recency, Frequency, Monetary (RFM)*

Model Recency, Frequency, dan *Monetary (RFM)* merupakan salah satu pendekatan yang banyak digunakan di berbagai bidang, khususnya dalam pemasaran. Model ini memungkinkan pengambil keputusan untuk mengevaluasi nilai pelanggan berdasarkan tiga indikator utama, yaitu seberapa baru pelanggan melakukan transaksi *(Recency),* seberapa sering pelanggan melakukan transaksi *(Frequency),* dan total nilai transaksi yang dilakukan pelanggan *(Monetary*). Dengan memanfaatkan data transaksi, analisis RFM mampu memberikan pemahaman yang komprehensif mengenai pola pembelian pelanggan[9]. Definisi RFM yang digunakan dalam penelitian ini adalah sebagai berikut:

- *Recent(R)* merepresentasikan selisih waktu antara tanggal transaksi terakhir pelanggan dengan tanggal analisis. Nilai *Recency* yang lebih kecil menunjukkan bahwa pelanggan melakukan transaksi dalam waktu yang lebih dekat, sehingga mencerminkan tingkat aktivitas yang lebih tinggi. Sebagai contoh, apabila tanggal analisis adalah 22 Oktober 2024 dan pelanggan P133 terakhir melakukan transaksi pada 31 Januari 2024, maka selisih waktunya adalah 264 hari, yang menunjukkan bahwa pelanggan tersebut tidak melakukan pembelian dalam jangka waktu yang cukup lama.
- *Frequency(F)* menggambarkan seberapa sering pelanggan melakukan transaksi dalam periode tertentu. Nilai *Frequency* yang lebih tinggi menunjukkan tingkat aktivitas pelanggan yang lebih tinggi. Sebagai contoh, apabila pelanggan P131 melakukan tiga kali transaksi pada bulan Maret 2024, maka nilai *Frequency* pelanggan tersebut adalah 3, yang menandakan bahwa perusahaan tersebut relatif aktif dalam melakukan pembelian.
- *Monetary(M)* merepresentasikan total nilai pembelian yang dilakukan oleh pelanggan selama periode analisis. Nilai *Monetary* yang lebih tinggi menunjukkan kontribusi finansial pelanggan yang lebih besar terhadap total penjualan perusahaan. Sebagai contoh, apabila pelanggan P131 melakukan pembelian *Silicon Curtain* Kuning dengan nilai Rp10.000.000, maka nilai Monetary pelanggan tersebut adalah Rp10.000.000.

Berdasarkan hasil perhitungan, sebanyak 131 perusahaan teridentifikasi telah melakukan transaksi selama periode analisis. Contoh hasil perhitungan nilai RFM disajikan pada Tabel 2 berikut.

TABLE 2
(MODEL RECENCY, FREQUENCY, MONETARY)

No.	Company	R	F	M
1	P01	38	6	65744550
2	P02	392	2	4748580
3	P03	3	20	443983726
4	P04	208	2	129714600
…	…	…	…	…
131	P131	68	81	211079041

Tabel tersebut menggambarkan variasi perilaku pelanggan berdasarkan nilai RFM. Sebagai contoh, P131` memiliki nilai *Recency* sebesar 68 yang menunjukkan transaksi yang relatif baru, nilai *Frequency* yang tinggi yaitu sebanyak 81 transaksi, serta nilai *Monetary* yang besar sebesar Rp211.079.041. Hal ini menunjukkan bahwa perusahaan tersebut merupakan pelanggan yang aktif dan bernilai tinggi bagi perusahaan.

D. *Isolation Forest Algorithm*

Isolation Forest Algorithm merupakan metode pendeteksi anomali atau outlier yang bekerja berdasarkan premis bahwa data yang menyimpang biasanya muncul lebih jarang dan berada jauh dari pola mayoritas. *Outlier* yang tidak ditangani secara memadai dapat menurunkan kualitas analisis sehingga menimbulkan bias serta kesalahan dalam pengambilan kesimpulan statistik[10]. Cara kerja metode ini terdiri dari dua tahap utama, yaitu tahap pelatihan untuk membangun model *iForest* dan tahap evaluasi untuk menentukan skor anomali. Pada tahap pelatihan, algoritma menghasilkan sejumlah *isolation tree* sebanyak T pohon, dimana nilai T ditentukan oleh *hyperparameter n_trees*. Jika nilai skornya sama dengan atau melebihi batas *threshold* yang telah ditentukan, maka data tersebut diklasifikasikan sebagai anomali, jika tidak data dianggap normal. Titik anomali diberi label dengan skor -1, sedangkan titik normal diberi skor 1[11].

E. *Silhouette Coefficient*

Metode *Silhouette Coefficient* digunakan untuk mengevaluasi kualitas hasil *clustering* dengan menggabungkan dua aspek utama, yaitu *cohesion* dan *separation*. *Cohesion* mengukur tingkat kedekatan antar data dalam satu *cluster*, sedangkan separation mengukur jarak antar *cluster* yang berbeda. Jarak antar data dihitung menggunakan *Euclidean Distance*.

Nilai *silhouette* memberikan indikasi seberapa baik setiap data berada di dalam *cluster* yang telah ditetapkan. Nilai yang mendekati 1 menunjukkan bahwa data terkelompok dengan baik, nilai mendekati 0 menunjukkan bahwa data berada di sekitar batas antar *cluster*, sedangkan nilai negatif mengindikasikan bahwa data tersebut kemungkinan ditempatkan pada *cluster* yang tidak sesuai[12]. Perhitungan

Silhouette Coefficient dilakukan melalui beberapa tahapan sebagai berikut[13]:

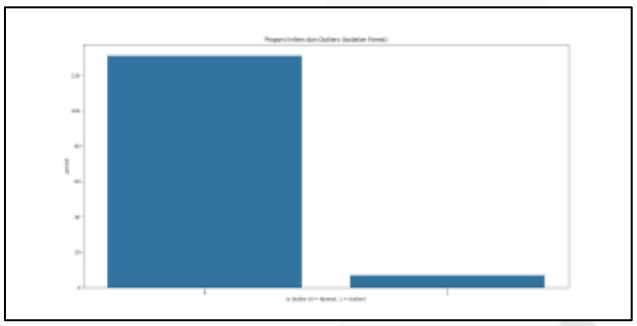
- Menghitung rata-rata jarak antara data ke-*i* dengan seluruh data lain dalam *cluster* yang sama, yang dinotasikan sebagai *a(i)*.
- Menghitung rata-rata jarak antara data ke-*i* dengan seluruh data pada *cluster* terdekat, yang dinotasikan sebagai *b(i)*.
- Menghitung nilai *Silhouette Coefficient* menggunakan persamaan berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)} \qquad (1)$$

Kualitas *clustering* secara keseluruhan ditentukan berdasarkan nilai rata-rata silhouette dari seluruh data yang digunakan dalam proses *clustering*.

III. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini berupa data transaksi pelanggan PT. AJM Global Pratama yang mencakup periode Agustus 2023 hingga Juni 2025. Proses diawali dengan penghapusan data duplikat, pencarian outlier, penyesuaian format data, serta proses normalisasi sebelum data ditransformasikan ke dalam format RFM. Setelah itu dilakukan pendeteksian outlier menggunakan *isolation forest algorithm*, sehingga mendapatkan hasil sebagai berikut:

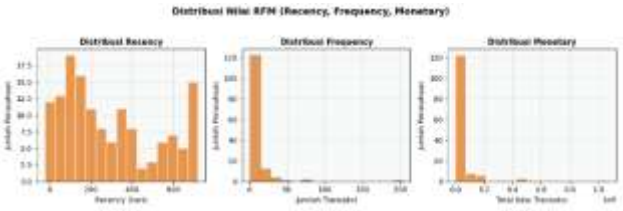


GAMBAR 3

(SCATTER PLOT DETEKSI OUTLIER)

Pada gambar 3 yaitu *Bar Chart* yang menunjukkan proporsi inliers dan *outliers* yang terdeteksi menggunakan algoritma *Isolation Forest*. Algoritma diinisialisasi dengan parameter kontaminasi 0.05, yang berarti ekspetasi bahwa sekitar 5% data merupakan anomali. Parameter ini dipilih untuk menjaga keseimbangan antara kualitas data dan menghindari penghapusan data yang berlebihan. Dari visualisasi, terlihat bahwa sebagian besar data sekitar 95% atau sebanyak 131 pelanggan termasuk *inliers* (0). Sementara hanya sebagian kecil sejumlah 5% atau sebanyak 7 pelanggan yang dikategorikan sebagai *outliers* (1).

Setelah dilakukan normalisasi seperti data tanggal diubah ke dalam format YYYY-MM-DD, sedangkan nilai transaksi dikonversi ke dalam tipe numerik guna menghindari kesalahan perhitungan., dataset disusun ke dalam metrik *Recency, Frequency*, dan *Monetary* (RFM) untuk merepresentasikan perilaku pelanggan berdasarkan waktu transaksi terakhir, frekuensi pembelian, serta total nilai pengeluaran. Tahapan ini bertujuan untuk menghasilkan data yang bersih, konsisten, dan terstruktur dengan baik sehingga layak digunakan dalam analisis segmentasi pelanggan berbasis RFM.



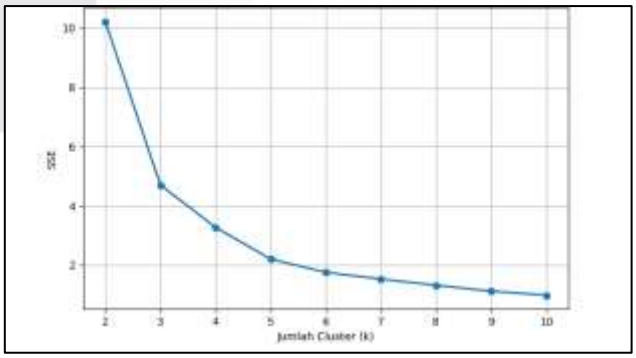
GAMBAR 2

(DISTRIBUSI DATA RFM)

Gambar 2 menunjukkan distribusi data pada tiga dimensi RFM, yaitu *Recency, Frequency, dan Monetary*, dalam skala yang telah distandarisasi. Sebagian besar pelanggan berada pada rentang *Recency* yang rendah (0–200 hari), yang mengindikasikan adanya transaksi yang relatif baru dan aktivitas pelanggan yang masih aktif. Namun demikian, terdapat sebagian kecil pelanggan dengan nilai *Recency* yang tinggi (sekitar 700 hari) yang merepresentasikan pelanggan tidak aktif. Distribusi *Frequency* dan *Monetary* menunjukkan kemencengan positif (*positive skewness*) yang kuat, di mana sebagian besar pelanggan (sekitar 115 pelanggan) memiliki jumlah transaksi yang rendah dan nilai pengeluaran yang kecil. Hal ini mengindikasikan bahwa meskipun sebagian besar pelanggan tergolong masih aktif, mayoritas pelanggan melakukan transaksi dengan frekuensi yang rendah dan memberikan kontribusi nilai yang minimal. Dengan demikian, dataset didominasi oleh pelanggan dengan tingkat keterlibatan dan nilai ekonomi yang rendah.

A. *Modelling*

Sebelum memasuki tahap pemodelan, penentuan jumlah *cluster* yang tepat merupakan hal yang sangat penting untuk memastikan hasil segmentasi yang optimal dan merepresentasikan struktur data dengan baik. Berdasarkan hasil analisis menggunakan metode tersebut, diperoleh jumlah *cluster* terbaik yang disajikan pada Gambar 3 sebagai dasar dalam proses *clustering* selanjutnya.

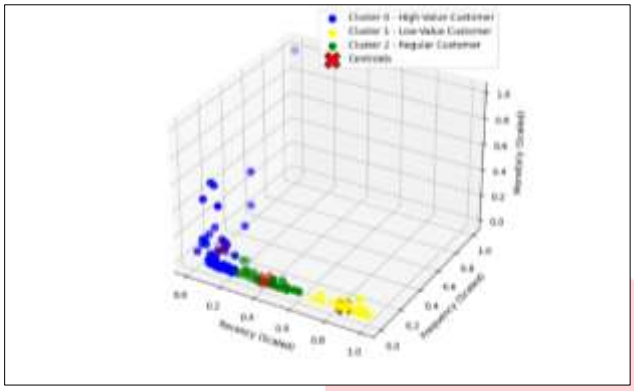


GAMBAR 4

(GRAFIK ELBOW METHOD)

Grafik *Elbow Method* menunjukkan hubungan antara jumlah *cluster* (k) dan nilai *Sum of Squared Errors (SSE*) pada algoritma *K-Means*. Nilai SSE mengalami penurunan yang

signifikan hingga k = 3, kemudian laju penurunannya melambat dan membentuk pola siku (elbow). Pola ini menunjukkan bahwa penambahan jumlah *cluster* setelah titik tersebut tidak memberikan peningkatan yang signifikan terhadap kualitas pengelompokan. Oleh karena itu, jumlah *cluster* optimal ditentukan sebanyak k = 3.



GAMBAR 5
(SCATTER PLOT 3D)

Setelah jumlah *cluster* optimal ditetapkan, algoritma K-Means diterapkan pada dataset RFM (Recency, Frequency, dan Monetary) yang telah dinormalisasi. Proses ini menghasilkan tiga *cluster* pelanggan dengan karakteristik perilaku transaksi yang relatif serupa. Segmentasi ini bertujuan untuk mengidentifikasi pola perilaku pelanggan yang dapat mendukung penyusunan strategi hubungan pelanggan *(customer relationship)* serta meningkatkan efektivitas pengambilan keputusan perusahaan.

TABLE 3
(DISTRIBUSI RFM TIAP *CLUSTER)*

Cluster	R	F	M	nSize(%)	RFM Score
0	0.12	0.070	0.091	44.2	↓R↑F↑M
1	0.86	0.003	0.014	23.91	↑R↓F↓M
2	0.41	0.007	0.017	31.88	R|F|M

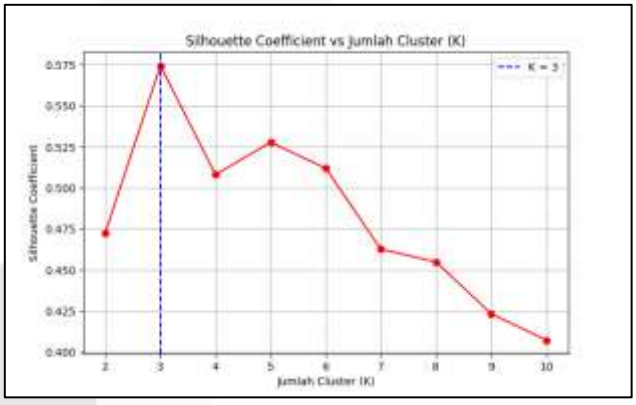
Berikut ini merupakan deskripsi dari masing-masing *cluster* berdasarkan tabel di atas.

- *Cluster* 0 *(High Value Customers*), *Cluster* ini memiliki nilai Recency yang rendah (0,12), Frequency yang tinggi (0,070), dan Monetary yang tinggi (0,091). *Cluster* ini terdiri dari 61 perusahaan atau sebesar 44,2% dari total populasi. Pola ↓R↑F↑M menunjukkan bahwa perusahaan dalam *cluster* ini merupakan pelanggan yang aktif, sering melakukan transaksi, serta memberikan kontribusi nilai transaksi yang tinggi. *Cluster* ini merepresentasikan pelanggan paling bernilai yang memberikan kontribusi pendapatan signifikan bagi perusahaan.
- *Cluster* 1 (*Low Value Customers*), *Cluster* ini menunjukkan nilai Recency yang tinggi (0,86), Frequency yang rendah (0,003), dan Monetary yang rendah (0,014). *Cluster* ini terdiri dari 33 perusahaan atau sebesar 23,91% dari total populasi. Pola ↑R↓F↓M mengindikasikan bahwa perusahaan dalam *cluster* ini jarang melakukan transaksi dan hanya memberikan kontribusi nilai transaksi yang minimal.
- *Cluster* 2 (*Regular Customers*), *Cluster* ini memiliki nilai Recency yang sedang (0,41), Frequency yang rendah (0,007), dan Monetary yang rendah (0,017). *Cluster* ini terdiri dari 44 perusahaan atau sebesar 31,88% dari total populasi. Pola R|F|M menunjukkan bahwa perusahaan dalam *cluster* ini melakukan transaksi secara relatif konsisten, namun dengan kontribusi nilai transaksi yang sedang jika dibandingkan dengan *cluster High Value Customers*.

Analisis distribusi RFM pada setiap *cluster* memberikan pemahaman yang lebih mendalam mengenai pola perilaku pembelian pelanggan perusahaan. Informasi ini dapat dijadikan sebagai dasar dalam pengambilan keputusan yang lebih tepat terkait strategi transaksi dan pengelolaan hubungan pelanggan.

B. *Evaluation*

Evaluasi hasil *clustering* dilakukan secara objektif menggunakan *Silhouette Coefficient* untuk mengukur kualitas pemisahan dan kekompakan antar*cluster*.



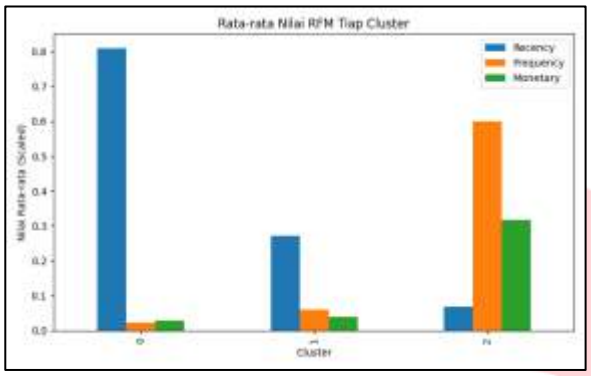
GAMBAR 6
(GRAFIK EVALUASI SILHOUTTE COEFFICIENT)

Berdasarkan gambar 6 diatas Berdasarkan hasil analisis, jumlah *cluster* yang dipilih menunjukkan kualitas pemisahan dan kekompakan data yang optimal. Nilai Silhouette yang diperoleh sebesar 0,5744 mengindikasikan bahwa struktur *cluster* yang terbentuk cukup baik, dengan tingkat homogenitas yang tinggi di dalam setiap *cluster* serta pemisahan yang jelas antar*cluster*. Hasil ini juga sejalan dengan temuan dari metode *Elbow*, sehingga jumlah *cluster* yang digunakan dinilai paling representatif untuk menggambarkan pola segmentasi pelanggan. Dengan struktur *cluster* yang terbentuk secara optimal, hasil pengelompokan ini dapat memberikan gambaran yang lebih jelas mengenai karakteristik pelanggan dan menjadi dasar

yang kuat dalam penyusunan strategi analisis dan pengambilan keputusan selanjutnya.

C. Identifikasi *Cluster* pelanggan Hasil Segmentasi RFM

Bagian ini menyajikan identifikasi *cluster* hasil segmentasi RFM untuk pola pembelian pelanggan berdasarkan nilai Recency, Frequency, dan Monetary beserta rekomendasi strategi bagi perusahaan.



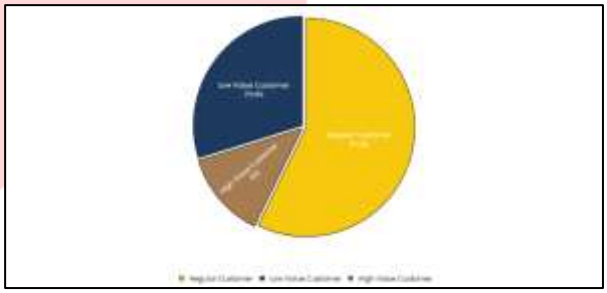
GAMBAR 7

(HISTOGRAM RATA-RATA TIAP *CLUSTER)*

- *Cluster* 0 (*Low Value Customer*) memiliki nilai Recency tinggi serta nilai Frequency dan Monetary yang rendah, dengan jumlah 39 perusahaan atau sekitar 29,77% dari total populasi. Pola RFM ↑R↓F↓M menunjukkan bahwa pelanggan dalam *cluster* ini sudah lama tidak melakukan transaksi dan memiliki frekuensi serta nilai pembelian yang rendah. Kondisi ini mengindikasikan rendahnya kontribusi nilai serta potensi perpindahan ke penyedia lain. Strategi yang direkomendasikan meliputi promosi harga, komunikasi pemasaran yang lebih intensif, serta penawaran produk dengan harga yang lebih kompetitif.
- *Cluster* 1 (*Regular Customer*) memiliki nilai Recency, Frequency, dan Monetary pada tingkat menengah dan terdiri dari 75 perusahaan atau sekitar 57,25% dari total populasi. Pola RFM yang relatif stabil menunjukkan bahwa pelanggan dalam *cluster* ini melakukan pembelian secara cukup konsisten, meskipun dengan nilai transaksi yang tidak terlalu besar. Jarak waktu pembelian ulang cenderung lebih panjang karena usia pakai produk yang relatif lama. Strategi yang disarankan meliputi pemeliharaan hubungan pelanggan melalui program pelanggan tetap, pemberian insentif pembelian berulang, serta penerapan upselling dan cross-selling untuk meningkatkan nilai transaksi.
- *Cluster* 2 (*High Value Customer*) memiliki nilai Recency rendah serta nilai Frequency dan Monetary yang tinggi, dengan jumlah 17 perusahaan atau sekitar 12% dari total populasi. Pola RFM ↓R↑F↑M menunjukkan bahwa pelanggan dalam *cluster* ini sering melakukan transaksi dengan nilai pembelian yang besar dan memberikan kontribusi signifikan terhadap pendapatan perusahaan. Oleh karena itu, strategi yang direkomendasikan berfokus pada mempertahankan loyalitas pelanggan melalui program loyalitas eksklusif, layanan premium, serta penguatan hubungan jangka panjang melalui kemitraan strategis.

D. *Deployment*

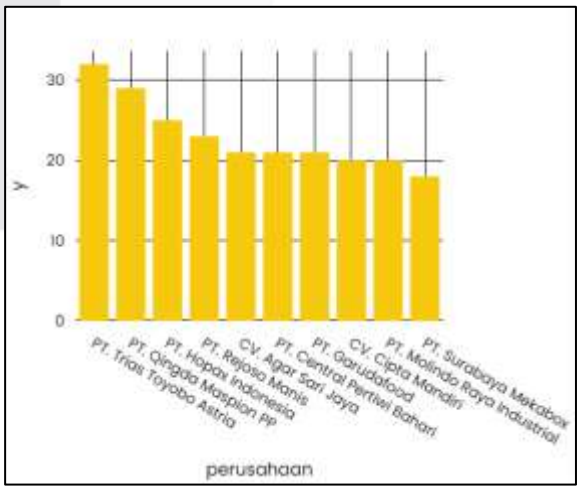
Setelah diperoleh hasil *clustering*, dilakukan visualisasi data untuk mempermudah pembacaan dan pemahaman terhadap pola data. Visualisasi data pada penelitian ini dikembangkan menggunakan Streamlit. Visualisasi *cluster* yang dihasilkan memungkinkan pengguna untuk mengamati pola dan hubungan antar data secara intuitif. Melalui dashboard ini, perusahaan dapat dengan mudah mengeksplorasi pola pembelian pelanggan pada setiap *cluster*.



GAMBAR 8

(PIE CHART DISTRIBUSI *CLUSTER)*

Visualisasi ini menampilkan perbandingan proporsi antara *Cluster* 0 (*Low Value Customers*), *Cluster* 1 (*Regular Customers*), dan *Cluster* 2 *(High Value Customer*), sehingga perusahaan dapat memahami komposisi masing-masing segmen pelanggan secara lebih jelas. Berdasarkan hasil visualisasi, *Cluster* 0 memiliki proporsi sebanyak 29.8% dengan jumlah 39 pelanggan, kemudian *Cluster* 1 sebanyak 57.25% dengan jumlah pelanggan yaitu 75, dan terkahir yaitu *Cluster* 2 sebagai kelompok dengan persentasi terkecil yaitu 13% dengan jumlah yaitu 17 pelanggan.

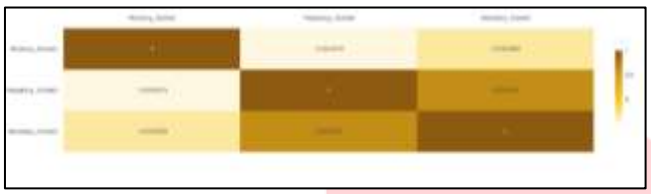


GAMBAR 9

(BAR CHART TOP 10 FREQUENCY)

Visualisasi *Bar Chart* menunjukkan bahwa PT. Trias Toyobo Astria memiliki jumlah transaksi tertinggi dengan sekitar 32

transaksi, diikuti oleh PT. Qingda Maspion PP dengan sekitar 29 transaksi, yang menandakan keduanya sebagai pelanggan dengan frekuensi transaksi paling tinggi. Pada kelompok menengah, PT. Hopax Indonesia dan PT. Rejoso Manis masing-masing mencatat sekitar 25 dan 23 transaksi, sementara beberapa perusahaan lain seperti PT. Garudafood, PT. Agar Sari Jaya, dan PT. Central Pertiwi Bahari menunjukkan jumlah transaksi yang relatif serupa. Perusahaan dengan jumlah transaksi terendah dalam visualisasi ini adalah PT. Surabaya Mekabox dengan sekitar 19 transaksi.



GAMBAR 10
(HEATMAP KORELASI RFM)

Heatmap korelasi menunjukkan hubungan antar variabel RFM, di mana terdapat korelasi positif yang kuat antara Frequency dan Monetary, menandakan bahwa pelanggan dengan frekuensi pembelian tinggi cenderung memiliki nilai belanja yang lebih besar. Sebaliknya, *Recency* menunjukkan korelasi negatif sedang terhadap *Frequency* dan *Monetary*, yang mengindikasikan bahwa pelanggan yang melakukan pembelian lebih baru cenderung lebih sering bertransaksi dan memiliki nilai pembelian yang lebih tinggi. Temuan ini sejalan dengan konsep dasar analisis RFM.

IV. KESIMPULAN

Segmentasi pelanggan PT. AJM Global Pratama dilakukan menggunakan metode K-Means berbasis model RFM terhadap data transaksi periode Agustus 2023 hingga Juni 2025. Dari 1.106 data transaksi, diperoleh 131 data pelanggan setelah proses pembersihan data dan penghapusan outlier menggunakan *Isolation Forest*. Hasil penentuan jumlah *cluster* optimal melalui metode *Elbow* menunjukkan k = 3, dengan nilai *Silhouette Score* sebesar 0,5744. Segmentasi menghasilkan tiga kelompok pelanggan, yaitu *Low Value Customer (29,77%), Regular Customer (57,25%), dan High Value Customer (12,98%).*

Hasil segmentasi ini menjadi dasar dalam penyusunan strategi bisnis yang lebih tepat sasaran, di mana pelanggan bernilai rendah difokuskan pada peningkatan transaksi, pelanggan reguler pada pemeliharaan hubungan dan peningkatan nilai pembelian, serta pelanggan bernilai tinggi pada strategi retensi melalui layanan dan program loyalitas. Seluruh hasil analisis disajikan melalui dashboard Streamlit untuk mendukung pengambilan keputusan berbasis data.

REFERENSI

[1] A. K. Virgiawan, E. D. Absharina, and Fenando, "Peran Big Data Dalam Meningkatkan Daya Saing Bisnis Di Era Digital," *J. Sist. Inf. dan Tek. Komput.*, vol. 10, no. 1, 2025.

[2] R. E. Damanik and P. I. M. Nasution, "Evaluasi Penggunaan Teknologi Big Data Untuk Analisis Data Dan Pengambilan Keputusan," *J. Multidisiplin Saintek*, vol. 3, no. 10, pp. 1–23, 2024.

[3] Sujono and I. L. Sari, "Perancangan Sistem Informasi Akademik Pada SDIT Al-Qudwah Pangkalpinang Menggunakan Metode FAST," *J. IT CIDA*, vol. 9, no. 1, 2023.

[4] S. Santiastry, Asriyanik, and W. Apriandari, "Penerapan Algoritma Naive Bayes Dan Metode CRISP-DM Dalam Prediksi Hasil Tes Kemampuan Bahasa Inggris Mahasiswa," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 10432–10439, 2024.

[5] S. F. Djun, I. G. A. Gunadi, and S. Sariyasa, "Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means *Clustering* pada Model Data RFM," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 5, no. 4, pp. 354–364, 2024, doi: 10.35746/jtim.v5i4.434.

[6] M. F. Fadhillah, A. Lovely, A. Suyoso, and I. Puspitasari, "Customer Segmentation with *Clustering* Algorithm Based on Recency, Frequency, and Monetary (RFM) Attributes," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 1, pp. 48–56, 2025, doi: 10.57152/malcom.v5v1.1491.

[7] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means *Clustering* untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.

[8] R. A. Indraputra and R. Fitriana, "K-Means *Clustering* Data COVID-19," *J. Tek. Ind.*, vol. 10, no. 3, pp. 275–282, 2020, doi: 10.25105/jti.v10i3.8428.

[9] D. David, M. D. Lauro, and D. E. Herwindiati, "Sistem Prediksi Customer Loyalty Dengan Metode RFM dan Fuzzy C-Means," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 4, no. 1, p. 33, 2020, doi: 10.24912/computatio.v4i1.7099.

[10] J. Elektronika and D. A. N. Komputer, "Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI," vol. 17, no. 2, pp. 657–666, 2024.

[11] R. Y. Pramudita, B. D. Setiawan, and M. Data, "Kinerja Deteksi Tautan Website Phishing Menggunakan Isolation Forest," vol. 9, no. 10, pp. 1–9, 2025.

[12] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi *Clustering* Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.

[13] C. Sukmayadi, A. Primajaya, and I. Maulana, "Penerapan Algoritma K-Medoids dalam Menentukan Daerah Rawan Banjir di Kabupaten Karawang," *INFORMAL Informatics J.*, vol. 6, no. 3, p. 187, 2021, doi: 10.19184/isj.v6i3.25423.