

IMPLEMENTASI K-MEANS CLUSTERING DALAM SEGMENTASI PELANGGAN BERDASARKAN USIA, PENDAPATAN, DAN MODEL RFM (STUDI KASUS: LANTIKYA STORE JOMBANG)

Beni Tiyas Kristanti^{1*}, Achmad Junaidi², Eka Prakarsa Mandyartha³

^{1,2,3}Universitas Pembangunan Nasional “Veteran” Jawa Timur, Jl. Rungkut Madya No.1, Gn. Anyar, Surabaya

Received: 28 Juni 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Segmentasi Pelanggan;

K-Means;

RFM;

AHP;

AHP.

Correspondent Email:

tiyaskristanti12@gmail.com

ABSTRAK. Lantikya Store Jombang merupakan salah satu toko retail di Kabupaten Jombang yang belum menerapkan strategi pemasaran yang memberikan pelayanan yang berbeda berdasarkan karakteristik dan tipe pelanggan. Segmentasi pelanggan adalah proses membagi pelanggan kedalam kelompok berdasarkan karakteristik atau perilaku yang berbeda untuk menunjukkan tingkat keragaman di antara pelanggan. Karakteristik dan perilaku pelanggan dilihat berdasarkan usia, pendapatan, *recency* (terakhir melakukan transaksi), *frequency* (jumlah kedatangan), dan *monetary* (jumlah uang yang dikeluarkan) atau disebut dengan RFM (*Recency, Frequency, dan Monetary*). Pengelompokan dilakukan dengan salah satu algoritma yaitu K-Means dengan jumlah data yaitu 1140 data. Pada penelitian ini menghasilkan 4 kelompok berdasarkan perbandingan hasil dari metode pencarian kelompok yaitu *elbow method*, *silhouette method*, dan *gap statistic*. Analisis yang dilakukan untuk menyusun strategi pemasaran dihitung berdasarkan nilai variabel dengan metode *Analytic Hierarchy Process* (AHP) dan *Customer Lifetime Value* (CLV). Karakteristik dan tipe pelanggan yang dihasilkan dari 4 kelompok pelanggan yaitu usia generasi milenial yang memiliki pendapatan tinggi dengan tipe pelanggan untuk retensi umum dan pelanggan bernilai penting, serta usia generasi Z yang memiliki pendapatan rendah dengan tipe pelanggan untuk pengembangan umum dan pelanggan yang hilang.

Abstract. Lantikya Store Jombang is a retail store located in Jombang Regency that has yet to implement a marketing strategy providing differentiated services based on customer characteristics and types. Customer segmentation is the process of dividing customers into groups based on different characteristics or behaviors to illustrate the diversity among customers. Customer characteristics and behaviors are assessed based on age, income, *recency* (last transaction), *frequency* (number of visits), and *monetary* (amount spent), also known as RFM (*Recency, Frequency, and Monetary*). The clustering was performed using the K-Means algorithm with a total of 1,140 data points. This study resulted in four clusters based on the comparison of results from the *elbow method*, *silhouette method*, and *gap statistic*. The analysis for developing marketing strategies was calculated based on variable values using the *Analytic Hierarchy Process* (AHP) and *Customer Lifetime Value* (CLV). The characteristics and types of customers resulting from the four clusters are millennials with high income categorized as customers for general retention and high-value customers, and Generation

Z with low income categorized as customers for general development and lost customers.

1. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat telah mengubah paradigma bisnis. Dalam konteks ini, data menjadi nilai atau aset penting bagi perusahaan dalam mendukung pengambilan keputusan yang akurat [1]. Data yang dikelola secara efektif mampu memberikan pandangan mendalam tentang tren pasar, perilaku konsumen, dan kinerja operasional perusahaan. Oleh karena itu, perusahaan yang mampu mengoptimalkan pemanfaatan data memiliki keunggulan kompetitif dalam mengambil keputusan yang baik dan berorientasi pada hasil [2].

Salah satu pendekatan yang dapat digunakan untuk memanfaatkan data secara lebih mendalam adalah dengan melakukan *data mining*. *Data mining* adalah proses penggalian dan pencarian pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar [3]. Proses yang dilakukan dalam data mining yaitu estimasi, prediksi, klasifikasi, klustering, dan asosiasi [4].

Data mining tidak hanya memungkinkan identifikasi pola dan informasi berharga dari data besar, tetapi juga mendukung pemahaman mendalam terhadap perilaku konsumen. Segmentasi merupakan proses membagi pasar potensial kedalam kelompok-kelompok konsumen yang berbeda dan memilih satu segmen atau lebih sebagai pasar target yang akan dicapai dengan marketing yang berbeda [5].

Segmentasi pelanggan merupakan salah satu faktor penting dalam peningkatan penjualan karena tingkat signifikansi segmentasi akan mempengaruhi sejauh mana peningkatan penjualan dapat dicapai [6].

Berdasarkan beberapa penelitian sebelumnya, K-Means dianggap sebagai metode yang cocok untuk melakukan segmentasi pelanggan. Penelitian yang dilakukan oleh Wardani menunjukkan metode K-Means memiliki kinerja yang optimal dalam segmentasi penjualan otomotif [7]. Hal tersebut dibuktikan dengan tingginya nilai Silhouette Coefficient mencapai 0.716 mendekati nilai 1,

dibandingkan metode DBSCAN dengan nilai 0.298, serta Hierarchical yang menunjukkan performa rendah. Selain itu, penelitian [8] menunjukkan bahwa K-Means memiliki waktu eksekusi lebih cepat yaitu 0,4 detik dengan 7050 data. Sedangkan DBSCAN 0,8 dan Hierarchical 1,2 menit.

CV Lantikya atau biasa disebut dengan Lantikya Store merupakan perusahaan yang bergerak di sektor retail di Kabupaten Jombang sejak 2020. Lantikya Store menjual berbagai macam alat tulis, aksesoris, boneka, *babyshop*, *toys*, dan peralatan rumah tangga. Seiring dengan peningkatan persaingan dalam industri bisnis, perusahaan dituntut untuk tidak hanya fokus pengembangan produk. Namun, perusahaan juga harus memahami kebutuhan pelanggan secara mendalam untuk meningkatkan penjualan dan pelayanan.

Gambar 1 Pendapatan Lantikya Store Jombang 2022-2023



Dari data pada Gambar 1 dapat dilihat bahwa omset penjualan Lantikya Store pada tahun 2022 hingga 2023 terdapat fluktuasi dari bulan ke bulan yang cukup signifikan. Pada tahun 2022 terdapat fluktuasi yang signifikan dalam pendapatan bulanan dengan perubahan yang cukup besar dari bulan ke bulan. Terlihat pada bulan Mei mencapai puncak tertinggi, sedangkan bulan Juni mengalami penurunan yang tajam. Selain itu, fluktuasi pendapatan yang signifikan juga terjadi pada bulan-bulan lainnya, seperti peningkatan yang signifikan pada bulan Juli dan penurunan yang cukup besar pada bulan Agustus.

Sementara itu pada tahun 2023 terjadi fluktuasi peningkatan tajam terjadi pada bulan Juli. Dimana peningkatan tersebut dapat diakibatkan karena pada bulan Juli merupakan tahun ajaran baru bagi pelajar atau mahasiswa. Selain itu terjadi fluktuasi yang lebih moderat dibulan selain Juli. Dapat disimpulkan bahwa pendapatan Lantikya Store Jombang dari tahun 2022 hingga 2023 menunjukkan ketidakstabilan yang tinggi selama periode dua tahun. Maka dari itu diperlukannya perencanaan bisnis dan mengoptimalkan strategi pemasaran untuk meningkatkan stabilitas pendapatan Lantikya Store Jombang.

Penggunaan metode *clustering* khususnya K-Means dapat menganalisis dan memahami karakteristik pelanggan Lantikya Store secara lebih mendalam. Dengan melakukan segmentasi pelanggan menggunakan metode *clustering* dapat mengidentifikasi kelompok-kelompok pelanggan yang memiliki kesamaan dalam perilaku pembelian, preferensi produk, dan karakteristik demografis. Hal ini memungkinkan perusahaan untuk menyesuaikan strategi pemasaran secara lebih spesifik dan efektif untuk setiap kelompok pelanggan, sehingga meningkatkan retensi pelanggan dan memaksimalkan keuntungan.

Berdasarkan latar belakang yang telah dijabarkan diatas, maka peneliti tertarik untuk melakukan penelitian tentang segmentasi pelanggan untuk mengoptimalkan strategi pemasaran dengan metode K-Means *clustering*. Dengan memanfaatkan data pendukung lainnya seperti data jenis kelamin, usia, dan pendapatan yang saat ini masih belum dimanfaatkan secara optimal oleh Lantikya Store. Oleh karena itu diperlukan analisis terhadap data-data yang ada untuk memahami pengetahuan yang terkandung didalamnya.

2. TINJAUAN PUSTAKA

2.1. Segmentasi Pelanggan

Segmentasi adalah proses di mana data dikelompokkan ke dalam komponen yang tidak saling tumpang tindih [9]. Sedangkan segmentasi pelanggan adalah proses membagi kelompok pelanggan ke dalam subkelompok yang memiliki kebutuhan, karakteristik, dan perilaku yang berbeda satu sama lain untuk menunjukkan tingkat keberagaman atau heterogenitas di antara pelanggan [10].

2.2. Clustering

Cluster merupakan sekelompok objek data yang memiliki kemiripan satu sama lain dan berbeda dengan objek yang terdapat di *cluster* lainnya. Objek-objek tersebut dikelompokkan ke dalam satu atau lebih *cluster*, sehingga objek yang berada dalam satu *cluster* memiliki tingkat kesamaan yang tinggi [11]. Sedangkan *clustering* merupakan proses pengelompokan dan pembagian pola data ke dalam beberapa set data sehingga membentuk pola yang serupa dan tergabung dalam *cluster* yang sama, sementara pola yang berbeda terisolasi dalam *cluster* yang berbeda.

2.3. Algoritma K-Means

Algoritma K-means adalah salah satu metode dalam teknik *clustering* non-hierarki yang berupaya membagi data yang ada ke dalam satu atau lebih *cluster* [12]. Data dalam *cluster* dikelompokkan ke dalam *cluster* yang memiliki karakteristik yang sama. Algoritma ini mengelompokkan data ke dalam k kelompok atau parameter input.

2.4. Metode Elbow

Metode elbow adalah suatu pendekatan yang digunakan untuk mendapatkan informasi tentang penentuan jumlah klaster terbaik dengan melihat titik di mana presentase variasi jumlah klaster membentuk siku pada grafik [13]. Dasar metode elbow adalah menggunakan kuadrat jarak antara titik-titik sampel di setiap *cluster*, dan pusat *cluster* memberikan sekumpulan nilai k . Indikator kinerja metode ini dapat dilakukan dengan perhitungan *Within-Cluster Sum Square Error* (WSS) menggunakan persamaan (1).

$$WSS = \sum_{k=1}^n \sum_{x_i \in C_k} |x_i - c_k|^2 \quad (1)$$

Dimana n adalah jumlah *cluster*, x_i adalah titik data, dan c_k adalah centroid dari *cluster* k . WSS dihitung dengan menjumlahkan kuadrat jarak Euclidean antara setiap titik data dan centroidnya dalam setiap *cluster*, kemudian menjumlahkan hasilnya untuk semua *cluster*.

2.5. Metode Silhouette Coefficient

Metode *Silhouette Coefficient* merupakan kombinasi dua metode yaitu metode *cohesion* untuk mengukur kedekatan antar objek dalam suatu *cluster* dan metode *separation* untuk mengukur seberapa jauh suatu *cluster* terpisah dari *cluster* lainnya [14]. Hasil dari perhitungan metode ini memiliki rentan -1 sampai 1. Jika

nilai mendekati angka 1 maka pengelompokkan dalam suatu kluster semakin membaik.

2.6. Metode Gap Statistic

Gap Statistic adalah suatu metode yang digunakan untuk memperkirakan jumlah optimal kelompok dalam analisis kluster [15]. Metode Gap statistik mengevaluasi kualitas pengelompokan dengan membandingkan nilai log W dengan log E untuk setiap jumlah *cluster*. Gap statistik membantu menentukan jumlah *cluster* yang optimal dengan menganalisis grafik gap statistik, yang menunjukkan perbedaan antara nilai log W dan log E untuk berbagai jumlah *cluster* [16].

2.7. Recency, Frequency, Monetary (RFM)

Recency Frequency Monetary (RFM) adalah sebuah alat strategi pemasaran yang umumnya dipakai untuk mengidentifikasi segmentasi pelanggan berdasarkan perilaku, dengan memanfaatkan informasi dari data transaksi [17]. Model RFM memiliki tiga atribut yang dapat digunakan untuk segmentasi pelanggan sebagai berikut [18]:

1. Recency

Waktu (dalam satuan seperti hari/bulan/tahun) sejak transaksi pembelian atau kunjungan belanja terakhir.

2. Frequency

Total jumlah transaksi pembelian atau kunjungan belanja dalam periode yang dianalisis.

3. Monetary

Nilai total pembelian dalam periode yang dianalisis atau nilai rata-rata (misalnya, nilai rata-rata bulanan) per satuan waktu.

2.8. Normalisasi Min-Max

Normalisasi Min-Max adalah suatu teknik normalisasi yang melibatkan transformasi linier terhadap data awal. Metode ini mengubah rentang nilai data asli sehingga rentangnya berkisar antara 0 dan 1 [19]. Perhitungan dapat dilakukan dengan persamaan 2.

$$d' = \frac{[d - \min(p)] * [new_{\max(p)} - new_{\min(p)}]}{[\max(p) - \min(p)]} + new_{\min(p)} \quad (2)$$

Dimana:

d' = nilai yang telah

dinormalisasi

d = nilai asli

$\min(p)$ = nilai terkecil atribut

$\max(p)$ = nilai terbesar atribut

$new_{\min(p)}$ = rentang minimal

$new_{\max(p)}$ = rentang maksimal

2.9. Analisis Korelasi

Analisis korelasi merupakan salah satu teknik yang digunakan untuk mengukur arah dan kekuatan hubungan antaran dua variabel. Dua variabel dianggap berkorelasi jika perubahan pada salah satu variabel diikuti oleh perubahan pada variabel lainnya, baik dalam arah yang sama maupun berlawanan [20]. Dalam analisis *cluster*, koefisien korelasi yang jauh lebih rendah dapat dianggap penting. Standar aturan dalam mengevaluasi dan menginterpretasikan nilai koefisien korelasi terdapat pada Tabel 1 [21].

Tabel 1 Interpretasi Nilai Korelasi

Nilai Korelasi	Interpretasi
0,90 – 1,00	Korelasi Sangat Tinggi
0,70 – 0,89	Korelasi Tinggi
0,50 – 0,69	Korelasi Moderat
0,30 – 0,49	Korelasi Lemah
0,00 – 0,29	Sedikit Jika Ada Korelasi

Untuk menghitung nilai korelasi dapat dilakukan dengan koefisien korelasi Pearson pada persamaan 3.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (3)$$

Dimana:

n : jumlah pasangan data

x dan y : nilai dari dua variabel

$\sum xy$: jumlah hasil kali pasangan

nilai x dan y

$\sum x$: jumlah nilai x

$\sum y$: jumlah nilai y

$\sum x^2$: jumlah kuadrat nilai x

$\sum y^2$: jumlah kuadrat nilai y

2.10. Analytic Hierarchy Process (AHP)

Analytic Hierarchy Process (AHP) merupakan suatu teknik pengambilan keputusan yang menggabungkan berbagai faktor seperti logika, intuisi, pengalaman, pengetahuan, emosi, dan rasa secara sistematis untuk mencapai optimalitas dalam proses pengambilan keputusan [22]. Metode ini dikembangkan oleh Thomas L. Saaty pada tahun 1998 dari Wharston Business school untuk menemukan peringkat atau urutan prioritas dari berbagai alternatif dalam menyelesaikan suatu masalah. Proses *Analytic Hierarchy Process (AHP)* dapat dilakukan dengan langkah berikut [23]:

1. Menyusun struktur keputusan *hierarki* yang mencakup faktor pertimbangan dan alternatif keputusan, dengan fokus pada model *RFM* dalam kasus pembobotan.
2. Melakukan perbandingan berpasangan untuk menilai kepentingan dalam model *RFM* menggunakan skala yang tercantum dalam Tabel 2.

Tabel 2 Skala Penilaian Perbandingan Berpasangan

Intensitas Kepentingan	Keterangan
1	Kedua elemen sama penting
3	Elemen yang satu sedikit lebih penting daripada elemen yang lainnya
5	Elemen yang satu lebih penting daripada elemen yang lainnya
7	Satu elemen jelas lebih mutlak penting daripada elemen lainnya
9	Satu elemen mutlak penting daripada elemen lainnya
2,4,6,8	Nilai-nilai antara dua nilai pertimbangan-pertimbangan yang berdekatan
Kebalikan	Jika aktivitas i mendapat satu angka dibanding dengan aktivitas j, maka j mempunyai nilai kebalikannya dibanding dengan i

3. Uji Konsistensi

Sebelum menghitung bobot, perlu dilakukan perhitungan konsistensi menggunakan indeks inkonsistensi. Konsistensi yang optimal dicapai ketika nilai inkonsistensi adalah 0. Meskipun konsistensi sempurna jarang tercapai karena adanya bias dan ketidak konsistenan manusia dalam penilaian, nilai indeks konsistensi kurang dari 0.1 masih dianggap dapat diterima. Uji Konsistensi dilakukan dengan mencari C_i dan C_r melalui persamaan 4 dan 5.

$$C_i = \frac{(\lambda_{max} - \text{jumlah kriteria})}{\text{jumlah kriteria} - 1} \quad (4)$$

$$C_r = \frac{C_i}{IR} \quad (5)$$

Dimana:

C_i = Consistency index

C_r = Consistency Ratio

IR = Indeks Random

Nilai IR didapatkan dari eksperimen Oak Ridge National Laboratory dan dikembangkan oleh Wharton School dengan nilai pada Tabel 3.

Tabel 3 Nilai IR

n	Keterangan
1	0
2	0
3	0.58
4	0.9
5	1.12
6	1.24
7	1.32
8	1.4
9	1.45
10	1.49

2.11. Customer Lifetime Value (CLV)

Customer Lifetime Value (CLV) adalah nilai saat ini dari aliran laba yang diharapkan dari pelanggan selama periode pembelian sepanjang hidupnya [24]. Rumus untuk menghitung *Customer Lifetime Value (CLV)* dapat dilakukan dengan mengalikan bobot prioritas dengan nilai masing-masing kelompok pelanggan Atau dengan rumus pada persamaan 6 [25].

$$CLV = NR \times WR + NF \times WF + NM \times WM \quad (6)$$

Dimana:

CLV = nilai CLV yang dicari

NR = nilai normalisasi variabel *recency*

WR = nilai bobot variabel *recency*

NF = nilai normalisasi variabel *frequency*

WF = nilai bobot variabel *frequency*

NM = nilai normalisasi variabel *monetary*

WM = nilai bobot variabel *monetary*

2.12. Usia

Usia dapat digunakan sebagai faktor penentu dalam berbagai konteks, seperti dalam analisis demografis, penelitian pasar, dan aspek-aspek lain dari pemahaman populasi atau kelompok tertentu [26]. Usia dapat dikelompokkan berdasarkan generasi untuk

menyusun strategi pemasaran. Kelompok dapat dilihat pada Tabel 4 dan Tabel 5 [27].

Tabel 4 Generasi Usia

* Usia dihitung dari tahun kelahiran hingga 2024

Tahun Kelahiran	Usia	Generasi
1928 - 1945	79 - 96 tahun	Silent Generation
1946 - 1964	60 – 78 tahun	Baby Boomers
1965 - 1980	44 – 59 tahun	Generation X
1981 - 1996	43 – 28 tahun	Mileneal
1997 - 2012	12 – 27 tahun	Generation Z

Tabel 5 Penyusunan Strategi Pemasaran Berdasarkan Usia

Generasi	Strategi Pemasaran
Generasi Baby Boomers	Mengoptimalkan penggunaan e-mail, telepon dan facebook; menargetkan boomers melalui desktop dan gawai oleh karena itu pemasaran melalui Whatsapp for business
Generasi X	Membangun loyalitas terhadap brand dengan menyederhanakan proses pembelian; memanfaatkan iklan video; menawarkan hadiah dan program loyalitas.
Generasi Y (milenial)	Menarik pengguna melalui influencer; Review online untuk meningkatkan reputasi dan pendapatan brand; Memastikan identitas brand konsisten
Generasi Z (Zilenial)	Meningkatkan personalisasi; menggunakan pemasaran melalui influencer untuk menjangkau target pasar; Membuat personal brand lintas media sosial

2.13. Pendapatan

Pendapatan yang diterima oleh konsumen memiliki dampak signifikan terhadap perilaku konsumen dalam pengambilan keputusan pembelian produk. Tingkat pendapatan yang diterima oleh seseorang sangat memengaruhi seberapa besar mereka bersedia mengeluarkan uang dan pola konsumsi mereka. Individu dengan pendapatan tinggi cenderung memiliki pengeluaran yang lebih besar dan pola konsumsi yang lebih tinggi,

sementara mereka dengan pendapatan rendah cenderung memiliki pengeluaran dan pola konsumsi yang lebih terbatas. Kesimpulan ini menunjukkan bahwa pendapatan memainkan peran penting dalam membentuk keputusan konsumen dan mengarahkan pola konsumsi masing-masing individu [28].

2.14. Nilai RFM

Karakteristik pelanggan dapat dipetakan menjadi matriks nilai pelanggan berdasarkan perbandingan antara *Frequency* (*F*) pembelian dan nilai *Monetary* (*M*) [29]. Selain itu, variabel kebaruan transaksi atau *Recency* (*R*) digunakan sebagai indikator kesetiaan pelanggan.

Klasifikasi pelanggan dan karakteristik pelanggan terbagi menjadi 8 tipe pelanggan berdasarkan nilai kebaruan transaksi pelanggan (*Recency*), *Frequency* pembelian, dan nilai *Monetary* pada Tabel 5 [30]. Nilai variabel yang berada di atas rata-rata dilambangkan dengan ↑ dan nilai variabel yang berada di bawah rata-rata dilambangkan dengan ↓.

Tabel 6 Tipe Pelanggan

Tipe Pelanggan	Simbol
Pelanggan bernilai penting	R ↑ F ↑ M ↑
Pelanggan yang Penting untuk Pengembangan	R ↑ F ↓ M ↑
Pelanggan yang Perlu Dilindungi	R ↓ F ↑ M ↑
Pelanggan yang Penting untuk Retensi	R ↓ F ↓ M ↑
Pelanggan dengan Nilai Umum	R ↑ F ↑ M ↓
Pelanggan untuk Pengembangan Umum	R ↑ F ↓ M ↓
Pelanggan untuk Retensi Umum	R ↓ F ↑ M ↓
Pelanggan yang Hilang	R ↓ F ↓ M ↓

Dari tipe pelanggan di atas, disusun dasar untuk strategi operasional untuk kelompok pelanggan yang berbeda pada Tabel 7.

Tabel 7 Dasar Penyusunan Strategi Pemasaran

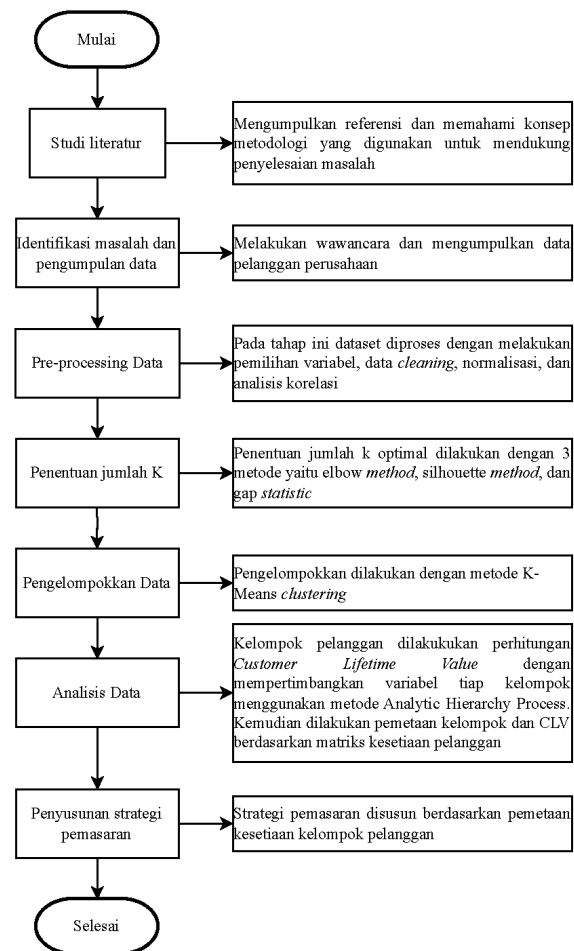
Tipe Pelanggan	Strategi Pemasaran
Pelanggan bernilai penting	Tingkatkan ke pelanggan orang yang sangat penting (VIP), berikan layanan yang dipersonalisasi, dan berikan lebih banyak sumber daya.
Pelanggan yang Penting	Memberikan layanan poin anggota dan memberikan diskon tertentu untuk

untuk Pengembangan	meningkatkan tingkat retensi pelanggan.
Pelanggan yang Perlu Dilindungi	Memperkenalkan produk/fungsi/layanan terbaru melalui SMS dan email untuk meningkatkan konsumsi pelanggan
Pelanggan yang Penting untuk Retensi	Memperkenalkan produk/fungsi/layanan upgrade terbaru, promosi diskon, dll melalui SMS, email, telepon, dll, untuk menghindari kehilangan pelanggan.
Pelanggan dengan Nilai Umum	Memperkenalkan produk/fungsi/layanan terbaru untuk meningkatkan konsumsi pelanggan.
Pelanggan untuk Pengembangan Umum	Memberikan layanan masyarakat, memperkenalkan produk/fungsi baru, dan mempromosikan konsumsi pelanggan.
Pelanggan untuk Retensi Umum	Memperkenalkan produk/fungsi baru untuk membangkitkan minat pelanggan ini.
Pelanggan yang Hilang	Bagian pelanggan ini dapat dibangkitkan dengan promosi dan diskon. Ketika alokasi sumber daya tidak mencukupi, bagian pengguna ini dapat ditinggalkan untuk sementara.

3. METODE PENELITIAN

Metode penelitian adalah proser atau langkah-langkah sistematis dalam menyusun ilmiah atau ilmu pengetahuan [31]. Penelitian ini menggunakan pendekatan kuantitatif, di mana data diungkapkan dalam bentuk angka dan dianalisis menggunakan teknik statistik. Adapun tahapan yang dilakukan untuk menyelesaikan penelitian ini dapat ditunjukkan pada Gambar 2.

Gambar 2 Tahapan Penelitian



4. HASIL DAN PEMBAHASAN

4.1. Studi Literatur

Studi literatur dalam penelitian ini diambil dari berbagai penelitian sebelumnya yang relevan, serta referensi yang diperoleh dari sejumlah jurnal, *e-book*, dan situs *website* terpercaya mengenai penerapan algoritma K-Means untuk *clustering*. Teori penelitian sebelumnya dan relevan terhadap penelitian ini yaitu oleh Alamsyah dan Pangestu terkait segmentasi pelanggan menggunakan metode K-Means *clustering* dan mode RFM [32][33].

4.2. Identifikasi Masalah dan Pengumpulan Data

Data pelanggan yang dibutuhkan adalah riwayat transaksi pada bulan November tahun 2022 hingga Desember 2023. Selain itu, dilakukan wawancara terkait profil perusahaan dan strategi pemasaran dengan Manager Lantikya Store Jombang. Data yang diperoleh

berupa atribut yang terdapat pada Tabel 8 dengan jumlah data yaitu 1140 data.

Tabel 8 Atribut Data

Atribut	Keterangan
Nama	Nama lengkap pembeli
Jenis kelamin	Informasi jenis kelamin berupa laki-laki dan perempuan
Usia	Usia pembeli
Pekerjaan	Status pekerjaan atau profesi pembeli
Pendapatan	Pendapatan pembeli selama satu tahun
Tanggal Pembelian	Tanggal dilakukan pembelian
Kunjungan	Jumlah kunjungan atau seberapa sering melakukan pembelian di tahun 2022-2023
Total Belanja	Total pembelian yang dilakukan selama berbelanja di Lantikya Store

4.3. Pre-Processing Data

Setelah proses pengumpulan data, maka tahap selanjutnya yaitu *pre-processing* dengan mengolah data yang telah dikumpulkan menjadi data siap olah sesuai dengan atribut yang dibutuhkan. Tahap yang dilakukan dalam *pre-processing* data adalah pemilihan data, *data cleaning*, normalisasi, dan analisis korelasi dengan bantuan *Microsoft Excel*.

4.3.1. Pemilihan Data

Proses pemilihan data melibatkan pemahaman terhadap setiap variabel atau kolom data yang sesuai dengan kebutuhan penelitian. Data pelanggan terdiri dari 8 kolom yang dijelaskan pada Tabel 8. Dari 8 kolom data pelanggan, dipilih 6 kolom yaitu nama, usia, pendapatan, tanggal, kunjungan, dan total belanja. Pada tahap ini dilakukan dengan menghapus kolom yang tidak diperlukan.

Selain itu, pemilihan data digunakan untuk mengidentifikasi variabel yang diperlukan seperti usia, pendapatan dan model RFM (*Recency, Frequency, Monetary*). Variabel *recency* diperoleh melalui perhitungan selisih waktu antara tanggal terakhir pelanggan melakukan transaksi sampai dengan tanggal akhir yang ditentukan. Dalam data pelanggan Lantikya Store, variabel *recency* didapatkan melalui perhitungan selisih atribut tanggal pembelian dengan tanggal akhir yang ditentukan yaitu 31 Desember 2023. Perhitungan memanfaatkan fungsi dalam *Microsoft Excel* yaitu *days*.

4.3.2. Data Cleaning

Dataset yang sudah melalui proses pemilihan data akan dilakukan pembersihan data dari nilai-nilai yang tidak valid atau hilang, penanganan *outlier*, serta penyesuaian data. Data *cleaning* dilakukan dengan menghapus data yang memiliki nilai *missing value*. Pada tahap ini dilakukan dengan menggunakan fitur *sort* pada *Microsoft Excel*.

Tabel 9 Data Pelanggan Tidak Valid

Nama	Usia	Pendapatan	R	F	M
Kristina Nadila	18	0	5	47000	
A.. Dika	17	0	2	61000	
W..	11	0	5	42000	
Feonia Anggun	19	0	3	54500	
D..	33	70000000	7	230000	

Dari data *cleaning*, terdapat beberapa data yang tidak valid. Diantara lain pada variabel *recency* terdapat 5 data yang memiliki nilai tidak valid. Total seluruh data nilai tidak valid dihapus dari data pelanggan. Sehingga tersisa 1135 data yang memiliki nilai dan valid.

4.3.3. Normalisasi

Data yang telah disiapkan dan sudah melalui proses pemilihan data dan data *cleaning* akan diubah melalui normalisasi data dengan menyetarakan skala nilai variabel 0 sampai 1. Perhitungan normalisasi dapat dilakukan dengan persamaan 2.

Tabel 10 Nilai Minimal dan Maksimal Variabel

Variabel	Min	Max
Usia	11	50
Pendapatan	0	200000000
Recency	1	364
Frequency	1	30
Monetary	10000	1450000

4.3.4. Analisis korelasi

Analisis korelasi merupakan teknik untuk mengukur kekuatan hubungan antar variabel. Adapun perhitungan dan aturan dapat dilihat pada dasar teori 2.9. Hasil perhitungan analisis korelasi terdapat pada Tabel 4.7.

Setelah dilakukan analisis korelasi dan tidak ada korelasi yang signifikan, maka langkah selanjutnya yaitu mengubah hasil normalisasi format *Excel Worksheet* (.xlsx) kedalam format *Comma Separated Value* (.csv).

Tabel 11 Hasil Analisis Korelasi Variabel

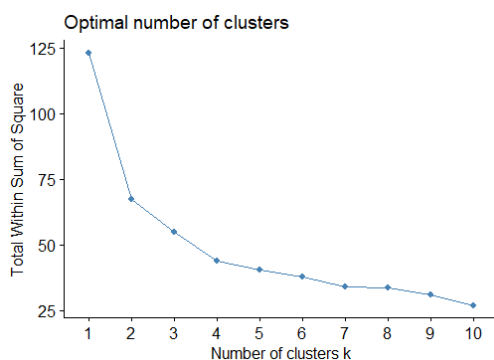
Variabel	Usia	Pendapatan	R	F	M
Usia	1	0,5317	0,0275	-0,0009	0,0670
Pendapatan	0,531	1	0,0030	0,0242	0,0214
R	0,027	0,0030	1	-0,0011	0,0115
F	-0,0009	0,0242	-0,0011	1	0,1565
M	0,0670	0,0214	0,0115	0,1565	1

Berdasarkan hasil analisis korelasi variabel usia, pendapatan, *recency*, *frequency*, dan *monetary* menunjukkan bahwa hubungan antar variabel tidak memiliki korelasi tinggi. Dengan demikian, hasil korelasi antar variabel memiliki hubungan beragam dan tidak konsisten diantara satu sama lain.

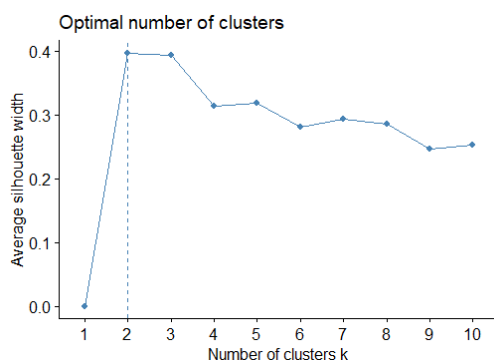
4.4. Penentuan Jumlah k

Penentuan jumlah kelompok dilakukan dengan membandingkan hasil tiga metode pencarian *cluster* yaitu metode siku (*elbow method*), metode siluet (*silhouette method*), dan statistik kesenjangan (*gap statistic*).

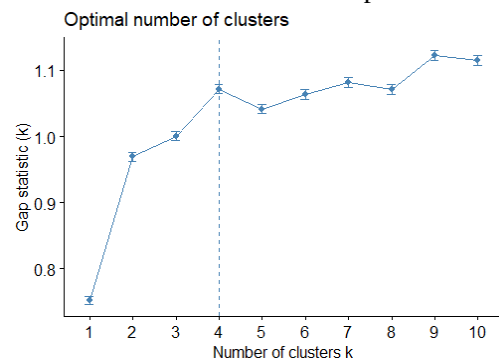
Gambar 3 Hasil Metode Elbow



Gambar 4 Hasil Rata-Rata Metode Silhouette



Gambar 5 Hasil Metode Gap Statistic



Berdasarkan hasil grafik pada Gambar 4.5, metode elbow dan metode gap statistik menunjukkan jumlah *cluster* optimal di $k=4$. Sedangkan jumlah *cluster* optimal metode siluet yaitu $k=2$. Dengan demikian dapat disimpulkan jumlah *cluster* optimal berdasarkan perbandingan ketiga metode pencarian *cluster* yaitu $k=4$. Langkah selanjutnya yaitu pengelompokkan data kedalam 4 *cluster*.

4.5. Pengelompokkan Data

Metode yang digunakan dalam melakukan pengelompokkan yaitu metode K-Means. Dataset yang digunakan adalah dataset yang sebelumnya telah dilakukan normalisasi setiap variabelnya. Proses pengelompokkan memanfaatkan perangkat lunak Rstudio. Sejumlah 1135 data dibagi menjadi 4 *cluster* dengan anggota *cluster* 1 yaitu 419 anggota, *cluster* 2 dengan 97 anggota, *cluster* 3 dengan 542 anggota, dan *cluster* 4 dengan 77 anggota.

Tabel 12 Cuplikan Hasil Kelompok Data

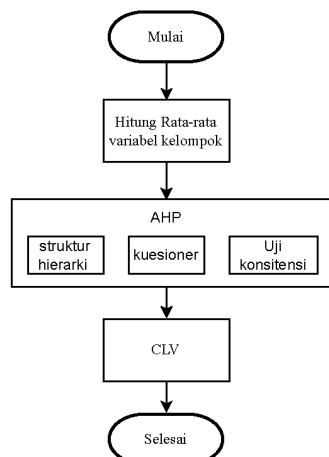
Nama	Usia	Pendapatan	R	F	M	Cluster
Adityas	0,28	0,07	0,19	0	0,25	3
Anjeli S	0,58	0,09	0,19	0,06	0,01	2
Nurma	0,35	0,06	0,32	0,03	0,10	3
Agus T	0,20	0,06	0,32	0,06	0,02	3
Karina	0,46	0,03	0,32	0,06	0,14	2

4.6. Analisis Hasil Cluster

Hasil dari kelompok pelanggan dilakukan perhitungan rata-rata untuk mendapatkan nilai *Customer Lifetime Value* (CLV). Perhitungan *Customer Lifetime Value* (CLV) dilakukan dengan mengalikan nilai rata-rata tiap atribut dengan bobot prioritas. Nilai bobot prioritas didapatkan melalui perhitungan

bobot atribut menggunakan metode *Analytic Hierarchy Process (AHP)* dengan langkah yang telah dibahas pada dasar teori bagian 2.10. Berikut langkah dalam analisis hasil *cluster*.

Gambar 6 Langkah Analisis Hasil Cluster



4.6.1. Rata-Rata Variabel Kelompok

Perhitungan nilai rata-rata dilakukan melalui perangkat lunak Rstudio dengan hasil pada Tabel 13. Sedangkan nilai rata-rata setiap *cluster* yang belum melalui proses normalisasi dihitung melalui *Microsoft Excel* dengan nilai dapat dilihat pada Tabel 14.

Tabel 13 Hasil Rata-Rata Cluster Normalisasi

Cluster	Usia	Penda patan	R	F	M
1	0,24	0,04	0,74	0,07	0,08
2	0,61	0,10	0,36	0,07	0,09
3	0,25	0,04	0,29	0,07	0,08
4	0,54	0,096	0,80	0,07	0,14
Rata-rata	0,41	0,07	0,55	0,07	0,10

Tabel 14 Hasil Rata-Rata Cluster

Cluster	Usia	Pendapatan	R	F	M
1	21	Rp 14.873.866	270	3	Rp 132.185
2	35	Rp 36.823.216	133	3	Rp 153.371
3	21	Rp 15.673.107	108	3	Rp 139.265
4	32	Rp 34.629.870	293	3	Rp 211.831
Rata-rata	27	Rp 25.500.015	201	3	Rp 159.163

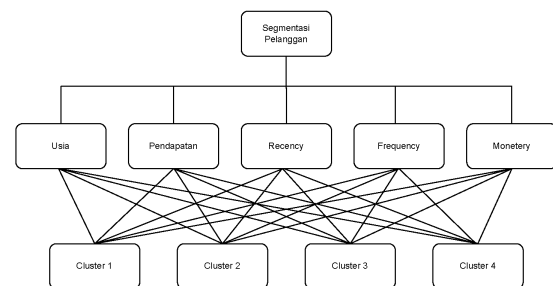
4.6.2. Analytic Hierarchy Process (AHP)

Analytic Hierarchy Process (AHP) digunakan untuk menentukan atribut mana yang

menjadi prioritas dalam mempengaruhi pembelian customer atau penyusunan strategi pemasaran. Tahap dalam proses ini yaitu penyusunan struktur hierarki, pengisian kuesioner, dan pengujian konsistensi.

Struktur hierarki disusun untuk memberikan pemahaman tentang bagaimana data dibagi menjadi kelompok atau *cluster*. Setiap *cluster* mewakili sekelompok data yang memiliki karakteristik serupa berdasarkan kriteria variabel yang telah ditentukan. Struktur hierarki dalam segmentasi dapat dilihat pada Gambar 7.

Gambar 7 Struktur Hierarki



Setelah dilakukan penyusunan struktur hierarki, maka langkah selanjutnya yaitu perhitungan bobot kriteria variabel. Penentuan bobot prioritas dilakukan dengan memberikan kuesioner yang berisikan perbandingan variabel atau kriteria penelitian kepada pihak perusahaan yaitu manager Lantikya Store Jombang.

Gambar 8 Hasil Pengisian Kuesioner

Kriteria	Skala Penilaian Perbandingan																	Kriteria
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Usia	V																	Pendapatan
Usia					V													Recency
Usia	V																	Frequency
Usia			V															Monetary
Pendapatan						V												Recency
Pendapatan							V											Frequency
Pendapatan								V										Monetary
Recency										V								Frequency
Recency											V							Monetary
Frequency												V						Monetary

Hasil kuesioner disusun matriks perbandingan berpasangan dengan hasil pada Tabel 15. Selanjutnya dilakukan rekapitulasi hasil pengisian pada Tabel 16

Tabel 15 Matriks Perbandingan Hasil Kuesioner

Kriteria	Matriks Perbandingan Berpasangan				
	Usia	Pendapatan	R	F	M
Usia	1	9	5	9	7
Pendapatan	1/9	1	3	3	3

Kriteria	Matriks Perbandingan Berpasangan				
	Usia	Pendapatan	R	F	M
R	1/5	1/3	1	1/5	1/5
F	1/9	1/3	5	1	5
M	1/7	1/3	5	1/5	1

Tabel 16 Rekapitulasi Kuesioner

Kriteria	Rekapitulasi Matriks Perbandingan Berpasangan				
	Usia	Pendapatan	R	F	M
Usia	1,00	9,00	5,00	9,00	7,00
Pendapatan	0,11	1,00	3,00	3,00	3,00
R	0,20	0,33	1,00	0,20	0,20
F	0,11	0,33	5,00	1,00	5,00
M	0,14	0,33	5,00	0,20	1,00
Total	1,57	11,00	19,0	13,4	16,2

Setelah dilakukan rekapitulasi nilai, maka langkah selanjutnya yaitu normalisasi nilai setiap variabel. Perhitungan normalisasi dilakukan dengan membagi nilai setiap variabel dengan total jumlah kolom variabel. Hasil normalisasi dijumlahkan untuk menentukan prioritas setiap variabel. Nilai prioritas diperoleh dengan membagi jumlah setiap variabel dengan jumlah kriteria. Hasil penentuan nilai prioritas setiap variabel dapat dilihat pada Gambar 9.

Gambar 9 Hasil Nilai Bobot Prioritas

Normalisasi Matriks Perbandingan Berpasangan							
Kriteria	Usia	Pendapatan	Recency	Frequency	Monetary	Jumlah	Prioritas
Usia	0,64	0,82	0,26	0,67	0,43	2,82	0,56
Pendapatan	0,07	0,09	0,16	0,22	0,19	0,73	0,15
Recency	0,13	0,03	0,05	0,01	0,01	0,24	0,05
Frequency	0,07	0,03	0,26	0,07	0,31	0,75	0,15
Monetary	0,09	0,03	0,26	0,01	0,06	0,46	0,09
Jumlah						5,00	1,00

Sebelum dilakukan uji konsistensi, hitung nilai jumlah per baris setiap variabel. Perhitungan dilakukan dengan mengalikan bobot prioritas dengan nilai setiap variabel. Kemudian, hasil perhitungan nilai setiap variabel baris dijumlahkan dengan hasil yang ditunjukkan pada Gambar 10.

Gambar 10 Jumlah Matriks Baris

Kriteria	Usia	Pendapatan	Recency	Frequency	Monetary	Jumlah per Baris
Usia	0,56	1,31	0,24	1,35	0,65	4,11
Pendapatan	0,06	0,15	0,14	0,45	0,28	1,08
Recency	0,11	0,05	0,05	0,03	0,02	0,26
Frequency	0,06	0,05	0,24	0,15	0,46	0,96
Monetary	0,08	0,05	0,24	0,03	0,09	0,49

Nilai jumlah kriteria setiap baris dijumlahkan dengan nilai prioritas untuk mendapatkan nilai ratio konsistensi. Dimana

jika nilai ratio konsistensi dijumlahkan, maka akan menghasilkan λ_{max} . Hasil perhitungan perhitungan dapat dilihat pada Tabel 4.19

Tabel 17 Hasil Maximum Value (λ_{max})

Kriteria	Jumlah per Baris	Prioritas	Ratio
Usia	4,11	0,56	4,67
Pendapatan	1,08	0,15	1,22
Recency	0,26	0,05	0,31
Frequency	0,96	0,15	1,11
Monetary	0,49	0,09	0,58

$$\lambda_{max} = \frac{4,16 + 1,22 + 0,31 + 1,11 + 0,58}{5} = 1,58$$

Langkah terakhir yaitu menghitung indeks inkonsistensi untuk menilai apakah kuesioner yang telah di isi oleh responden bernilai konsisten atau inkonsisten. Langkah perhitungan konsistensi dapat dilihat pada persamaan 4 dan 5. Hasil perhitungan konsistensi sebagai berikut:

$$Ci = \frac{1,58 - 5}{5 - 1} = -0,86$$

$$Cr = \frac{-0,86}{1.12} = -0,76$$

Berdasarkan hasil nilai *consistency ratio* pengisian kuesioner oleh manager Lantikya Store Jombang bernilai -0,76 dan lebih kecil daripada 0,1. Hal tersebut membuktikan bahwa nilai inkonsisten dan dapat diterima.

4.6.3. Customer Lifetime Value (CLV)

Perhitungan *Customer Lifetime Value* (CLV) dilakukan dengan persamaan 6. Nilai variabel yang digunakan yaitu nilai variabel rata-rata setiap *cluster* yang ditunjukkan pada Tabel 13. Sedangkan nilai bobot yang digunakan yaitu nilai bobot prioritas yang sebelumnya telah dilakukan perhitungan. Nilai tersebut yaitu 0.56, 0.15, 0.05, 0.15, 0.09 dengan variabel berturut-turut usia, pendapatan, *recency*, *frequency*, dan *monetary*. Hasil perhitungan *Customer Lifetime Value* (CLV) ditunjukkan pada Tabel 18.

Tabel 18 Customer Lifetime Value (CLV)

Cluster	Usia	Pendapatan	R	F	M	CLV
1	0.24	0.04	0.74	0.07	0.08	0.20
2	0.61	0.10	0.36	0.07	0.09	0.39
3	0.25	0.04	0.29	0.07	0.08	0.18

Cluster	Usia	Pendapatan	R	F	M	CLV
4	0.54	0.09	0.80	0.07	0.14	0.38

Berdasarkan hasil perhitungan *Customer Lifetime Value (CLV)*, didapatkan bahwa *cluster* yang memiliki nilai tertinggi yaitu dimulai dari *cluster 2*, *cluster 4*, *cluster 1*, dan *cluster 3*.

4.6.4. Penyusunan Strategi Pemasaran

Penyusunan diprioritaskan berdasarkan nilai tertinggi hingga terendah *Customer Lifetime Value (CLV)*. Strategi pemasaran ditentukan berdasarkan karakteristik usia, pendapatan, dan tipe pelanggan setiap *cluster*. Karakteristik usia dan pendapatan pelanggan dipetakan dengan teori 2.12 dan 2.13. Berdasarkan hasil nilai rata-rata variabel setiap *cluster* yang ditunjukkan pada Tabel 13, karakteristik usia dalam *cluster 2* dan *4* memiliki usia yang termasuk kedalam generasi milenial. *Cluster* tersebut memiliki pendapatan tinggi diatas rata-rata dibandingkan dengan *cluster* yang lain. Sedangkan *cluster 1* dan *3* memiliki karakteristik yang tergolong kedalam generasi Z dan memiliki penghasilan rendah dibawah rata-rata.

Sedangkan tipe pelanggan dipetakan berdasarkan nilai variabel *recency*, *frequency*, dan *monetary* (RFM) yang telah dihitung berdasarkan rata-rata variabel. Penentuan tipe pelanggan telah dibahas pada teori 2.14 dan hasil ditunjukkan pada Gambar 11. Nilai variabel yang berada diatas rata-rata dilambangkan dengan ↑ dan nilai variabel yang berada dibawah rata-rata dilambangkan dengan ↓.

Gambar 11 Nilai Variabel Pelanggan

Cluster	Usia	Pendapatan	Recency	Frequency	Monetary	Nilai Variabel
2	0.6103621	0.10227393	0.3623300	0.07785283	0.09956329	UTPTR↓ F ↑ M↓
4	0.5471195	0.09618206	0.8043362	0.07971339	0.14016053	UTPTR↑ F ↑ M↑
1	0.2489444	0.04131113	0.7405669	0.07538474	0.08485067	U↓P↓R↑ F ↓ M↓
3	0.2524837	0.04353097	0.2951318	0.07360987	0.08976710	U↓P↓R↓ F ↓ M↓
Rata-rata	0,414727425	0,070824523	0,550591225	0,076640208	0,103585398	

Tabel 19 Tipe Pelanggan

Cluster	Nilai Variabel	Tipe Pelanggan
2	U↑P↑R↓ F ↑ M↓	Pelanggan untuk retensi umum
4	U↑P↑R↑ F ↑ M↑	Pelanggan bernilai penting
1	U↓P↓R↑ F ↓ M↓	Pelanggan untuk pengembangan umum

Cluster	Nilai Variabel	Tipe Pelanggan
3	U↓P↓R↓ F ↓ M↓	Pelanggan yang hilang

Setelah dipetakan berdasarkan tipe pelanggan dan karakteristik usia dan pendapatan pelanggan pada Tabel 19, maka langkah selanjutnya yaitu penyusunan strategi pemasaran berdasarkan karakteristik dan tipe pelanggan masing-masing *cluster* dengan rekomendasi yang telah dibahas pada teori 2.14.

Cluster 2 terdiri dari generasi milenial dengan pendapatan tinggi yang tergolong pelanggan untuk retensi umum. Strategi pemasaran yang disarankan adalah menggunakan influencer, mendorong ulasan online positif, dan memperkenalkan produk baru secara berkala untuk meningkatkan visibilitas dan daya tarik. Cluster 4, juga generasi milenial dengan pendapatan tinggi, merupakan pelanggan bernilai penting. Strategi pemasaran mencakup layanan personalisasi, penawaran eksklusif, dan pengalaman belanja khusus untuk mempertahankan loyalitas mereka. Cluster 1 terdiri dari generasi Z dengan pendapatan rendah yang tergolong pelanggan untuk pengembangan umum. Strategi melibatkan influencer dengan kredibilitas, personalisasi kampanye, dan promosi menarik untuk meningkatkan keterlibatan dan loyalitas. Cluster 3 adalah generasi Z dengan pendapatan rendah yang merupakan pelanggan yang hilang. Strategi pemasaran melibatkan personalisasi, penggunaan influencer relevan, dan penawaran promosi khusus untuk menarik kembali minat belanja mereka. Dengan demikian, berbagai strategi pemasaran disesuaikan dengan karakteristik dan kebutuhan masing-masing *cluster* untuk meningkatkan keterlibatan dan loyalitas pelanggan.

5. KESIMPULAN

Penelitian mengenai clustering segmentasi pelanggan di Lantikya Store Jombang berdasarkan usia, pendapatan, dan model RFM menggunakan metode siku, siluet, dan statistik kesenjangan menunjukkan bahwa jumlah *cluster* optimal adalah empat. Cluster 1 (generasi Z, pendapatan rendah) tergolong pelanggan untuk pengembangan umum dengan strategi pemasaran menggunakan influencer dan promosi menarik. Cluster 2 (generasi

milenial, pendapatan tinggi) adalah pelanggan untuk retensi umum dengan strategi melibatkan influencer dan ulasan positif online. Cluster 3 (generasi Z, pendapatan rendah) merupakan pelanggan yang hilang, di mana strategi yang efektif mencakup personalisasi pemasaran dan promosi khusus. Cluster 4 (generasi milenial, pendapatan tinggi) adalah pelanggan bernilai penting yang memerlukan layanan dan penawaran eksklusif untuk mempertahankan loyalitas mereka.

Dengan demikian, penelitian ini dapat memudahkan perusahaan dalam menentukan strategi pemasaran berdasarkan karakteristik dan tipe pelanggan dalam setiap kelompok. Selain itu penggunaan metode siku, siluet, dan statistik kesenjangan dapat menentukan jumlah cluster optimal dengan lebih akurat dan mengurangi risiko bias dari satu metode saja. Namun metode yang digunakan untuk menentukan jumlah cluster dapat menghasilkan variasi hasil jika diterapkan pada dataset yang berbeda atau dalam kondisi yang berbeda. Oleh karena itu, untuk analisis lebih lanjut dapat menggunakan data terbaru dan berkualitas tinggi untuk memastikan hasil yang lebih relevan dan akurat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] A. Prahendratno and G. S. Mahendra, *Business Intelegent*, no. June. 2023.
- [2] H. & Tahir, *Transformasi Bisnis di Era Digital (Teknologi Informasi dalam Mendukung Transformasi Bisnis di Era Digital)*, no. August. 2023.
- [3] Efori Buulolo, *Data Mining Untuk Perguruan Tinggi*. Sleman: Deepublish, 2020.
- [4] Rachma, "Analisis Data untuk Maksimalkan Peluang Bisnis di Marketplace," smkdev. Accessed: Jan. 21, 2023. [Online]. Available: <https://www.smk.dev/analisis-data-untuk-maksimalkan-peluang-bisnis-di-marketplace/>
- [5] L. L. K. Leon G. Schiffman, *Consumer Behavior*. Pearson Education, 2009.
- [6] Tiris Sudrartono, "Pengaruh Segmentasi Pasar Terhadap Tingkat Penjualan Produk Fashion Umk," *Coopetition J. Ilm. Manaj.*, vol. 10, no. 1, pp. 53–64, 2019, doi: 10.32670/coopetition.v10i1.40.
- [7] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan Hasil Metode Clustering K-Means, Db Scanner & Hierarchical Untuk Analisa Segmentasi Pasar," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, p. 191, 2023, doi: 10.26798/jiko.v7i2.796.
- [8] F. D. Wahyuningtyas, A. Arafat, A. Stiawan, and D. Rolliawati, "Komparasi Algoritma Hierarchical, K-Means, dan DBSCAN pada Analisis Data Penjualan Melalui Facebook," *Explor. J. Sist. Inf. dan Telemat.*, vol. 14, no. 1, p. 7, 2023, doi: 10.36448/jsit.v14i1.2931.
- [9] D. S. Alireza Bab-Hadiashar, *Data Segmentation and Model Selection for Computer Vision*. 2012.
- [10] S. N. Ashma et al., "Segmentasi Pelanggan Berdasarkan Keluhan Dengan Menggunakan K-Means Cluster Analysis Pada PT Infomedia Nusantara," pp. 276–280, 2020.
- [11] Mulaab, *Data Mining: Konsep dan Aplikasi*. Malang: Media Nusa Creative, 2017.
- [12] G. W. N. Dicky Nofriansyah, *Algoritma Data Mining dan Pengujian*. Sleman: Deepublish, 2019.
- [13] T. S. Madhulatha, "An overview of clustering methods," *IOSR J. Eng.*, vol. 2, no. 4, pp. 719–725, 2012, doi: 10.3233/ida-2007-11602.
- [14] S. Asmiatun, N. Wakhidah, A. N. Putri, U. Semarang, and K. Jalan, "Penerapan Metode K-Medoids Untuk Pengelompokkan Kondisi Jalan Di Kota Semarang 1,2," vol. 6, no. 2, 2020.
- [15] H. L. Ramadhania and L. Zakaria, "Aplikasi Metode Silhouette Coefficient, Metode Elbow dan Metode Gap Statistic dalam Menentukan K Optimal pada Analisis K-Medoids," vol. 04, no. 01, pp. 1–10, 2023.
- [16] F. A. Dewa and M. T. Jatipaningrum, "Segmentasi E-Commerce Dengan Cluster K-Means Dan Fuzzy C-Means (Studi Kasus: Media Sosial Di Indonesia yang diunduh di Play Store)," *J. Stat. Ind. dan Komputasi*, vol. 4, no. 1, pp. 53–67, 2019.
- [17] E. H.-V. Rocío González Martínez, Ramón Alberto Carrasco, Jesús García-Madariaga, Carlos Porcel Gallego, "A comparison between Fuzzy Linguistic RFM Model and traditional RFM model applied to Campaign Management. Case study of retail business," no. Itqm, pp. 0–8, 2019, doi: 10.1016/j.procs.2019.11.286.
- [18] K. Tsipstsis and A. Chorianopoulos, *Data Mining Techniques in CRM: Inside Customer Segmentation*. 2009.
- [19] Y. K. Jain and S. K. Bhandare, "International

- Journal of Computer and Communication Technology Min Max Normalization Based Data Perturbation Method for Privacy Protection,” vol. 4, no. 4, 2013, doi: 10.47893/IJCCT.2013.1201.
- [20] S. Azwar, “Asumsi-Asumsi Dalam Inferensi Statistika,” *Bul. Psikol.*, vol. 9, no. 1, pp. 8–17, 2015.
- [21] G. G. Agustin Garcia Asuero, Ana Sayago, “The Correlation Coefficient: An Overview,” *Crit. Rev. Anal. Chem.*, 2006, doi: 10.1080/10408340500526766.
- [22] N. I. Rumi, “USULAN STRATEGI PEMASARAN BERDASARKAN KRITERIA PRIORITAS MARKETING MIX DENGAN METODE MULTIDIMENSIONAL SCALLING (MDS) DAN ANALYTICAL HIERARCHY PROCESS (AHP) (STUDI KASUS DI UD.BUBUHAN),” 2023.
- [23] D. R. Liu and Y. Y. Shih, “Integrating AHP and data mining for product recommendation based on customer lifetime value,” *Inf. Manag.*, vol. 42, no. 3, pp. 387–400, 2005, doi: 10.1016/J.IM.2004.01.008.
- [24] K. L. K. Philip Kotler, *Manajemen Pemasaran*. Erlangga, 2009.
- [25] D. B. Saputra, “Implementasi Fuzzy C-Means Dan Model Rfm Untuk Segmentasi Pelanggan (Studi Kasus : Pt . Xyz) Implementation of Fuzzy C-Means and Rfm Model for Customer Segmentation (Case Study : Pt . Xyz) Implementasi Fuzzy C-Means Dan Model Rfm Untuk Segmentasi Pela,” 2018.
- [26] C. Patriajaya, “Karakteristik demografi sebagai dasar segmentasi dan target pasar untuk penetapan strategi bauran pemasaran pasar lem Dove di Surabaya [CD-ROM],” 2012.
- [27] N. K. B. Ela Elliyana, Tri Siwi Agustina, Habibie Kadir, Yuliani, Bayu Imanuddin, H. Mohdari, Febrisi Dwita, Y. S. Lidiana, Abdurrohman, and S. R. H. Gazali, E. Elia Ardyan, Lili Marlinah, *Dasar-dasar Pemasaran*. Malang: Ahlimedia Press, 2022.
- [28] Rizky and Rosalina Noor, “Pengaruh Harga, Pendapatan, dan Halal Awareness Terhadap Keputusan Pembelian Produk Kosmetik Impor Pada Konsumen Pengguna E-commerce Sociolla (Studi Pada Mahasiswa di Universitas Brawijaya),” *J. Ilm. Fak. Ekon. dan Bisnis*, pp. 1–14, 2020.
- [29] C. Marcus, “A practical yet meaningful approach to customer segmentation,” *J. Consum. Mark.*, 1998.
- [30] Y.-Y. Shih and C.-Y. Liu, “A method for customer lifetime value ranking — Combining the analytic hierarchy process and clustering analysis,” *J. Database Mark. Cust. Strateg. Manag.*, vol. 11, no. 2, pp. 159–172, 2003, doi: 10.1057/palgrave.dbm.3240216.
- [31] Suryana, “Metodologi Penelitian: Model Praktis Penelitian Kuantitatif dan Kualitatif,” 2010.
- [32] A. Alamsyah *et al.*, “Customer Segmentation Using the Integration of the Recency Frequency Monetary Model and the K-Means Cluster Algorithm,” *Sci. J. Informatics*, vol. 9, no. 2, pp. 189–196, 2022, doi: 10.15294/sji.v9i2.39437.
- [33] D. I. Panji Indra Pangestu, Teguh Iman Hermanto, “Analisis Segmentasi Pelanggan Berbasis Model Recency Frequency Dan Monetary (Rfm) Menggunakan Algoritma K-Means,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, 2023.