**IMPLEMENTASI MODEL *RECENCY*, *FREQUENCY*, *MONETARY* (RFM) DAN ALGORITMA K-MEANS DALAM SEGMENTASI PELANGGAN PADA GALERI MAGRAN LIVING**

**Fabio Rayhan1\*, Gandung Triyono2**

1,2 Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: 1\*2012500621@student.budiluhur.ac.id, 2gandung.triyono@budiluhur.ac.id

(\* : corresponding author)

**Abstrak-** Selama sebelas tahun berdiri, Galeri Magran *Living* belum menerapkan sebuah metode segmentasi pelanggan yang mendalam untuk mengetahui karakteristik pelanggan dari data yang ada. Kondisi ini dapat mengakibatkan pemasaran produk yang kurang tepat sasaran, dan dapat mengurangi kecenderungan pelanggan untuk berbelanja kembali di masa depan. Tujuan dari penelitian ini adalah untuk mengetahui ada berapa jumlah kelompok pelanggan dan bagaimana karakteristik pelanggan di tiap-tiap kelompoknya. Manfaat dari penelitian ini adalah untuk membantu pihak Galeri Magran *Living* khususnya bagian pemasaran untuk dapat merumuskan sebuah strategi pemasaran secara tepat agar dapat meningkatkan kecenderungan pelanggan dalam kembali berbelanja dan meningkatkan keuntungan. Penelitian ini menggunakan model RFM (*recency*, *frequency*, dan *monetary*) dan algoritma *clustering* K-Means, serta *elbow method* atau metode siku untuk pengujian jumlah *cluster* terbaik. Model RFM digunakan untuk bisa mengetahui skor dari kebiasaan pelanggan yang meliputi *recency*, *frequency*, dan *monetary*. Algoritma K-Means dipilih karena memiliki hasil *clustering* yang lebih baik dibanding algoritma lain khususnya pada segmentasi pelanggan. Hasil dari penelitian ini mendapatkan 4 buah *cluster* berdasarkan pengujian yang dilakukan menggunakan *elbow method*. *Cluster* 0 merupakan *cluster* dengan ciri-ciri frekuensinya cukup besar dan biaya yang dikeluarkan juga cukup banyak, tetapi cenderung jarang untuk kembali berbelanja. Akan diberikan label *Hibernating Customer*. *Cluster* 1 merupakan ini merupakan *cluster* terbaik. Pelanggan pada *cluster* ini sangat mungkin untuk kembali berbelanja, karena dari segi frekuensi berbelanja sangat sering, dan biaya yang dikeluarkan sangat banyak. Akan diberikan label *Loyal Customer*. *Cluster* 2 merupakan *cluster* terburuk. Pelanggan pada *cluster* ini kecil kemungkinannya untuk kembali, karena terakhir berbelanja sudah sangat lama, frekuensi yang sangat kecil, dan biaya yang dikeluarkan sangat sedikit. Akan diberikan label *Lost Customer*. *Cluster* 3 merupakan *cluster* khusus pelanggan baru dengan skor *recency* yang sangat tinggi, namun untuk frekuensi dan biaya yang dikeluarkan masih sedikit sekali. Akan diberikan label *New Customer*.

**Kata Kunci:** Segmentasi Pelanggan, Model RFM, Algoritma K-Means, *Elbow Method*

*IMPLEMENTATION OF THE RFM MODEL AND K-MEANS ALGORITHM IN CUSTOMER SEGMENTATION AT MAGRAN LIVING GALLERY*

**Abstract-** *For eleven years since its establishment, Magran Living Gallery has not implemented an in-depth customer segmentation method to understand customer characteristics from existing data. This situation can lead to less targeted product marketing and may reduce the tendency of customers to make repeat purchases in the future. The aim of this study is to determine the number of customer groups and the characteristics of customers in each group. The benefit of this research is to assist Magran Living Gallery, particularly the marketing department, in formulating an accurate marketing strategy to increase customer tendency to return for future purchases and to boost profits. This research utilizes the RFM model and the K-Means algorithm. The RFM model is used to assess customer habits scores, including recency, frequency, and monetary value. The K-Means algorithm was chosen because it provides better clustering results compared to other algorithms, particularly in customer segmentation. The results of this study identified four clusters based on tests conducted using the elbow method to determine the optimal number of clusters. Cluster 0 is a cluster characterized by a relatively high frequency and significant expenditure, but the customers in this cluster tend to rarely return for repeat purchases. This cluster will be labeled as Hibernating Customer. Cluster 1 is the best cluster. Customers in this cluster are highly likely to return for repeat purchases because they shop very frequently and spend a lot. This cluster will be labeled as Loyal Customer. Cluster 2 is the worst cluster. Customers in this cluster are unlikely to return, as they made their last purchase a long time ago, shop very infrequently, and spend very little. This cluster will be labeled as Lost Customer. Cluster 3 is a cluster specifically for new customers with a very high recency score, but with very low frequency and expenditure so far. This cluster will be labeled as New Customer.*

***Keywords****: Customer Segmentation, RFM Model, K-Means Algorithm, Elbow Method*

# PENDAHULUAN

Segmentasi pelanggan merupakan hal yang esensial pada suatu perusahaan, terutama untuk mengetahui peta persaingan bisnis. Perusahaan tidak hanya dituntut memanfaatkan strategi pemasaran produk, tetapi juga memanfaatkan strategi pemasaran pelanggan. Hal ini sangat berpengaruh guna menciptakan hubungan baik dengan pelanggan sehingga kepuasan dan loyalitas pelanggan dapat tercapai[1]. Dalam penyusunan strategi pemasaran pelanggan, perusahaan dapat memanfaatkan segmentasi pelanggan untuk mengidentifikasi karakteristik pelanggan dan mengelompokkan pelanggan berdasarkan perilaku atau kebiasaan yang sama. Akhir-akhir ini, kemampuan untuk menghasilkan informasi yang bermanfaat dari data telah ada menjadi isu penting dalam pengelolaan industri. Hal ini menekankan bahwa betapa pentingnya bagi pelaku industri untuk menerapkan teknik penggalian data guna mengungkap informasi tersembunyi mengenai pelanggan dari data pelanggan sebelumnya dan menentukan strategi pengelolaan hubungan pelanggan yang efektif[2].

Pemanfaatan data menjadi poin penting dalam keberlangsungan sebuah perusahaan dan faktor ini juga berlaku pada Galeri Magran *Living*, yang hingga saat ini juga bertumpu kepada pemanfaatan data. Namun sejak 11 tahun berdiri, Galeri Magran *Living* belum menerapkan segmentasi pelanggan yang mendalam untuk mengetahui karakteristik pelanggan dari data yang ada. Kondisi ini dapat mengakibatkan pemasaran produk yang kurang tepat sasaran, dan dapat mengurangi kecenderungan pelanggan untuk berbelanja kembali di masa depan[3]. Untuk mengatasi hal tersebut, perlu diterapkan metode yang dapat digunakan oleh Galeri Magran *Living* dalam menganalisa, mengelompokkan, dan menyusun strategi pemasaran berdasarkan tiap karakteristik pelanggan, maka dilakukan pengolahan data sehingga agar mendapatkan informasi yang tersembunyi dari data-data tersebut.

Beberapa penelitian telah dilakukan untuk menentukan model dan metode terbaik dalam segmentasi pelanggan. Penelitian terdahulu menyebutkan bahwa model yang banyak digunakan dan memiliki tingkat akurasi tinggi adalah *Recency*, *Frequency*, *Monetary* (RFM). Model ini menilai perilaku pelanggan berdasarkan waktu transaksi terakhir (*Recency*), jumlah transaksi (*Frequency*), dan jumlah uang yang dikeluarkan (*Monetary*)[4]. Dengan mempelajari rekaman interaksi dan transaksi pelanggan di masa lalu, perusahaan dapat mengevaluasi tiga aspek dimensi tersebut. RFM secara luas diterapkan pada *database* pemasaran dan merupakan alat umum untuk membangun strategi pemasaran[5].

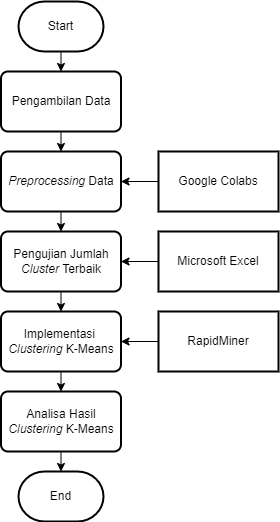
Dalam mengimplementasikan penambangan data, terdapat banyak teknik yang dapat dipilih, salah satunya adalah teknik *unsupervised learning*. Salah satu jenis teknik *unsupervised learning* yang paling sesuai dengan segmentasi adalah *clustering*. *Clustering* adalah teknik *unsupervised learning* dimana data dibagi ke dalam beberapa kelompok, data yang sama atau sejenis dalam satu kelompok lebih besar dibandingkan dengan kesamaan data tersebut dengan data dalam kelompok lain. Salah satu metode *clustering* yang sering digunakan adalah *Clustering* K-Means, *Clustering* K-Means adalah metode non-hierarki yang menggunakan teknik partisi untuk membagi atau memisahkan objek-objek ke dalam daerah-daerah terpisah [6].

Berdasarkan beberapa penelitian sebelumnya, metode RFM (*Recency*, *Frequency*,dan *Monetary)* dan *clustering* K-Means dapat dikombinasikan dan diimplementasikan dalam segmentasi pelanggan. Hasil dari *clustering* K-Means ini dapat memberikan hasil *clustering* dengan berbagai karakteristik pelanggan didalamnya. Hal ini dapat dijadikan referensi dalam menentukan strategi pemasaran pelanggan kepada Galeri Magran *Living* agar pemasaran produk dapat berjalan secara tepat sasaran.

# METODE PENELITIAN

# Tahapan Penelitian

Tahapan penelitian yang ditampilkan dalam bentuk diagram alir akan menjelaskan secara lengkap langkah demi langkah secara sistematis yang dilakukan dalam penelitian dari awal hingga akhir. Dapat dilihat pada gambar 1 berikut.



**Gambar 1.** Alur Penelitian

Tahap yang pertama dilakukan pengambilan data dengan melakukan wawancara untuk mendapatkan izin dan mengambil data historis melalui sumber internal Galeri Magran *Living*. Selanjutnya dilakukan *preprocessing* data mulai dari pengecekan *null value* dan pemilihan atribut serta menetapkan skor RFM untuk tiap pelanggan. Setelah itu melakukan pengujian untuk mencari jumlah *cluster* terbaik dengan *elbow method*. Lalu dilakukan implementasi *clustering* K-Means untuk melakukan pengelompokkan pelanggan berdasarkan skor RFM. Terakhir, melakukan analisa hasil *clustering* untuk melihat karakteristik dari tiap-tiap *cluster* yang didapat.

# Pengambilan Data

Wawancara dilakukan kepada dua anggota perusahaan yaitu koordinator HCGA dan manajer IT. Wawancara ini bertujuan untuk meminta izin melakukan penelitian tugas akhir, untuk mengetahui apakah terdapat permasalahan pada bagian pemasaran produk, dan untuk meminta izin untuk menggunakan data transaksi pelanggan perusahaan. Dari hasil wawancara, diharapkan dapat mengetahui dan mempelajari apa saja permasalahan yang akan disimpulkan pada rumusan masalah, bagaimana cara menyelesaikannya, dan menyampaikan solusi atau rekomendasi dari permasalahan tersebut.

Setelah memperoleh wawasan tentang permasalahan dan cara-cara penyelesaiannya, akan dilakukan tahapan pengumpulan dan pengambilan data. Data pada penelitian ini didapat melalui data historis atau *historical* data melalui sumber internal Galeri Magran *Living*.

# *Preprocessing* Data

Data yang telah dikumpulkan selanjutnya akan melalui proses preprocessing untuk memastikan apakah masih ada *noise* atau missing value. Pada penelitian ini akan menggunakan bahasa pemrograman *Python* dengan editor teks *Google Colab*. Pertama dilakukan pengecekan *missing value* dan yang kedua menghilangkan atribut yang tidak digunakan.

Berdasarkan topik penelitian ini yaitu segmentasi pelanggan, maka kolom yang tidak digunakan merupakan atribut yang tidak ada korelasinya yaitu: Prd, KdBarang, NmBarang, Qty, dan Sales. Sedangkan atribut yang digunakan yaitu: Date, CustomerID, SalesOrderID, dan Amount.

Setelah itu, data akan dimasukkan ke dalam *pivot table* di dalam *Microsoft Excel* untuk dilakukan implementasi RFM. Analisa ini mengidentifikasi pelanggan dari kumpulan data dengan menggunakan tiga atribut yaitu selisih waktu transaksi, frekuensi transaksi, dan jumlah uang yang dibelanjakan.

Nilai RFM dari semua atribut dapat ditentukan menggunakan skor dengan skala 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 yang mencerminkan kontribusi pelanggan terhadap keuntungan perusahaan. Skor 10 menunjukkan kontribusi yang sangat besar, sedangkan nilai 0 menunjukkan kontribusi yang sangat kecil. Pada penelitian ini akan menggunakan rumus “*PERCENTRANK.EXC*” yang merupakan cabang rumus “*RANK*” dengan menerapkan sistem persen (%) pada penghitungan peringkat. Sebelum dilakukan pengujian jumlah *cluster* optimal, data akan melewati proses normalisasi dengan metode *z-transformation* agar proses *clustering* dapat berjalan secara efisien[7].

# Pengujian Jumlah *Cluster*

Setelah mendapatkan skor RFM tiap pelanggan, pengujian jumlah *cluster* akan dilakukan dengan menggunakan *Elbow Method* atau metode siku dengan mencari rata-rata jarak dari titik centroid[8]. *Elbow Method* atau metode siku merupakan salah satu metode pengujian dalam *clustering*. Pada metode siku, nilai k optimal dapat diketahui dari penurunan inersia sehingga membentuk sebuah bentuk berupa siku-siku[9].

Penggunaan aplikasi *RapidMiner* dilakukan untuk mencari angka rata-rata jarak tersebut, dengan memasukkan data yang telah diimplementasi RFM dan memasukkan operator *clustering* K-Means serta penambahan “*Cluster Distance Performance Operator*”. Kemudian angka rata-rata jarak akan dimasukkan ke dalam *Microsoft Excel* untuk menerapkan grafik *Elbow Method*. Setelah grafik terlihat, maka akan dipilih mana jumlah *cluster* yang paling optimal dan jumlah tersebut akan diimplementasikan pada tahap selanjutnya.

# Implementasi *Clustering* K-Means

Teknik data *mining* telah diaplikasikan dalam berbagai bidang. Aktivitas transaksi jual beli yang terjadi secara rutin dalam sebuah organisasi menghasilkan volume data transaksi yang besar dan kompleks. Dengan menggunakan teknik data mining, terutama teknik *clustering*, memungkinkan untuk membagi seluruh pelanggan menjadi beberapa kelompok (*cluster*) berdasarkan karakteristik yang serupa.

*Clustering* K-Means adalah metode non-hierarki yang menggunakan teknik partisi untuk membagi atau memisahkan objek-objek ke dalam daerah-daerah terpisah[10]. *Clustering* K-Means menggunakan sebuah persamaan untuk mengukur jarak minimum antara objek-objek yang disebut *Euclidean Distance*. *Euclidean Distance* sebuah adalah komponen dari K-Means *Clustering* yang biasa digunakan untuk menghitung jarak antara objek-objek atau titik data. Rumus dari *Euclidean Distance* dapat dilihat pada persamaan (1) yaitu :

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Keterangan:

*D* = *Euclidean distance*

*X*1 = *Centroid* awal data A

*Y*1 = *Centroid* awal data B

*X*2 = Titik data A

*Y*2 = Titik data B

# Analisa Hasil *Clustering* K-Means

Langkah terakhir yaitu melakukan penggunaan terhadap visualisasi hasil *clustering* K-Means dalam bentuk *heat map* dan *scatter plot*. Analisa hasil *clustering* dilakukan berdasarkan visualisasi yang didapat dan akan dijadikan acuan dalam segmentasi. Dari hasil visualisasi, akan dianalisis apa saja karakteristik atau perilaku dari tiap-tiap *cluster* yang ada secara lengkap dan saksama.

# HASIL DAN PEMBAHASAN

# Pengambilan Data

Setelah melalui tahap wawancara kepada HRD dan manajer IT dari Galeri Magran Living dengan tujuan untuk meminta izin melakukan penelitian, dan menggunakan data internal. Dataset yang didapatkan merupakan data penjualan historis Galeri Magran Living dari tahun 2018 sampai dengan 2022. Di dalam *dataset* ini terdapat 2.296 data dengan 9 atribut dengan nama Date, CustomerID, SalesOrderID, Prd, KdBarang, Nmbarang, Qty, Amount, dan Sales. Serta berisi informasi-informasi penting dalam sebuah transaksi jual-beli seperti *furniture*, kompor, mesin pembuat kopi, dan lain sebagainya. Nama *file* dari *dataset* yang didapat adalah “Data Customer PAW 1 2018 – 2022” dengan tipe *file* .csv.

# *Preprocessing* Data

## 3.2.1 Memilih Atribut

Berdasarkan topik penelitian ini yaitu segmentasi pelanggan, maka kolom yang tidak digunakan merupakan atribut yang tidak ada korelasinya. Cuplikan dari beberapa data terdapat pada tabel 1 berikut.

Tabel 1. Pemilihan Atribut

|  |  |  |  |
| --- | --- | --- | --- |
| Date | CustomerID | SalesOrderID | Amount |
| 06/01/2018 | 10001 | SO1 | Rp7.897,000 |
| 06/01/2018 | 10001 | SO2 | Rp31.952,000 |
| 06/01/2018 | 10001 | SO3 | Rp30.296,000 |
| 06/01/2018 | 10001 | SO4 | Rp19.525,000 |
| 06/01/2018 | 10002 | SO5 | Rp32.253,000 |
| … | … | … | … |
| 24/12/2022 | 10789 | SO2289 | Rp14.378,000 |
| 24/12/2022 | 10289 | SO2295 | Rp29.926,000 |
| 24/12/2022 | 10289 | SO2294 | Rp28.406,000 |
| 24/12/2022 | 10289 | SO2293 | Rp14.255,000 |
| 24/12/2022 | 10289 | SO2292 | Rp23.721,000 |

Atribut-atribut yang dipilih yaitu: Prd, KdBarang, NmBarang, Qty, dan Sales. Sedangkan atribut yang digunakan yaitu: Date, CustomerID, SalesOrderID, dan Amount.

## 3.2.2 Implementasi RFM

Data akan dimasukkan ke dalam *pivot table* di dalam Microsoft Excel untuk dilakukan implementasi RFM. Pada analisa RFM, mengidentifikasi pelanggan dari kumpulan data dengan menggunakan tiga atribut yaitu selisih waktu transaksi, frekuensi transaksi, dan jumlah uang yang dibelanjakan. Maka, terdapat 3 nilai utama yang diperlukan yaitu menambahkan *max of date*, *count of sales order*, dan *sum of amount* pada *field values*. *Max of date* meliputi tanggal, bulan, dan tahun transaksi terakhir pelanggan. *Count of sales order* adalah jumlah transaksi suatu pelanggan. *Sum of amount* merupakan total harga tiap transaksi dalam rupiah pada satu pelanggan.

Nilai RFM dari semua atribut dapat ditentukan menggunakan skor dengan skala 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 yang mencerminkan kontribusi pelanggan terhadap keuntungan perusahaan. Pada penelitian ini akan menggunakan rumus “PERCENTRANK.EXC” yang merupakan cabang rumus “RANK” dengan menerapkan sistem persen (%) pada penghitungan peringkat. Hasil implementasi pada lima data pertama terdapat pada tabel 2 berikut.

Tabel 2. Implementasi RFM

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CustomerID | Max of Date | Count of SalesOrderID | Sum of Amount | Days Since Last Order | R | F | M |
| 10001 | 06/01/2021 | 6 | Rp112.616,000 | 724 | 7 | 8 | 9 |
| 10002 | 06/01/2018 | 1 | Rp32.253,000 | 1820 | 1 | 0 | 4 |
| 10003 | 10/01/2018 | 1 | Rp28.751,000 | 1816 | 1 | 0 | 3 |
| 10004 | 10/01/2020 | 5 | Rp71.212,000 | 1086 | 5 | 8 | 7 |
| 10005 | 10/01/2018 | 8 | Rp121.454,000 | 1816 | 1 | 9 | 9 |

Penjelasan dari skala tiap-tiap skor RFM yang didapat akan dijabarkan pada tabel 3 berikut. Skor 10 menunjukkan kontribusi yang sangat besar, sedangkan nilai 0 menunjukkan kontribusi yang sangat kecil.

Tabel 3. *Range* Skala Skor RFM

|  |  |  |  |
| --- | --- | --- | --- |
| Skor | R (hari) | F | M (Rp) |
| 0 | 1710-1820 | 1 | 2.000,000 - 10.000,000 |
| 1 | 1626-1709 | 2 | 11.000,000 - 17.000,000 |
| 2 | 1522-1625 | 3 | 18.000,000 - 24.000,000 |
| 3 | 1403-1521 | 4 | 25.000,000 - 29.000,000 |
| 4 | 1262-1402 | 5 | 38.000,000 - 30.000,000 |
| 5 | 1039-1261 | 6 | 39.000,000 - 48.000,000 |
| 6 | 839-1038 | 7 | 49.000,000 - 58.000,000 |
| 7 | 677-838 | 8 | 59.000,000 - 76.000,000 |
| 8 | 446-676 | 9 | 77.000,000 - 107.000,000 |
| 9 | 206-445 | 10 | 108.000,000 - 349.000,000 |
| 10 | 5-205 | 10-15 | 350.000,000 - 403.000,000 |

## 3.2.3 Normalisasi Data

Ketiga atribut yang sudah ditambah yaitu R, F, dan M akan melalui proses normalisasi data dengan menggunakan *z-tranformation*. Metode ini akan mentransformasikan atribut dimana nilai dari atribut A tertentu ditransformasikan berdasarkan rata-rata dan standar deviasi dari atribut A. Rumus dari metode ini terdapat pada persamaan (2) berikut.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Keterangan :

𝑣′ = *z-tranform*

𝑣 = nilai dari atribut A

α = nilai rata-rata dari atribut A

𝜎 = standar deviasi dari atribut A

Hasil perhitungan normalisasi *z-tranformation* pada lima data pertama terdapat pada tabel 4 berikut.

Tabel 4. Normalisasi Data

|  |  |  |  |
| --- | --- | --- | --- |
| CustomerID | R | F | M |
| 10001 | 0,524 | 1,404 | 1,566 |
| 10002 | -1,568 | -1,117 | -0,176 |
| 10003 | -1,568 | -1,117 | -0,524 |
| 10004 | -0,173 | 1,404 | 0,869 |
| 10005 | -1,568 | 1,719 | 1,566 |

Pada hasil normalisasi hasilnya bisa terdapat negatif dikarenakan pembagian yang dilakukan pada perhitungan normalisasi *z-tranformation*.

# Pengujian Jumlah *Cluster*

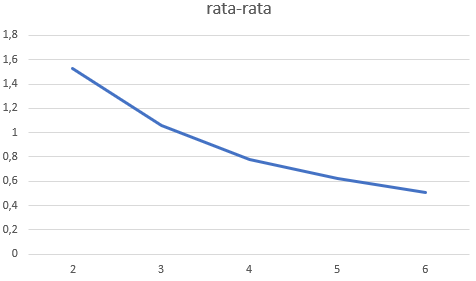
Pada tahap ini pertama menggunakan RapidMiner dengan pemanfaatan operator “*Cluster Distance Performance Operator*” untuk mencari angka rata-rata dalam jarak *centroid*. Di dalam operator “*Clustering* K-Means”, dimasukkan *cluster* dua pada kotak “*Min*” dan sepuluh pada kotak “*Max*” ini berarti pengujian yang akan dilakukan adalah *cluster* berjumlah 2, 3, 4, 5, 6.

Setelah angka rata-rata jarak didapat, angka tersebut akan dimasukkan kedalam Microsoft Excel untuk melihat *cluster* berapa yang merupakan *cluster* terbaik. Mula-mula dibuat dua buah kolom dengan nama “k” dan “rata-rata” lalu dimasukkan nilai-nilai yang dapat dilihat pada tabel 5 berikut.

Tabel 5. Angka Rata-Rata Jarak *Centroid*

|  |  |
| --- | --- |
| K | Rata- rata |
| 2 | 1,525 |
| 3 | 1,057 |
| 4 | 0,778 |
| 5 | 0,624 |
| 6 | 0,502 |

Angka-angka yang didapat merupakan hasil angka rata-rata dari jarak tiap *centroid*. Setelah tabel dibuat, selanjutnya akan dimasukkan grafik x,y untuk melihat seperti apa visualisasi dari angka tersebut. Hasil dari pengujian jumlah k terbaik terdapat pada gambar 2 berikut.



**Gambar 2.** Grafik *Elbow Method*

Untuk menentukan jumlah *cluster* optimal dari grafik elbow method, perlu mencari titik dengan laju penurunannya melambat secara drastis. Berdasarkan gambar 2, dapat disimpukan bahwa laju penurunan yang drastis terjadi setelah titik ke-4. Oleh karena itu, jumlah *cluster* optimal yang akan digunakan adalah 4 *cluster*.

# Implementasi *Clustering* K-Means

Tahap pertama merupakan penentuan jumlah *cluster* (k). Berdasarkan pengujian yang telah dilakukan dengan penggunaan *elbow method*, telah didapat jumlah *cluster* yang paling optimal yaitu 4, maka dalam penghitungan ini akan menggunakan 4 *cluster* dimana k=4. Tahap kedua adalah menentukan nilai *centroid* awal secara acak dari titik data yang ada seperti pada tabel 6 berikut.

**Tabel 6.** Inisialisasi *Cluster*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Atribut | C0 | C1 | C2 | C3 |
| R | 7 | 2 | 1 | 10 |
| F | 8 | 7 | 0 | 9 |
| M | 9 | 7 | 2 | 9 |

Tahap ketiga adalah menghitung jarak tiap titik data terhadap titik *centroid* yang telah ditentukan. Penghitungan jarak akan dilakukan dengan menggunakan *Euclidean Distance* untuk mencari jarak dari tiap titik dengan titik *centroid*. Proses pencarian jarak untuk ID pelanggan pertama (10001) dengan skor RFM 7, 8, dan 9 dapat dilihat pada persamaan (3) berikut.

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Tahap keempat adalah memasukkan titik data ke dalam *cluster*, pada contoh pelanggan pertama sebelumnya masuk ke dalam *cluster* 0 (C0) karena jaraknya nol (0). Tahap kelima adalah menghitung nilai *centroid* baru berdasarkan rata-rata dari setiap data di dalam *cluster*. Setelah didapat hasil jarak serta anggota tiap *cluster*, selanjutnya dilakukan penghitungan rata-rata dari setiap data di dalam *cluster* untuk mencari nilai *centroid* baru yang akan digunakan pada iterasi berikutnya. Proses perhitungannya dapat dilihat pada persamaan (4) berikut.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Tahap keenam dan terakhir adalah mengulang langkah ketiga hingga kelima. Langkah ini dimaksudkan untuk mengulang kembali langkah sebelumnya sampai tidak ada perubahan anggota *cluster*. Pada tahap ini, terdapat dua iterasi tambahan hingga tidak ada perubahan anggota *cluster*. Pada iterasi kedua ketika digunakan nilai *centroid* yang baru, tetapi masih terdapat perbedaan anggota *cluster*. Bagian ini juga menggunakan aplikasi bernama *RapidMiner* untuk memudahkan dalam pengimplementasian *clustering* K-Means. Hasil yang didapat ada pada tabel 7 berikut.

Tabel 7. Jumlah Pelanggan Tiap *Cluster*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | CustomerID | R | F | M | Cluster |
| 1 | 10001 | 7 | 8 | 9 | 0 |
| 2 | 10002 | 1 | 0 | 4 | 2 |
| 3 | 10003 | 1 | 0 | 3 | 2 |
| 4 | 10004 | 5 | 8 | 7 | 0 |
| 5 | 10005 | 1 | 9 | 9 | 0 |
| … | … | … | … | … | … |
| 788 | 10788 | 10 | 8 | 6 | 1 |
| 789 | 10789 | 10 | 5 | 5 | 1 |

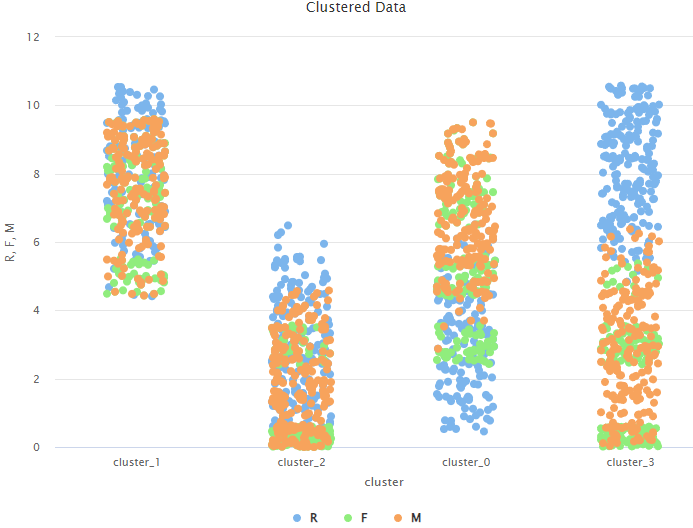
# Analisa Hasil *Clustering* K-Means

Berdasarkan proses *clustering* yang telah dilakukan dengan algoritma K-Means, jumlah *cluster* terbentuk sebanyak 4 *cluster*. *Cluster* dengan jumlah terbesar terdapat pada *cluster* 2 sebanyak 216 pelanggan dengan persentase 28%. *Cluster* 3 memiliki 203 pelanggan dengan persentase 26%. *Cluster* 0 sebanyak 193 pelanggan dengan persentase 24%. Terakhir *cluster* terkecil terdapat pada *cluster* 1 dengan 177 pelanggan dan persentase sebesar 23%. Para pelanggan akan dikelompokkan sesuai dengan *cluster* mereka masing-masing. Pada tabel 8 akan diperlihatkan jumlah pelanggan dan persentase dari tiap *cluster*.

Tabel 8. Jumlah Pelanggan Tiap *Cluster*

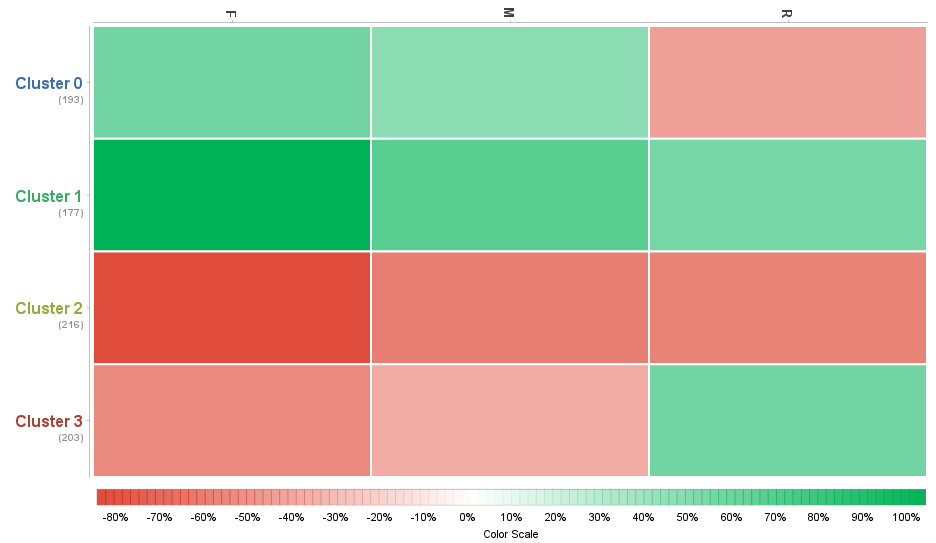
|  |  |  |
| --- | --- | --- |
| Cluster | Jumlah Pelanggan | Persentase |
| 0 | 193 | 24% |
| 1 | 177 | 22% |
| 2 | 216 | 28% |
| 3 | 203 | 26% |

Tahap ini akan menggunakan tiga bentuk visualisasi sebagai acuan analisa. Pertama digunakan *scatter plot* dan *heat map* untuk melihat hubungan antar atribut R, F, dan M dalam *clustering* yang telah dilakukan. Visualisasi scatter plot dapat dilihat pada gambar 3 berikut.



**Gambar 3.** *Scatter Plot Clustering* K-Means

Titik biru menyatakan *Recency* (R), titik hijau menyatakan *Frequency* (F), dan titik oranye menyatakan *Monetary* (M). Terlihat bahwa *cluster* 1 dan *cluster* 2 persebarannya cukup rapat sedangkan *cluster* 0 dan *cluster* 3 persebarannya cukup lebar.



**Gambar 4.** *Heat Map Clustering* K-Means

Selanjutnya ada visualisasi dalam bentuk *heat map* dengan tiga buah warna. Hijau berarti terbaik, putih berarti netral, merah berarti terburuk. Terlihat bahwa *cluster* 1 merupakan *cluster* dengan hijau di ketiga atribut yang artinya *cluster* ini adalah *cluster* terbaik. Setelah itu ada *cluster* 0, dimana pada *cluster* ini terdapat hijau di atribut F dan M sedangkan atribut R merah. *Cluster* 3 adalah *cluster* yang terdapat satu hijau pada atribut R tetapi F dan M menunjukkan warna merah. Terakhir adalah *cluster* 2 yang merupakan *cluster* terburuk dengan ketiga atribut menujukkan warna merah. Penjelasan lengkap mengenai karakteristik tiap pelanggan ada pada tabel 9 berikut.

Tabel 9. Deskripsi Karakteristik Pelanggan

|  |  |
| --- | --- |
| Cluster | Karakteristik Pelanggan |
| 0 | * Pelanggan pada *cluster* ini frekuensinya cukup besar dan biaya yang dikeluarkan juga cukup banyak, tetapi jarang untuk kembali berbelanja. * Skor rata-rata R adalah 3,456 * Skor rata-rata F adalah 5,585 * Skor rata-rata M adalah 6,617 |
| 1 | * *Cluster* ini merupakan *cluster* terbaik. Pelanggan pada *cluster* ini sangat mungkin untuk kembali, dari segi frekuensi berbelanja sangat besar, dan biaya yang dikeluarkan sangat banyak. * Skor rata-rata R adalah 7,994 * Skor rata-rata F adalah 7,232 * Skor rata-rata M adalah 7,582 |
| 2 | * *Cluster* ini merupakan *cluster* terburuk. Pelanggan pada *cluster* ini sangat tidak mungkin untuk kembali, frekuensi yang sangat kecil, dan biaya yang dikeluarkan sangat sedikit. * Skor rata-rata R adalah 2,852 * Skor rata-rata F adalah 0,556 * Skor rata-rata M adalah 1,769 |
| 3 | * Pelanggan pada *cluster* ini merupakan pelanggan baru dengan skor *recency* yang sangat tinggi, namun untuk frekuensi dan biaya yang dikeluarkan masih sedikit sekali. * Skor rata-rata R adalah 8,074 * Skor rata-rata F adalah 1,571 * Skor rata-rata M adalah 2,729 |

# KESIMPULAN

Berdasarkan analisa dan pembahasan mengenai segmentasi pelanggan menggunakan model RFM dan *clustering* K-Means pada data transaksi Galeri Magran Living, maka dapat disimpulkan bahwa dapat menentukan jumlah *cluster* ideal yaitu 4 buah *cluster* dengan berbagai macam karakteristik. *Cluster* 0 merupakan *cluster* dengan ciri-ciri frekuensinya cukup besar dan biaya yang dikeluarkan juga cukup banyak, tetapi cenderung jarang untuk kembali berbelanja. Akan diberikan label *Hibernating Customer*. *Cluster* 1 merupakan ini merupakan *cluster* terbaik. Pelanggan pada *cluster* ini sangat mungkin untuk kembali berbelanja, karena dari segi frekuensi berbelanja sangat sering, dan biaya yang dikeluarkan sangat banyak. Akan diberikan label *Loyal Customer*. *Cluster* 2 merupakan *cluster* terburuk. Pelanggan pada *cluster* ini kecil kemungkinannya untuk kembali, karena terakhir berbelanja sudah sangat lama, frekuensi yang sangat kecil, dan biaya yang dikeluarkan sangat sedikit. Akan diberikan label *Lost Customer*. *Cluster* 3 merupakan *cluster* khusus pelanggan baru dengan skor *recency* yang sangat tinggi, namun untuk frekuensi dan biaya yang dikeluarkan masih sedikit sekali. Akan diberikan label *New Customer*.

# DAFTAR PUSTAKA

[1] I. Maskanah, “Segmentasi Pelanggan Toko Purnama dengan Algoritma K-Means dan Model RFM untuk Perancangan Strategi Pemasaran,” *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, pp. 218-288, , 2020, doi: 10.35314/isi.v5i2.1443.

[2] S. Jun Lee and K. Siau, “A review of data mining techniques,” *Ind. Manag. Data Syst.*, vol. 101, no. 1, pp. 41–46, 2001.

[3] A. Satriawan, R. Andreswari, and O. N. Pratiwi, “Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan RFM Model dan Algoritma K-Means,” *e-Proceeding Eng.*, vol. 8, no. 2, pp. 2876–2883, 2021, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/14687

[4] K. Z. Wijaya, A. Djunaidi, and F. Mahananto, “Segmentasi Pelanggan Menggunakan Algoritma K-Means dan Analisis RFM di Ova Gaming E-Sports Arena Kediri,” *J. Tek. ITS*, vol. 10, no. 2, pp. 230- 237, 2021, doi: 10.12962/j23373539.v10i2.67707.

[5] A. T. Widiyanto and A. Witanti, “Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global),” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 204–215, 2021, doi: 10.24002/konstelasi.v1i1.4293.

[6] D. Zheng, “Application of silence customer segmentation in securities industry based on fuzzy cluster algorithm,” *J. Inf. &COMPUTATIONAL Sci.*, vol. 10, no. 13, pp. 4337–4347, 2013.

[7] E. Barannikova, “Digital Marketing Strategy Bachelor of Business Administration,” 2023.

[8] E. Umargono, J. E. Suseno, and V. G. S. K, “K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula,” Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019), vol. 474, 2019, pp. 121–129, 2020.

[9] R. H. Zakariyya, “Customer Segmentation by using RFM Model and K-Mean Clustering in PT XYZ,” *Telkom Univ*, pp. 1–10, 2020.

[10] S. Sajidah, R. Herdiana, and D. Solihudin, “Segmentasi Pelanggan Salon Nuii Beauty Glow Menggunakan K-Means Clustering,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 558–566, 2023, doi: 10.36040/jati.v7i1.6333.