

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Studi literatur dilakukan untuk mendapatkan pedoman dalam melakukan sebuah penelitian, meskipun begitu ada beberapa perbedaan yang ada di dalam studi literatur pada penelitian terkait mulai dari metode yang dipakai serta sistem yang dibuat. Berikut ini merupakan penelitian sebelumnya yang saling berkaitan dengan tema yang akan diangkat oleh peneliti diantaranya yaitu :

Penulis melakukan studi literatur pada penelitian milik Aji Setiono Agung dkk di tahun (2023) dengan judul “Analisis *Recency Frequency Monetary* Dan *K-Means Clustering* Pada Klinik Gigi Untuk Menentukan Segmentasi Pasien”. Peneliti ingin membuat pasien lebih banyak dan kepuasan pelayanan terhadap pasien meningkat, maka perusahaan harus mempunyai strategi. Palapa Dentists belum mengadopsi strategi CRM (*Customer Relationship Management*) masih memperlakukan semua pasien dengan pendekatan yang sama. Penelitian ini menggunakan metode RFM (*Recency Frequency Monetary*) dan *K-Means Clustering* dengan tujuan menentukan segmentasi pasien dan memilih kelompok pasien mana yang paling menguntungkan bagi perusahaan. Penentuan jumlah cluster menggunakan *elbow method* yang menghasilkan jumlah *cluster* terbaik adalah 2. *Silhouette score* menghasilkan jumlah 2 *cluster* dengan *score* 0.6014345457538962. Sedangkan hasil *davies-bouldin score* menunjukan cluster optimal dengan 3 *cluster* tapi skornya 0.7500785223208264 masih jauh dari 0. *Cluster* 1 memiliki 17.413 anggota dan *cluster* 2 memiliki 2.068 anggota. *Cluster* 1 memiliki nilai rata-rata *recency* 641,63, *frequency* 3,21, dan *monetary* Rp. 2.424.251,98. Sedangkan *cluster* 2 memiliki nilai rata-rata *recency* 286,87, *frequency* 19,32, dan *monetary* Rp. 20.087.467,49. Dapat disimpulkan *cluster* 2 adalah kelompok pasien yang lebih menguntungkan dibandingkan *cluster* 1 (Aji Setiono dkk 2023).

Setelah itu peneliti melakukan studi literatur pada penelitian milik Sri Wahyuni dkk (2023) dengan judul “Segmentasi Pelanggan Berdasarkan Analisis *Recency, Frequency, Monetary* Menggunakan Algoritma K-Means Pada CV. Toedjoe Sinar Group”. Peneliti ingin membuat percetakan pada CV. Toedjoe Sinar Group ini tidak hanya fokus pada strategi yang mengutamakan produk, tetapi juga melakukan strategi yang mengutamakan pelanggan. Hal ini diperlukan untuk mengetahui perilaku pelanggan sehingga akan membantu dalam penerapan strategi pemasaran yang tepat untuk meningkatkan pendapatan perusahaan. Sebanyak 8.020 data transaksi pelanggan dari Bulan Januari hingga Maret tahun 2022 digunakan dalam penelitian ini. Pemodelan data yang digunakan dalam pengelompokan pelanggan adalah *Recency, Frequency, Monetary* (RFM) dan metode *clustering* K-Means melalui teknik data mining. Penentuan jumlah cluster berdasarkan nilai k menggunakan metode *elbow*. Penelitian ini menghasilkan nilai  $k=7$  sehingga *cluster* yang terbentuk sebanyak 7 *cluster*. Kelompok terbaik yaitu *cluster* 5 dengan nilai rata-rata *Recency* sebesar 0,0, *Frequency* sebesar 140,33 serta rata-rata *monetary* sebesar 4.3328.833 (Sri Wahyuni dkk 2023).

Selanjutnya peneliti melakukan studi literatur pada penelitian dengan judul “Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means *Clustering* pada Model Data RFM”. Peneliti melakukan pengembangan strategi bisnis intelijen untuk UMKM, serta merancang strategi pendampingan yang tepat sasaran bagi lembaga pendamping UMKM dan dikorelasikan dengan strategi *Customer Relationship Management* (CRM). Data transaksi diolah ke dalam bentuk model data (RFM). Segmentasi pelanggan dilakukan dengan proses klasterisasi dengan algoritma K Means, dan hasil dari segmentasi ini menghasilkan kelompok profil dari pelanggan UMKM. Proses evaluasi klaster optimum menggunakan *Elbow Method*, *Silhouette Scores*, dan Davies-Bouldin Index. Hasil evaluasi menunjukkan bahwa klaster optimum adalah 3 dengan nilai *Silhouette Scores* terbaik yaitu 0.548, dan Davies-Bouldin Index dengan nilai 0.76. Segmen pelanggan pertama memiliki *frequency* belanja tertinggi dan nilai

monetary tertinggi, sehingga dapat dikategorikan sebagai segmen pelanggan aktif dan profit, segmen ini perlu diberikan layanan loyalitas khusus. Segmen kedua adalah segmen dengan jumlah pelanggan terbanyak meski *frequency* belanja hanya 1 –2 kali namun rata-rata baru berbelanja atau *recency* sekitar 2 bulan terakhir, pelanggan segmen ini wajib diberikan layanan purna jual yang baik. Segmen ketiga adalah segmen pelanggan yang waktu berbelanja sudah lebih dari 6 bulan lalu, segmen ini dapat dikategorikan sebagai segmen prioritas akhir untuk didekati kembali, misal dengan menggunakan teknik email marketing. Lembaga pendamping dapat berfokus pada pendampingan CRM untuk menyasar 3 segmen tersebut (Sisilia Fhelly Djun dkk 2024).

Studi literatur selanjutnya dilakukan pada penelitian dengan judul “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi *Online Retail*”. Peneliti ingin melakukan segmentasi pelanggan berdasarkan data, banyak model dan metode *clustering* menggunakan algoritma K-means menjadi salah satu pilihan yang efektif. Metode ini telah menjadi tren dan banyak digunakan dimana hal tersebut dibuktikan dengan banyak nya jurnal terkait dari rentang tahun 2018 - 2022. Penelitian ini menggunakan pemrograman python untuk proses data mining dan *pre-processing* yang dilakukan pada data melalui *exploratory data analysis* untuk memahami informasi dari data yang digunakan sebelum melakukan klasterisasi. Dalam penerapan metode K-means, digunakan metode *elbow* untuk menentukan jumlah klaster yang optimal. Hasil dari metode elbow menunjukkan bahwa penggunaan 4 klaster adalah pilihan yang tepat dalam kasus ini. Selanjutnya, pemodelan K-means dengan 4 klaster dilakukan menggunakan variabel *quantity*, *unit price*, dan *customer id*, dan menghasilkan 4 klaster yang berbeda dengan karakteristik yang spesifik pada masing-masingnya. dapat diamati bahwa kuantitas dan harga satuan berperan penting dalam mempengaruhi perilaku pelanggan (Eriskiannisa Febrianty dkk 2023).

Setelah itu peneliti melakukan studi literatur dengan judul “Penerapan Metode K-Means Dengan Metode *Elbow* Untuk Segmentasi Pelanggan Menggunakan Model RFM (*Recency, Frequency & Monetary*)”. Peneliti menerapkan model RFM dan mengelompokkan pelanggan menggunakan metode K-Means menjadi beberapa *Cluster*. Dalam menentukan jumlah *Cluster* yang optimal maka diterapkan metode *Elbow*. Algoritma yang digunakan dalam pembentukan *Cluster* terdiri dari *Euclidean Distance* dan *Manhattan Distance*. Kedua algoritma akan dibandingkan kualitas pembentukan *Clusternya* menggunakan metode *Silhouette Coefficient*. Hasil yang diberikan pada penelitian ini berupa data yang terbagi atas 5 kelompok dengan dilakukannya lima kali pengujian untuk menentukan *centroid* yang unggul. *Cluster* yang unggul akan dibuatkan visualisasi datanya untuk memudahkan perusahaan dalam mengambil keputusan. Berdasarkan penerapan *Silhouette Coefficient*, algoritma yang lebih unggul yaitu *Manhattan Distance* dengan nilai  $s(i)$  sebesar 0.152695 (Adnan Burhan Hidayat Kia dkk 2020).

Studi literatur berikutnya dilakukan pada penelitian dengan judul “Analisis Segmentasi Pelanggan Menggunakan K-Means *Clustering* studi Kasus Aplikasi Alfagift”. Peneliti menentukan segmentasi pelanggan, dilakukan menggunakan algoritma K-Means *clustering*. Proses *clustering* dilakukan dengan mengelompokkan data histori transaksi pelanggan pengguna aplikasi Alfagift selama bulan Juni 2021 dengan lima kategori yaitu umur, jenis kelamin, frekuensi pembelian, tipe pembayaran dan kota. Proses analisis segmentasi pelanggan menggunakan metode *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) yang meliputi enam tahapan yaitu pemahaman bisnis (*Business Understanding*), pemahaman data (*Data Understanding*), persiapan data (*Data Preparation*), pemodelan (*Modeling*), evaluasi (*Evaluation*), dan penerapan (*Deployment*). Algoritma K-Means digunakan untuk membentuk kluster dan metode *Elbow* digunakan untuk mengevaluasi kluster-kluster yang terbentuk. Hasil  $k$  optimum yang terbentuk dari hasil perhitungan SSE (*Sum*

*Squared Error*) dengan metode *Elbow* adalah tiga cluster dengan selisih nilai SSE terbesar yaitu 1119.28644. Hasil pengelompokan pelanggan sebanyak tiga *cluster*; *cluster* pertama berjumlah 7.219 pelanggan, *cluster* kedua sebanyak 6.902 pelanggan, dan *cluster* ketiga sebanyak 5.371 pelanggan. Hasil K-Means clustering tersebut selanjutnya diinterpretasikan sebagai data yang akan digunakan dalam menentukan strategi pemasaran yang tepat (Satria Ardi Perdana dkk 2022).

Setelah itu peneliti melakukan studi literatur dengan judul “Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means *Clustering Segmentation of New Student Admission Promotion Locations Using RFM and K-Means Clustering*”. Peneliti melakukan segmentasi tingkat potensial sekolah sebagai strategi untuk memilih lokasi promosi penerimaan mahasiswa baru Universitas Bumigora menggunakan analisis model RFM dan metode K-means. Tahapan penelitian terdiri dari persiapan data penerimaan mahasiswa baru tahun 2019 dan 2020, prapengolahan data, penerapan model Recency (R), Frequency (F), dan Monetary (M), implementasi metode K-means, dan analisa hasil. Hasil penelitian ini adalah terbentuk 3 klaster tingkat potensial sekolah yang dapat dijadikan skala prioritas untuk lokasi promosi penerimaan mahasiswa baru Universitas Bumigora yaitu kurang potensial, potensial, dan sangat potensial. Klaster sangat potensial (C2) terdapat 28 sekolah, klaster potensial (C3) terdapat 90 sekolah, dan klaster kurang potensial (C1) terdapat 152 sekolah (Dyah Susilowati dkk 2022).

Penelitian selanjutnya yang digunakan sebagai studi literatur yaitu pada penelitian dengan judul “Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi *E-Commerce*”. Peneliti ingin mengungkap tentang segmentasi atau kelompok pelanggan berdasarkan kesamaan karakteristik dan perilaku pelanggan masing-masing. Salah satu teknik yang dapat digunakan untuk mensegmentasi pelanggan adalah Data Mining dengan menggunakan metode *clustering* menggunakan algoritma K-Means dan K-Medoids. Hasil penelitian menunjukkan bahwa algoritma K-Means dan K-Medoids sama-

sama menunjukkan hasil cluster optimal  $k = 3$ . Hasil tersebut juga sesuai dengan hasil metode *elbow* dan uji validitas Davies Bouldin Index yang menunjukkan bahwa jumlah cluster optimal adalah 3 (tiga). Hasil pengujian menunjukkan K-Medoids memiliki performa terbaik dengan nilai rasion sebesar 0,337575 dibandingkan K-Means 0,3380724, sehingga K-Medoids digunakan dalam clustering data sebagai *cluster* optimal. Hasil segmentasi pelanggan sesuai *Customer Loyalty Matrix* terdiri dari *core customer*, *new customers*, dan *lost customer* (Romadansyah Siagian dkk 2022).

Landasan teori merupakan bagian dalam mengumpulkan literatur dari jurnal yang kemudian digunakan sebagai landasan dalam melakukan penelitian ini, dimana studi literatur masih relevan jika dihitung dengan jarak maksimal 5 tahun. Pada tabel 2.1 menunjukkan daftar studi literatur yang digunakan sebagai landasan teori dalam melakukan penelitian ini.

Tabel 2.1 Penelitian Terdahulu

| No | Peneliti                                                    | Judul Penelitian                                                                                                             | Metode Penelitian                                                       | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----|-------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Aji Setiono Agung Triayudi, Endah Tri Esti Handayani (2023) | Analisis <i>Recency Frequency Monetary</i> Dan K-Means <i>Clustering</i> Pada Klinik Gigi Untuk Menentukan Segmentasi Pasien | Metode RFM ( <i>Recency Frequency Monetary</i> ) dan K-Means Clustering | Penentuan klaster menggunakan metode elbow menghasilkan 2 klaster terbaik dengan Silhouette score 0,6014345457538962. Davies-Bouldin score menunjukkan klaster optimal adalah 3, namun skornya 0,7500785223208264 masih jauh dari ideal. Klaster 1 berisi 17.413 anggota dengan rata-rata <i>recency</i> 641,63, <i>frequency</i> 3,21, dan <i>monetary</i> Rp. 2.424.251,98. Klaster 2 berisi 2.068 anggota dengan <i>recency</i> 286,87, <i>frequency</i> 19,32, dan <i>monetary</i> Rp. 20.087.467,49. Klaster 2 terbukti lebih menguntungkan dibanding klaster 1. |

| No | Peneliti                                                       | Judul Penelitian                                                                                                                                         | Metode Penelitian                                                                                                                            | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|----|----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2  | Sri Wahyuni, Tina Tri Wulansari, Fahrullah (2023)              | Segmentasi Pelanggan Berdasarkan Analisis <i>Recency</i> , <i>Frequency</i> , <i>Monetary</i> Menggunakan Algoritma K-Means Pada CV. Toedjoe Sinar Group | Pemodelan data yang digunakan dalam pengelompokan pelanggan adalah RFM dan metode clustering K-Means melalui teknik data mining <i>Elbow</i> | Penelitian ini menganalisis 8.020 data transaksi pelanggan dari Januari hingga Maret 2022 menggunakan model RFM dan metode clustering K-Means. Metode elbow menentukan jumlah optimal kluster sebanyak 7. Kluster terbaik adalah kluster 5 dengan rata-rata <i>recency</i> 0,0, <i>frequency</i> 140,33, dan <i>monetary</i> Rp 4.332.833.                                                                                                                       |
| 3  | Sisilia Fhelly Djun, I Gede Aris Gunadi, dan Sariyasa (2024)   | Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means <i>Clustering</i> pada Model Data RFM                                        | Metode RFM, K-means clustering, Elbow Method, Silhouette Scores, dan Davies-Bouldin Index                                                    | Evaluasi menunjukkan kluster optimal adalah 3, dengan Silhouette Score 0,548 dan Davies-Bouldin Index 0,76. Segmen pertama memiliki <i>frequency</i> dan <i>monetary</i> tertinggi, segmen kedua mencakup pelanggan terbanyak dengan <i>frequency</i> rendah namun <i>recency</i> sekitar 2 bulan terakhir, dan segmen ketiga terdiri dari pelanggan yang terakhir berbelanja lebih dari 6 bulan lalu. Pendampingan CRM dapat difokuskan pada ketiga segmen ini. |
| 4  | Eriskiannisa Febrianty Luchia Awalina, Woro Isti Rahayu (2023) | Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi <i>Online Retail</i>                 | Klasterisasi metode K-means, dan metode elbow                                                                                                | Metode elbow menunjukkan bahwa 4 kluster adalah jumlah yang optimal. Pemodelan K-means dengan 4 kluster menggunakan variabel <i>quantity</i> , <i>unit price</i> , dan <i>customer id</i> menghasilkan 4 kluster dengan karakteristik unik. Dari hasil ini, terlihat bahwa kuantitas dan harga satuan mempengaruhi perilaku pelanggan.                                                                                                                           |

| No | Peneliti                                                                                                | Judul Penelitian                                                                                                                                                                                    | Metode Penelitian                                                                                                                                                                                                                                            | Hasil                                                                                                                                                                                                                                                                                                                 |
|----|---------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5  | Adnan Burhan Hidayat Kiat, Yufis Azhar, Vinna Rahmayanti (2020)                                         | Penerapan Metode K-Means Dengan Metode <i>Elbow</i> Untuk Segmentasi Pelanggan Menggunakan Model RFM ( <i>Recency, Frequency &amp; Monetary</i> )                                                   | Penerapan model RFM, dikelompokkan menggunakan metode K-Means. Dalam menentukan jumlah Cluster yang optimal maka diterapkan metode <i>Elbow</i> . Algoritma yang digunakan dalam pembentukan Cluster terdiri dari Euclidean Distance dan Manhattan Distance. | Penelitian ini menghasilkan data yang terbagi dalam 5 kelompok setelah lima kali pengujian untuk menentukan centroid terbaik. Cluster unggul akan divisualisasikan untuk membantu keputusan perusahaan. Berdasarkan <i>Silhouette Coefficient</i> , algoritma Manhattan Distance unggul dengan nilai $s(i)$ 0.152695. |
| 6  | Ardi Perdana, Sara Famayla Florentin, Agus Santoso (2022)                                               | Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering studi Kasus Aplikasi "Alfagift"                                                                                                        | Metode CRISP-DM, K-means clustering, metode <i>elbow</i> dan perhitungan SSE                                                                                                                                                                                 | Pengelompokan pelanggan menghasilkan tiga cluster: cluster pertama dengan 7.219 pelanggan, cluster kedua 6.902 pelanggan, dan cluster ketiga 5.371 pelanggan. Hasil K-Means ini akan digunakan untuk menentukan strategi pemasaran yang tepat.                                                                        |
| 7  | Dyah Susilowati, Hairani Hairani, Indah Puji Lestari, Khairan Marzuki, Lalu Zazuli Azhar Mardedi (2022) | Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means Clustering <i>Segmentation of New Student Admission Promotion Locations Using RFM and K-Means Clustering</i> | Analisis model RFM dan metode K-means. Menggunakan metode <i>Elbow</i> dengan bantuan python dan tools rapidminer.                                                                                                                                           | Penelitian ini menghasilkan 3 klaster tingkat potensial sekolah untuk prioritas promosi penerimaan mahasiswa baru Universitas Bumigora: kurang potensial, potensial, dan sangat potensial. Klaster sangat potensial (C2) terdiri dari 28 sekolah, potensial (C3) 90 sekolah, dan kurang potensial (C1) 152 sekolah.   |

| No | Peneliti                                               | Judul Penelitian                                                                                           | Metode Penelitian                                                                                                          | Hasil                                                                                                                                                                                                                                                                                                                                          |
|----|--------------------------------------------------------|------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 8  | Romadansyah Siagian, Pahala Sirait, Arwin Halim (2022) | Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi <i>E-Commerce</i> | Metode clustering menggunakan algoritma K-Means dan K-Medoids. Uji validitas dengan metode elbow dan Davies Bouldin Index. | Hasil pengujian menunjukkan K-Medoids memiliki performa terbaik dengan nilai rasio 0,337575, sedikit lebih rendah dari K-Means yang 0,3380724. Oleh karena itu, K-Medoids digunakan sebagai metode clustering optimal. Segmentasi pelanggan berdasarkan Customer Loyalty Matrix terdiri dari core customer, new customers, dan lost customers. |

Pembaruan dari penelitian ini lebih berfokus pada kelas pelanggan beserta treatment yang dapat diberikan dari perusahaan kepada segmentasi pelanggannya. Penelitian terdahulu yang berfokus pada segmentasi pelanggan jauh lebih sederhana tanpa adanya kelas maupun treatment pelanggan, dengan hanya menggunakan satu algoritma (K-Means) dan berfokus pada segmentasi berdasarkan atribut RFM tanpa pengujian lanjutan pada kualitas *cluster*. Dengan demikian, penelitian ini menunjukkan peningkatan dalam metode pengelompokan dan segmentasi pelanggan yang lebih akurat dan valid, memperkuat hasil segmentasi dengan validitas statistik dan pengujian lebih mendalam.

## 2.2 Dasar Teori

Pada dasar teori berisi beberapa penjelasan penelitian yang terdiri dari: segmentasi pelanggan, RFM, *K-means*, *Elbow methode*, dan koefisien *silhouette*.

### 2.2.1 Segmentasi Pelanggan

Segmentasi pelanggan adalah salah satu bagian paling penting dari identifikasi pelanggan dan mencakup pengelompokkan seluruh pelanggan sesuai dengan karakteristik pelanggan berdasarkan perilaku pembelian, informasi demografis dan geografis, serta atribut psikografis (Carudin, 2021). Segmentasi merupakan proses membagi pasar menjadi segmen yang lebih kecil dan segmentasi

pelanggan mengacu pada bagaimana membagi sasaran pasar menjadi kelompok yang sesuai dengan karakteristik untuk mengembangkan strategi bisnis yang tepat (Pramudiansyah, dan Munte, 2023). Pembagian segmentasi merupakan proses untuk membagi pelanggan menjadi beberapa kelompok atau segmen yang dapat dikelola berdasarkan dengan karakteristik seperti loyalitas, demografis, kebiasaan membeli atau frekuensi pembelian untuk mengembangkan strategi dalam pemasaran ke setiap kelompok tersebut (Wicaksono, dkk, 2021).

Segmentasi berfungsi sebagai taktik perusahaan untuk mendapatkan pemahaman yang lebih mendalam tentang konsumen yang dituju, sehingga memungkinkan bisnis yang sedang berjalan menghasilkan hasil yang optimal (Sajidah, dkk, 2023). Menurut (Heri Sudarsono, 2020) dalam bukunya yang berjudul Manajemen Pemasaran menyatakan bahwa segmentasi pelanggan juga dapat meningkatkan loyalitas pelanggan dan membantu perusahaan mempertahankan kendali atas pasar mereka. Perusahaan dapat memenuhi kebutuhan pelanggan secara spesifik dan membangun hubungan yang kuat dengan pelanggan melalui komunikasi yang berfokus pada personalisasi (Sarkar, dkk, 2024). Menurut penelitian (Wibowo, 2021), terdapat pengelompokan data yang memiliki karakter data yang sama dan masing-masing data yang sudah dikelompokkan akan dianalisis agar dapat dipahami karakternya. Pada penelitian ini, segmentasi pelanggan dibagi menjadi 10 kelas sesuai dengan karakteristiknya berdasarkan nilai RFM seperti yang ditunjukkan pada tabel berikut ini (Konstantinos dkk, 2010).

Tabel 2.2 Karakter dan Saran Strategi Segmentasi Pelanggan

| Kelas Pelanggan            | Karakteristik                                                                           | Saran Strategi                                                                        |
|----------------------------|-----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| <i>VIP Customer</i>        | Pelanggan paling berharga dan setia dengan transaksi tinggi                             | Berikan hadiah yang dipersonalisasi, program loyalitas, dan layanan pelanggan premium |
| <i>Frequent Buyers</i>     | Pelanggan yang sering bertransaksi, walaupun dengan nilai pembelian tidak selalu tinggi | Tawarkan hadiah loyalitas dan promosi khusus untuk meningkatkan frekuensi pembelian   |
| <i>High Spenders</i>       | Pelanggan dengan nilai transaksi tinggi namun tidak sering membeli                      | Upsell produk premium atau tawarkan pengalaman eksklusif                              |
| <i>Potential Loyalists</i> | Pelanggan yang menunjukkan potensi loyalitas namun belum sepenuhnya konsisten           | Tingkatkan keterlibatan dengan penawaran yang dipersonalisasi dan program loyalitas   |
| <i>New Customers</i>       | Pelanggan baru yang masih dalam tahap eksplorasi dan belum konsisten                    | Berikan penawaran selamat datang dan pandu mereka dalam penemuan produk               |
| <i>Occasional Buyers</i>   | Pelanggan yang membeli secara sporadis atau jarang.                                     | Berikan diskon atau rekomendasi yang sesuai untuk mendorong pembelian lebih sering    |
| <i>At-Risk Customers</i>   | Pelanggan yang pernah aktif tetapi sekarang menurun aktivitasnya                        | Re-engage dengan penawaran yang dipersonalisasi dan kampanye pemulihan pelanggan      |
| <i>Lost Customers</i>      | Pelanggan yang dulu aktif namun sudah lama tidak membeli                                | Tawarkan diskon besar untuk menarik mereka kembali                                    |
| <i>Bargain Hunters</i>     | Pelanggan yang sangat sensitif terhadap harga dan mencari diskon                        | Soroti promosi dan penawaran khusus                                                   |
| <i>Window Shoppers</i>     | Pelanggan yang sering melihat-lihat tetapi jarang membeli                               | Berikan insentif seperti diskon pertama atau rekomendasi produk yang dipersonalisasi  |

### 2.2.2 RFM

Konsep RFM diperkenalkan oleh Bult dan Wansbeek (1995) dan telah terbukti sangat efektif ketika diterapkan dalam database pemasaran. Konsep RFM dapat digunakan untuk mengukur perilaku pelanggan sesuai dengan riwayat transaksi mereka sebelumnya (Rahim, dkk, 2021). RFM merupakan suatu model perhitungan yang terdiri dari tiga atribut domain yang memperhatikan transaksi pelanggan berdasarkan transaksi terakhir *Recency*, jumlah transaksi *Frequency*, serta nominal transaksi *Monetary* (Shirole, dkk, 2021). Metodologi ini sangat bermanfaat dalam segmentasi pelanggan dengan membagi pelanggan ke dalam beberapa kelompok. RFM terdiri dari tiga dimensi, yaitu *Recency*, *Frequency*, *Monetary* (Bagul, dkk, 2021):

a. *Recency*

*Recency* adalah mengukur nilai pelanggan berdasarkan pembelian paling akhir yang dilakukan pelanggan. Data terpenting yang diperlukan untuk menghitung nilai *recency* adalah tanggal pembelian terakhir. Nilai *recency* berkaitan dengan jarak antara tanggal terakhir transaksi dengan periode analisis, sehingga semakin dekat proses pembelian terakhir maka pelanggan tersebut semakin loyal. Nilai *recency* ini biasanya dinyatakan dalam satuan hari.

b. *Frequency*

*Frequency* adalah mengukur nilai pelanggan berdasarkan seberapa sering pelanggan tersebut melakukan transaksi. Semakin sering melakukan transaksi ini memungkinkan bahwa pelanggan tersebut merupakan pelanggan potensial. Transaksi yang dimaksud tidak tergantung pada berapa banyak yang pelanggan beli, tetapi satu transaksi pun bisa mewakili bahwa pelanggan tersebut potensial.

c. *Monetary*

*Monetary* adalah mengukur nilai pelanggan dengan melihat nilai pembelian yang dilakukan pelanggan dalam periode tertentu. Semakin tinggi nilai pembelian maka semakin tinggi pula potensial pelanggan tersebut.

RFM telah berhasil digunakan dalam proses analisis nilai pelanggan dalam beberapa penelitian sebelumnya, sehingga timbul gagasan bahwa konsep RFM juga dapat diterapkan untuk segmentasi produk berdasarkan data transaksi penjualan sebelumnya (Monalisa dan Erza, 2021). Langkah-langkah implementasi untuk perhitungan komponen RFM adalah sebagai berikut (Tsiptsis & Chorianopoulos, 2010):

1. Pengambilan Data (*data acquisition*)

Data transaksi yang sesuai dengan periode analisis diambil, diaudit, dibersihkan, dan disiapkan untuk tahap analisis selanjutnya.

2. Pemilihan populasi yang akan disegmentasi (*selection of the population to be segmented*)

Data yang diambil hanya data pelanggan yang aktif melakukan pembelian selama rentang periode analisis untuk dimasukkan dalam daftar segmentasi RFM. Dalam penelitian ini, data yang digunakan adalah data transaksi produk dalam kurun waktu periode analisis.

3. Persiapan data dan perhitungan pengukuran R, F, dan M

Menurut Cheng dan Chen (2009), semakin besar nilai *R* dan *F* maka kemungkinan pelanggan akan melakukan transaksi kembali dengan perusahaan tersebut. Selain itu semakin besar nilai *M*, maka kecenderungan pelanggan dalam memberikan respon kepada produk dan layanan perusahaan (Wibowo dan Handoko, 2020). Sekumpulan data dengan jumlah yang besar dapat digali informasinya dengan istilah *data mining* dan *knowledge discovery in databases* (KDD) (Ernawati, dkk, 2020).

### 2.2.3 K-means

Algoritma K-means merupakan teknik *clustering* berbasis centroid. Teknik ini mengelompokkan data set ke k-cluster berbeda yang memiliki jumlah titik yang hampir sama (Aditya dan Fitriana, 2021). Setiap cluster pada algoritma clustering

k-means diwakili oleh sebuah titik centroid. Titik centroid adalah rata-rata dari semua titik di set dan akan berubah di setiap langkah dan akan dihitung (Anitha dan Patil, 2022). Dari kajian literatur yang dilakukan oleh (William Santoso, dkk, 1993) ditemukan bahwa kombinasi penggunaan algoritma K-Means menghasilkan performa yang tinggi.

Menurut (Rohmah, dkk, 2021) Algoritma *K-means Cluster Analysis* merupakan bidang penelitian dalam analisis dan data *mining*. Pada algoritma ini teknik pengelompokannya berdasarkan kemiripan data yang tidak memiliki acuan apapun (*unsupervised*), Tetapi, akan membagi keseluruhan data yang akan menjadi kelompok atau mempunyai kemiripan yang sama. Pada dasarnya algoritma ini menghitung jarak pada setiap data dengan pusat data (*centroid*) untuk mengukur kemiripan data. Metode ini digunakan bertujuan untuk meminimalisir fungsi objektif yang diatur pada proses cluster-ing atau pengklasteran dengan meminimalkan variasi antar data yang terdapat dalam suatu cluster dan memaksimalkan variasi dengan data yang terdapat pada cluster lain.

Adapun prosedur perhitungan Algoritma *K- Means Clustering analysis* diuraikan sebagai berikut (Vysala, dkk, 2020):

1. Menentukan banyaknya cluster

Dalam menentukan banyaknya *cluster* diharuskan tidak lebih dari jumlah kriteria yang ada.

2. Menentukan titik pusat *cluster (centroid)*

Dalam menentukan titik pusat *cluster (centroid)* dengan mengambil nilai minimal, rata-rata dan maksimal, jika akan dibuat 3 *cluster*, karna pada penelitian ini penulis membuat 2 *cluster* jadi hanya mengambil nilai minimal dan maksimal pada setiap kriteria.

3. Menghitung antara jarak titik data objek ke titik data pusat (*centroid*)

Pada tahap perhitungan jarak ini menggunakan rumus dengan *Euclidean Distance* sebagai berikut:

$$d_{Euclidean}(x,y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Keterangan:

d (x,y) : jarak antara data pada titik x dan y

x : titik data objek

y : titik data *centroid*

i : jumlah atribut data

4. Pengelompokan data objek untuk menentukan anggota *cluster* berdasarkan jarak minimum

Pada proses ini setelah menghitung jarak diambil nilai minimum yang diberi nilai 1 dan lainnya 0, dimana nilai 1 untuk data yang ditempatkan pada ke *cluster* dan nilai 0 untuk data yang tempatkan ke *cluster* yang lain hingga akan membentuk sebuah matriks.

5. Kembali pada tahap ke 2

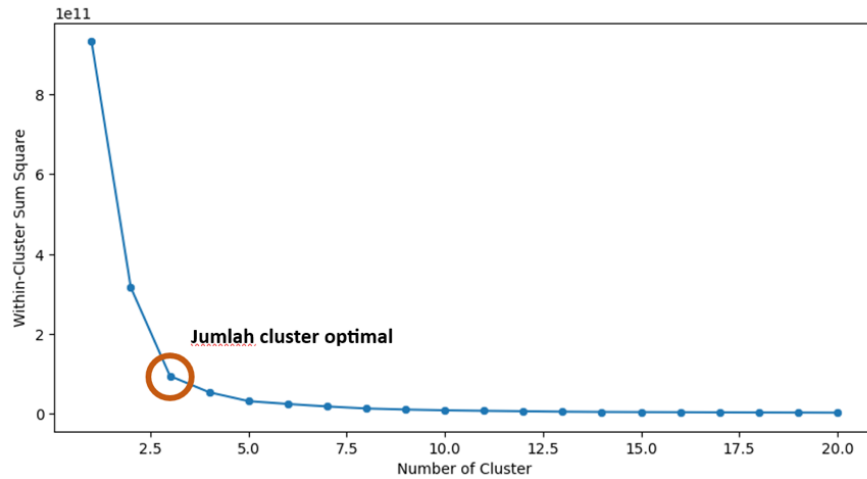
Lakukan pengulangan hingga poin 4 hingga nilai pada tiap *cluster* tidak berubah tempat.

#### 2.2.4 *Elbow Method*

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik (Ashari, dkk, 2023). Metode ini memberikan ide atau gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan cluster terbaik, selain itu persentase perhitungan yang dihasilkan menjadi pembanding antara jumlah *cluster* yang ditambah (Maori, 2023).

Hasil persentase yang berbeda dari setiap nilai *cluster* dapat ditunjukan dengan menggunakan grafik sebagai sumber informasinya (Ekasetya, dan jananto,

2020). Jika nilai cluster pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai *cluster* tersebut yang terbaik (Atira, dan Sari, 2023).



Gambar 2.1 Grafik *Elbow Method*

Algoritma Metode *Elbow* dalam menentukan nilai  $K$  pada K-Means (Merliana, Ernawati, & Santoso, 2008)

1. Mulai
2. Inisialisasi awal nilai  $K$
3. Naikkan nilai  $K$
4. Hitung hasil *sum of square error* dari tiap nilai  $K$
5. Melihat hasil *sum of square error* dari nilai  $K$  yang turun secara drastic
6. Tetapkan nilai  $K$  yang beberbentuk siku

Pendekatan grafis dan pendekatan lama untuk menemukan jumlah  $K$  yang optimal dalam algoritma K-means. Metode *elbow* biasa disajikan dalam bentuk grafik untuk mengetahui lebih jelas siku yang terbentuk (Kong, dkk, 2021). Tujuan dari metode *elbow* adalah untuk memilih nilai  $k$  yang kecil dan masih memiliki nilai *withinss* yang rendah. Nilai  $k$  pada kombinasi siku dengan *k-means* adalah grafik hubungan *cluster* dengan penurunan *error*. Jumlah *cluster*  $k$  yang dihasilkan dari pengujian dengan *k-means* dievaluasi dengan teknik SSE. SSE (*Sum of Square*

*Error*) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang telah dilakukan sebelumnya (Rocha, dkk, 2021). SSE didefinisikan sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2$$

Keterangan :

$K$  = jumlah cluster

$S_k$  = cluster ke- $k$ ,  $x_i$  adalah elemen dari cluster ke- $k$

$C_k$  = centroid dari cluster  $S_k$

$\| \cdot \|$  = jarak Euclidean antara dua pola data.

Kita harus memplot grafik diagram garis antara SSE dan nilai klaster yang sesuai dengan  $K$ .  $K$  dimulai dengan 2 dan akan bertambah 1 di setiap langkah. Jika grafik diagram garis menunjukkan penurunan SSE yang drastis, misalnya, seperti lengan, maka “siku” pada lengan adalah nilai yang menunjukkan jumlah cluster  $k$  yang sesuai dalam pengelompokan K-means (Alam, dkk, 2024).

### 2.2.5 Koefisien *Silhouette*

Salah satu metode evaluasi yang digunakan untuk melihat optimalisasi suatu kelompok atau *cluster* adalah metode koefisien *silhouette* (Atira, dan Sari, 2023). Metode ini digunakan untuk mengetahui seberapa tepat suatu data dikelompokkan. Metode ini merupakan metode validasi yang menggabungkan metode *cohesion* dan *separation* (Ashari, dkk, 2023). Pada metode ini dibutuhkan nilai jarak antar nilai pelanggan menggunakan rumus *Euclidean Distance* yang kemudian dilanjutkan dengan tahapan (Khan, dkk, 2024):

1. Untuk setiap objek  $I$ , hitung rata-rata jarak dari objek  $I$  dengan seluruh objek yang berada dalam satu kelompok sehingga diperoleh nilai rata-rata  $a_i$ .

2. Setiap objek I, hitung rata-rata jarak dari objek I dengan objek yang berada di kelompok lainnya dan diambil nilai yang paling kecil yaitu  $b_i$ .
3. Hitung nilai *silhouette coefficient* untuk objek I mengikuti persamaan berikut

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i) - b(i)\}}$$

**Keterangan:**

- $a_i$  : rata-rata objek ke-i dengan semua objek pada satu kelompok yang sama
- $b_i$  : rata-rata objek ke-i dengan semua objek pada satu kelompok yang berbeda
- $s_{(i)}$  : nilai koefisien *silhouette*

Hasil perhitungan nilai koefisien *silhouette* dapat bervariasi antara -1 hingga 1. Hasil pengelompokan dikatakan tepat jika nilai koefisien *silhouette* berada di antara dua kelompok dan objek tersebut tidak jelas harus dimasukkan ke dalam kelompok yang mana, jika nilai koefisien *silhouette* tersebut -1 artinya struktur kelompok yang dihasilkan tidak baik, maka objek tersebut lebih tepat dimasukkan ke dalam kelompok yang lain (Hidayanti, dkk, 2021).

SEKOLAH PASCASARJANA