

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Tempat Penelitian

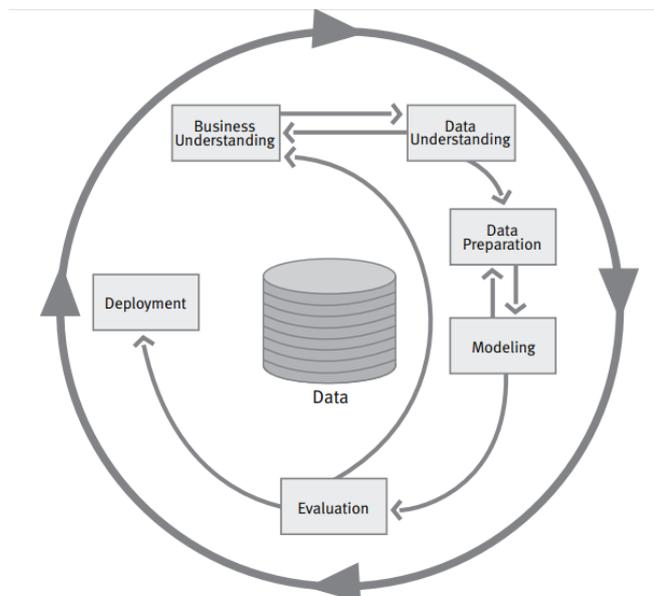
Tovi Kohi merupakan salah satu *coffee shop* yang berada di Kota Bandung. Tovi Kohi berdiri sejak tahun 2018. Tovi Kohi sudah memiliki *website* sistem informasi untuk membantu dalam pengelolaan transaksi. Sampai saat ini, Tovi Kohi memiliki data transaksi dengan jumlah besar yang disimpan sebagai arsip.

2.2 Landasan Teori

Landasan teori membahas mengenai materi atau teori yang digunakan sebagai acuan dalam penelitian. Landasan teori yang diuraikan berdasarkan hasil dari studi literatur yaitu mengumpulkan data dengan cara mempelajari jurnal, referensi buku, artikel dan bacaan-bacaan lainnya yang berhubungan dengan penelitian ini.

2.2.1 CRISP-DM

Metodologi *Cross-Industry Standard Process for Data Mining (CRISP-DM)* merupakan metode standar dalam pengolahan data untuk memecahkan masalah bisnis dan mencari solusi dari data yang ada [7].



Gambar 0.1 Metodologi CRISP-DM

Tahapan-tahapan yang dilakukan berdasarkan gambar di atas meliputi:

1. *Business Understanding*

Fase awal ini berfokus pada pemahaman tujuan dan persyaratan proyek dari perspektif bisnis, kemudian mengubah pengetahuan ini menjadi definisi masalah data mining dan rencana awal yang dirancang untuk mencapai tujuan tersebut [7].

2. *Data Understanding*

Pada tahap ini dilakukan pengumpulan data, menggambarkan data, eksplorasi data, dan verifikasi kualitas data [7]. Berikut penjelasan masing-masing tahapannya:

4) Pengumpulan Data

Buat daftar kumpulan data yang diperoleh termasuk lokasinya, metode yang digunakan untuk mendapatkannya, dan masalah apa pun yang dihadapi. Rekam masalah yang dihadapi dan solusi apa pun yang dicapai.

5) Gambaran Data

Menjelaskan data yang diperoleh, termasuk format data, volume data (misalnya, jumlah record dan field dalam setiap tabel), pengidentifikasi field, dan karakteristik permukaan lain yang ditemukan. Menilai apakah informasi yang diperoleh memenuhi persyaratan yang relevan.

6) Eksplorasi Data

Pada tahap ini membahas pertanyaan *data mining* menggunakan teknik kueri, visualisasi, dan pelaporan. Di dalamnya termasuk distribusi atribut kunci (misalnya atribut target dari tugas yang diprediksi), asosiasi antara pasangan atau sejumlah kecil atribut, hasil agregasi sederhana, karakteristik subpopulasi yang signifikan, dan analisis statistik sederhana. Analisis ini dapat secara langsung menjawab tujuan *data mining*. Selain itu juga, dapat berkontribusi atau meningkatkan deskripsi data dan laporan kualitatif, dan membantu dalam transformasi dan tahapan persiapan data lainnya yang diperlukan untuk analisis lebih lanjut [7].

7) Verifikasi Kualitas Data

Memeriksa kualitas data, apakah data lengkap (mencakup semua kasus yang diperlukan), apakah data sudah benar, atau mengandung kesalahan dan, jika ada kesalahan, seberapa umum kesalahan tersebut, dan memeriksa apakah ada nilai yang hilang dalam data. Pada tahap ini dilakukan seperti pemeriksaan data yang tidak relevan, data salah, data duplikasi, dan data kosong. Data duplikasi adalah data yang memiliki nilai sama pada dua atau lebih atribut. Data kosong adalah data yang tidak memiliki nilai pada suatu atribut. Cantumkan hasil verifikasi kualitas data. Jika ada masalah kualitas, buatlah daftar solusi yang memungkinkan. Solusi untuk masalah kualitas data umumnya sangat bergantung pada data dan pengetahuan bisnis [7].

3. *Data Preparation*

Pada tahap ini dilakukan seleksi data, pembersihan data, integrasi data, dan transformasi data [7]. Berikut penjelasan masing-masing tahapan:

6) Seleksi Data

Tentukan data yang akan digunakan untuk analisis. Kriteria meliputi relevansi data, kualitas, dan batasan teknis untuk tujuan *data mining*, seperti batasan volume data atau tipe data. Perhatikan bahwa pemilihan data melibatkan pemilihan atribut (kolom) dan rekaman (baris) dari sebuah tabel. Hasilnya berupa data yang akan dipilih beserta alasannya.

7) Pembersihan Data

Membersihkan data dari nilai-nilai data yang tidak diperlukan seperti data kosong, data duplikat, dan *outlier*. menjelaskan keputusan dan tindakan apa yang diambil untuk menyelesaikan masalah kualitas data yang dilaporkan selama tahap kualitas data dalam fase *data understanding*. Transformasi data untuk tujuan pembersihan dan dampak potensial pada hasil analisis harus dipertimbangkan.

8) Pembangunan Data

Tahap ini melibatkan aktivitas pemrosesan data yang konstruktif seperti membuat atribut turunan atau record yang sama sekali baru atau mengubah nilai

atribut yang ada. Atribut turunan adalah atribut baru yang dibangun dari satu atau lebih atribut yang ada dalam record yang sama.

9) Integrasi Data

Ini adalah metode menggabungkan data dari beberapa tabel atau catatan untuk membuat nilai atau catatan baru. Menggabungkan tabel berarti menggabungkan dua tabel atau lebih yang memiliki informasi berbeda tentang objek yang sama.

10) Transformasi Data

Transformasi data sebagian besar merujuk pada perubahan sintaksis yang dilakukan pada data yang tidak mengubah artinya tetapi dapat berguna untuk pemodelan. Penting untuk mengubah urutan atribut atau nilai dalam kumpulan data.

4. *Modeling*

Langkah pertama dalam pemodelan adalah memilih teknik pemodelan yang akan digunakan. Selanjutnya, dokumentasikan teknik pemodelan aktual yang akan digunakan. Sebelum membangun model, kita perlu membuat prosedur atau mekanisme untuk menguji kualitas dan validitas model. Jelaskan rencana yang dimaksudkan untuk melatih, menguji, dan mengevaluasi model. Komponen utama dari rencana tersebut adalah menentukan cara membagi kumpulan data yang tersedia menjadi kumpulan data pelatihan, pengujian, dan validasi.

5. *Evaluation*

Langkah-langkah evaluasi sebelumnya berurusan dengan faktor-faktor seperti akurasi dan keumuman model. Langkah ini menilai sejauh mana model memenuhi tujuan bisnis dan berusaha untuk menentukan apakah ada beberapa alasan bisnis mengapa model ini kurang [7].

6. *Deployment*

Tahap ini mengambil hasil evaluasi dan menentukan strategi untuk *deployment*. Jika prosedur umum telah diidentifikasi untuk membuat model yang relevan, prosedur ini didokumentasikan di sini untuk penerapan selanjutnya [7].

2.2.2 Data Mining

Data mining adalah tahap analisis data dalam *data science lifecycle* yang memfokuskan pada pencarian pola dan wawasan dari jumlah data besar yang terstruktur atau tidak terstruktur. Prosesnya melibatkan algoritma statistik dan machine learning untuk mencari pola (*pattern*) atau hubungan yang tidak diketahui sebelumnya, berguna sebagai pengetahuan untuk membuat keputusan bisnis [6]. Beberapa tugas data mining meliputi *classification*, *clustering*, *regression*, dan *association*.

Alur kerja atau tahapan pada *data mining* [6] pada umumnya meliputi beberapa proses sebagai berikut:

1. *Data Collection*

Pada tahap ini melakukan pengumpulan data dari berbagai sumber seperti *sensor network*, melakukan survei, *crawling / scraping web*, dan lain-lain. Tentunya harus dipastikan bahwa jumlah data cukup untuk digunakan dalam proses analisis dan memastikan bahwa data dari sumber yang terpercaya.

2. *Feature Extraction and Data Cleaning*

Setelah data terkumpul, belum tentu data dapat langsung dipakai untuk analisis. Maka dari itu, dilakukan proses *feature extraction* yaitu mengekstrak fitur yang penting dari data untuk digunakan dalam proses analisis. *Data cleaning* merupakan proses untuk membersihkan data dari *error*, *duplicate*, dan data yang tidak berguna.

3. *Analytical Preprocessing and Algorithms*

Analytical preprocessing adalah proses untuk menganalisis data dan menentukan hubungan antar variabel untuk memastikan bahwa data dapat digunakan untuk membuat keputusan bisnis. *Algorithms* adalah menentukan algoritma yang digunakan dalam proses analisis. Ada berbagai jenis algoritma seperti *classification*, *clustering*, *regression*, dan lainnya. Algoritma yang dipilih harus sesuai dengan masalah yang dianalisis.

2.2.3 Clustering

Clustering atau klasterisasi adalah teknik untuk melakukan pengelompokan terhadap data atau objek dari populasi atau data besar dengan membuat perbandingan kuantitatif dari beberapa karakteristik [6]. *Clustering* melakukan pengelompokan data tanpa adanya kelas data tertentu, sehingga dilakukan pembagian seluruh data menjadi kelompok yang memiliki kemiripan atribut [6]. Tujuan utama *clustering* adalah sebagai berikut:

1. Untuk mendapatkan pengetahuan yang berguna dari data, yaitu menghasilkan hipotesis, mendeteksi anomali, dan mengidentifikasi fitur yang menonjol dalam data yang diberikan.
2. Untuk mengidentifikasi tingkat kesamaan (*similarity*) antara bentuk atau titik yang terdiri dari data.
3. Sebagai metode untuk mengatur data dan meringkasnya melalui prototipe *cluster*.

Beberapa tipe-tipe yang umum digunakan dalam *clustering* [6], yaitu sebagai berikut:

1. *K-Means Clustering*, merupakan algoritma yang membagi data menjadi k klaster berdasarkan jarak antar titik data.
2. *Hierarchical Clustering*, merupakan algoritma yang membagi data menjadi klaster berdasarkan hubungan antar titik data.
3. *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)*, merupakan algoritma yang membagi data menjadi klaster berdasarkan densitas data.
4. *Fuzzy C-Means Clustering*, merupakan algoritma yang membagi klaster berdasarkan jarak antar titik data dan memperbolehkan titik data untuk bergabung dengan beberapa klaster.
5. *Expectation Maximization*, merupakan algoritma yang membagi klaster dengan mengasumsikan bahwa data dalam klaster memiliki distribusi gaussian.

2.2.4 K-Means

Algoritma *K-Means* membagi data menjadi k kluster berdasarkan jarak antar titik data. Cara kerjanya yaitu dengan menentukan k *centroid* awal secara acak dan kemudian menempatkan setiap titik data pada kluster yang paling dekat jaraknya dengan *centroid* tersebut. Proses ini dilakukan secara iteratif hingga *centroid* tidak berubah lagi. *K-Means* sangat efektif untuk menemukan struktur dalam data yang berdimensi tinggi, namun kurang efektif untuk data dengan struktur yang tidak teratur atau data dengan ukuran *cluster* yang berbeda [8].

Untuk menghitung iterasi jarak terdekat antar *cluster* menggunakan rumus *Euclidean Distance*. Dengan menentukan nilai *centroid* awal (nilai tengah) kemudian partisi data menggunakan nilai minimum. Lakukan iterasi sampai nilai atau posisi data tidak mengalami perubahan. Jika data sudah tidak berubah lagi maka hentikan iterasi. Maka hasil dari perhitungan iterasi merupakan *centroid* akhir dari data yang sudah dipartisi sebelumnya [9]. Berikut merupakan tahapan dari algoritma K-means *clustering*:

1. Menentukan jumlah *cluster* yang akan dibuat.
2. Menentukan titik *centroid* awal (titik pusat *cluster*) untuk dilakukannya proses *clustering*.
3. Mengelompokkan data dengan menghitung jarak data (distance measure) dengan masing – masing *centroid*. Adapun untuk menghitung jarak setiap data ke pusat *cluster* adalah dengan menggunakan rumus *euclidean distance*:

$$d(i, k) = \sqrt{(\sum_{j=1}^m (i_j - k_j)^2)} \quad (1)$$

Keterangan:

D(i,k) = perhitungan jarak data ke-i dengan pusat kluster ke-k

j = banyak data

m = maksimal data

4. Mengelompokkan data ke dalam *cluster* dengan jarak paling minimum dari setiap data terhadap masing – masing *centroid*.

5. Hitung *centroid* baru dengan cara menghitung nilai rata rata data yang terletak pada *cluster* yang sama menggunakan rumus berikut:

Persamaan 2:

$$\bar{v}_{ij} = \frac{1}{N_i} \sum_k^n x_{kj} \quad (2)$$

Keterangan:

\bar{v}_{ij} = *centroid cluster* ke-i untuk variabel ke-j

N_i = jumlah data yang menjadi anggota *cluster* ke-i

i, k = indeks dari *cluster*

j = indeks dari variabel

x_{kj} = indeks data ke-k yang ada di dalam *cluster* tersebut untuk variabel ke-j

6. Ulangi langkah 3 sampai 5 hingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain.

2.2.5 Davies Bouldin Index

Davies Bouldin Index ini adalah evaluasi yang berkerja di internal *cluster*, kualitas hasil *cluster* dilihat dari nilai DBI yang terkecil [10]. *Davies Bouldin Index* merupakan salah satu metode yang digunakan untuk mengukur sejauh mana ketepatan suatu cluster pada suatu metode *clustering*. Semakin rendah nilai DBI, semakin baik pengelompokan, karena menunjukkan bahwa *cluster-cluster* memiliki pusat yang jauh satu sama lain dan penyebaran yang lebih baik di dalam *cluster* [6]. Perhitungan evaluasinya di mulai dari menentukan nilai *SSW*, *SSB* dari titik *centroid* awal lalu akan menentukan nilai DBI nya. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within cluster (SSW)* yaitu:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (3)$$

Keterangan:

m = jumlah data dalam *cluster* ke i

x = data pada *cluster*

c = *centroid*

$d(x,c)$ = jarak data dengan *centroid*

Sum of Square Between-cluster (SSB) bertujuan untuk mengetahui separasi antar *cluster*. Persamaan yang digunakan untuk menghitung *Sum of Square Between-cluster (SSB)* yaitu:

$$SSB_{i,j} = d(c_i, c_j)$$

Keterangan:

c_i = *cluster* ke i

c_j = *cluster* ke j

$d(c_i, c_j)$ = jarak *centroid* i *centroid* j

Dalam perhitungan *SSW*, kita menghitung jumlah kuadrat selisih antara setiap titik data dalam *cluster* dengan rata-rata *cluster* itu sendiri. Kemudian, kita menjumlahkan jumlah kuadrat ini untuk semua *cluster*.

Sementara *SSB*, kita menghitung jumlah kuadrat perbedaan antara rata-rata *cluster* itu sendiri dengan rata-rata keseluruhan. Kita kemudian mengalikannya dengan jumlah titik data dalam *cluster* tersebut, dan menjumlahkan hasil ini untuk semua *cluster*. Selanjutnya menghitung nilai ratio seberapa baik nilai perbandingan antara *cluster* ke- i dengan *cluster* ke- j . Untuk menghitung ratio menggunakan rumus sebagai berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

Tujuan utamanya adalah untuk meminimalkan *SSW* dan memaksimalkan *SSB*. Hal ini menunjukkan bahwa titik-titik data dalam *cluster* seharusnya lebih serupa satu sama lain (*SSW* yang rendah), sementara *cluster* seharusnya berbeda satu sama lain (*SSB* yang tinggi). Persamaan *DBI* adalah sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (5)$$

Keterangan:

R_{ij} = rasio *SSW* dan *SSB* dari dua *cluster*

K = jumlah cluster terbentuk

2.2.6 Segmentasi Pasar

Perkembangan *marketing* dalam bisnis pada saat ini sudah mementingkan kepuasan pelanggan dalam menjual produknya. Karena tidak peduli seberapa baik atau bagus suatu produk, jika tidak dipasarkan secara baik maka tidak akan ada penjualan yang menguntungkan. Jika suatu instansi atau kedai kopi sudah memberikan konsumen pertimbangan yang layak mereka dapatkan dan mencari tahu lebih banyak tentang konsumen dan kebutuhannya, terkadang hal tersebut belum cukup untuk menjawab keputusan yang akan diambil dalam memanfaatkan peluang pasar. [11]. Segmentasi pasar memiliki beberapa keunggulan. Pada tingkat yang paling umum, Segmentasi memaksa perusahaan untuk mempertimbangkan di mana mereka berdiri dan di mana mereka ingin berada di masa depan. Hal itu memaksa organisasi untuk berpikir tentang keputusan yang baik untuk mendapatkan wawasan tentang apa yang konsumen inginkan [4].

Segmentasi pasar merupakan salah satu kunci penentu keberhasilan pemasaran, dan merupakan dasar untuk proses pencocokan antara keinginan dan kebutuhan konsumen dengan kemampuan organisasi / kedai kopi dalam memuaskan mereka. Dalam memahami segmentasi pasar, perlu untuk memperjelas perbedaan antara pelanggan (*customer*) dan konsumen (*consumer*) [11].

Pelanggan (*customer*) merupakan individu yang membeli produk atau layanan, sedangkan konsumen (*consumer*) merupakan individu yang menggunakan atau mengkonsumsi produk atau layanan. Pelanggan atau konsumen bisa saja dikatakan sama jika individu tersebut yang melakukan pembelian terhadap produk dan menggunakan produk tersebut oleh dirinya sendiri [11].

2.2.7 Kriteria Segmentasi Pasar

Analisis segmentasi pasar yang baik membutuhkan data empiris yang baik. Data empiris untuk studi segmentasi dapat berasal dari beberapa sumber seperti survei, observasi, pencatatan pembelian, kebiasaan konsumen, atau melakukan eksperimen terhadap objeknya. Kriteria segmentasi yang paling umum digunakan adalah geografis, sosiodemografi, psikografi dan perilaku. [4].

1. Segmentasi Demografik

Segmentasi demografik merupakan segmentasi berdasarkan karakteristik demografik seperti usia, jenis kelamin, pendapatan, status pekerjaan dan pendidikan. Kelebihan utama dari segmentasi geografis adalah, mudah untuk menargetkan pesan komunikasi, dan memilih saluran komunikasi (seperti surat kabar lokal, lokal stasiun radio dan TV) untuk menjangkau segmen geografis yang dipilih. Kekurangannya adalah walaupun tinggal di negara atau wilayah yang sama bukan berarti orang-orang memiliki karakteristik sama yang relevan terhadap pemasar [4].

2. Segmentasi Geografis

Ketika segmentasi geografis digunakan, lokasi atau tempat tinggal konsumen berfungsi sebagai satu-satunya kriteria untuk membentuk segmen pasar. Meskipun sederhana, pendekatan segmentasi geografis seringkali yang paling tepat [4].

3. Segmentasi Perilaku

Segmentasi perilaku atau bisa disebut *behavioral segmentation* merupakan segmentasi yang mempelajari ciri-ciri perilaku konsumen yang mencakup pengetahuan mengenai sikap, penggunaan, kebiasaan, atau tanggapan mereka terhadap suatu produk, layanan, promosi, termasuk pengalaman sebelumnya dengan produk, frekuensi pembelian, jumlah yang dihabiskan untuk membeli produk setiap kali (atau beberapa kali pembelian), dan perilaku saat mencari informasi [4]. Segmentasi perilaku dibagi menjadi beberapa jenis, yaitu sebagai berikut:

a. *Purchasing and Usage Behavioral*

Mempelajari mengenai perilaku pembelian dan penggunaan konsumen terhadap produk berdasarkan frekuensi atau intensitas, waktu pembelian, dan kebutuhan / keinginan. Dengan segmentasi pelanggan berdasarkan perilaku pembelian dan penggunaan mereka, perusahaan dapat lebih memahami kebutuhan dan preferensi pelanggan mereka dan menyesuaikan upaya pemasaran mereka. Misalnya, sebuah perusahaan dapat menargetkan pelanggan

setia dengan promosi khusus atau membuat produk khusus untuk pengguna pertama kali. Ini dapat membantu meningkatkan kepuasan pelanggan dan loyalitas merek, yang pada akhirnya dapat menyebabkan peningkatan penjualan dan profitabilitas bagi perusahaan [4].

b. Occasion Purchasing

Occasion purchasing atau pembelian untuk acara tertentu adalah jenis perilaku pembelian yang terjadi jika seseorang membeli produk atau layanan untuk menghadapi atau merayakan acara tertentu, seperti hari ulang tahun, pernikahan, tahun baru, hari raya idul fitri, dan lain sebagainya. *Occasion purchasing* dapat menjadi peluang untuk meningkatkan penjualan dan memperluas pangsa pasar. Perusahaan dapat menyesuaikan strategi pemasaran mereka untuk mendorong pembelian untuk acara tertentu dengan mengeluarkan promosi atau diskon khusus, menyediakan produk atau layanan yang sesuai dengan acara tertentu, atau menawarkan layanan pengiriman khusus [4].

2.2.8 Segmentasi Konsumen

Segmentasi konsumen adalah proses membagi konsumen menjadi kelompok-kelompok yang memiliki kesamaan satu sama lain. Konsumen dibagi menjadi beberapa segmen untuk mengoptimalkan signifikansi setiap konsumen dalam bisnis, yang berguna dalam memodifikasi produk atau menu sesuai dengan kebutuhan dan perilaku yang berbeda dari setiap konsumen [12].

Segmentasi konsumen memainkan peran penting untuk memungkinkan bisnis membuat pemasaran agar lebih efektif, melihat tren dalam perilaku pembelian konsumen dan membuat penawaran produk yang relevan kepada konsumen. Secara sederhana, pembagian segmentasi ini adalah proses membagi basis konsumen yang ada menjadi kelompok yang dapat dikelola dan layak berdasarkan karakteristik umum seperti usia, jenis kelamin, loyalitas, frekuensi pembelian, dan sebagainya [12].

2.2.9 Membership

Membership ini untuk salah satu cara mendapatkan konsumen tetap dan loyal dengan berisikan diskon atau potongan harga di dalam program member tersebut

[2]. Membership atau keanggotaan merupakan simbol sebagai konsumen dalam sebuah perusahaan yang termasuk bagian perusahaan atau organisasi yang diakui secara resmi [2]. Membership biasanya diberi tanda keanggotaannya dengan bentuk sim card atau kartu anggota. Yang dapat bergabung dengan program membership perusahaan dapat sebagai individu atau grup. Program membership ini dapat menguntungkan kepada pihak perusahaan [2]. Dengan membership card ini perusahaan dapat mendata seberapa banyak konsumen serta sebagai acuan bagian perusahaan yang dapat menjadi bahan evaluasi apakah member berguna atau tidak dalam program membership ini. Saat ini banyak perusahaan yang menggunakan membership guna mempertahankan konsumen tetap untuk mendapatkan loyalitas dalam bisnisnya [2]. Dengan melakukan membership ini perusahaan berharap mendapatkan loyalitas kepada perusahaan. Loyalitas konsumen kepada perusahaan ini berdampak kepada penjualan perusahaan yang menimbulkan profit yang menguntungkan, karena konsumen memilih perusahaan untuk sebagai pemenuh kebutuhan mereka atau membeli kembali produk atau jasa dari perusahaan [2].

2.2.10 Metode RFM

Metode RFM (*Recency, Frequency, dan Monetary*) adalah model untuk mengidentifikasi kelompok pelanggan dengan perlakuan khusus. Model RFM merupakan suatu teknik atau model yang dapat meramalkan perilaku konsumen dengan menerjemahkan perilaku konsumen ke dalam angka sehingga dapat mengarahkan suatu keputusan segmentasi yang lebih baik. Analisis RFM bergantung pada parameter *Recency, Frequency, dan Monetary* [13]. *Recency* menunjukkan waktu interaksi terakhir pelanggan dengan produk. *Frequency* jumlah transaksi yang dilakukan pelanggan. *Monetary* mengukur jumlah uang yang dihabiskan pelanggan untuk tiap transaksi pembelian [14]. Tahapan melakukan perhitungan model RFM untuk setiap pelanggan, hitung skor RFM berdasarkan tiga faktor utama [13]:

1. Recency

Pilih tanggal acuan terbaru untuk menghitung skor RFM. Misalnya, tanggal acuan bisa menjadi tanggal terakhir dalam dataset atau tanggal yang ditentukan secara manual. Hitung selisih antara tanggal terakhir pembelian pelanggan

dengan tanggal acuan. Berikan skor berdasarkan rentang waktu antara pembelian terakhir [13].

2. *Frequency*

Hitung jumlah total pembelian atau frekuensi pembelian pelanggan selama periode waktu tertentu. Berikan skor yang lebih tinggi untuk pelanggan yang melakukan pembelian lebih sering [13].

3. *Monetary*

Hitung total nilai pembelian pelanggan selama periode waktu tertentu. Berikan skor yang lebih tinggi untuk pelanggan yang memiliki nilai transaksi lebih tinggi [13].

Analisis RFM lebih dari sekadar memberikan skor kepada pelanggan berdasarkan perilaku pembelian pelanggan. Analisis RFM dapat membantu dalam melakukan segmentasi konsumen dengan alasan berbagai hal berikut ini [15]:

1. Fokus pada perilaku pembelian: Model RFM berfokus pada perilaku pembelian konsumen, yang merupakan indikator utama keterlibatan dan loyalitas konsumen. Dengan mempertimbangkan waktu, frekuensi, dan nilai pembelian, model RFM membantu mengidentifikasi konsumen yang kemungkinan besar akan menjadi pelanggan setia dan berpotensi meningkatkan penjualan [15].
2. Relevansi dengan tujuan bisnis: Model RFM membantu dalam mengarahkan strategi pemasaran dan promosi dengan memfokuskan upaya pada segmen konsumen yang paling berharga. Dengan mengetahui segmen konsumen dengan keterlibatan tinggi dan potensi nilai transaksi, bisnis dapat mengalokasikan sumber daya dan upaya pemasaran mereka secara efektif [15].
3. Personalisasi pemasaran: Dengan memahami perilaku pembelian konsumen, model RFM memungkinkan perusahaan untuk melakukan personalisasi pemasaran yang lebih efektif. Melalui segmentasi berdasarkan RFM, perusahaan dapat menyusun strategi pemasaran yang lebih relevan dan menargetkan penawaran promosi yang sesuai dengan preferensi dan kebutuhan setiap segmen konsumen [15].

4. Pengukuran yang mudah: Model RFM memberikan metrik yang jelas dan terukur untuk mengukur keterlibatan konsumen dan memberikan insight tentang tingkat kesuksesan upaya pemasaran. Hal ini memungkinkan perusahaan untuk melacak perubahan dalam perilaku pembelian konsumen dari waktu ke waktu dan mengevaluasi efektivitas strategi pemasaran yang diimplementasikan [15].

2.2.11 Profiling

Profiling adalah proses mengidentifikasi, menggambarkan, dan mengelompokkan konsumen berdasarkan karakteristik, perilaku, preferensi, dan kebutuhan mereka. Tujuan dari profiling konsumen adalah memahami secara lebih mendalam tentang konsumen, apa yang mereka inginkan, bagaimana mereka berperilaku, dan bagaimana memenuhi kebutuhan mereka dengan lebih baik [16].

Profiling dilakukan setelah segmentasi konsumen. Profil konsumen memungkinkan pemasar untuk memberikan layanan yang lebih baik kepada konsumen dan untuk berkomunikasi lebih efektif dengan konsumen yang ada. Data demografis dan pribadi konsumen dikumpulkan dan digunakan untuk membentuk profil konsumen. Profil konsumen juga digunakan untuk mencari pelanggan baru menggunakan sumber daya eksternal [3].

Profil konsumen dapat membantu bisnis untuk memahami dan mengenali peluang pasar, mengoptimalkan upaya pemasaran, meningkatkan kepuasan pelanggan, dan membangun hubungan jangka panjang yang lebih baik dengan konsumen [3].

2.2.12 Metode Transformasi Data

Data seringkali dapat dinyatakan dalam skala yang sangat berbeda (misalnya, usia dan gaji). Hal ini dapat menyebabkan beberapa atribut secara tidak sengaja diberi bobot terlalu banyak sehingga atribut lainnya diabaikan secara implisit. Oleh karena itu, penting untuk menormalkan atribut yang berbeda [16]. Ada beberapa metode untuk melakukan standarisasi data, diantaranya yaitu sebagai berikut:

1. Normalisasi Min-Max

Normalisasi min-maks adalah metode transformasi data yang mengubah nilai dalam kumpulan data menjadi rentang tertentu, biasanya 0 hingga 1. Metode ini berguna untuk mempertahankan skala yang konsisten antara atribut dengan rentang nilai yang berbeda. Rumus normalisasi min-max:

$$X_{baru} = \frac{X - \min(X)}{\text{range}(X)} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (6)$$

Normalisasi Min-Max Scaling berguna dalam beberapa situasi, seperti:

a. Data numerik dengan rentang nilai yang besar

Jika data memiliki rentang nilai yang luas, seperti atribut yang diukur dalam unit yang berbeda, normalisasi skala min-maks dapat membantu menyeimbangkan skala dan memastikan bahwa semua atribut memiliki efek yang sebanding pada analisis.

b. Algoritma yang sensitif terhadap skala

Beberapa algoritma seperti algoritma pengelompokan k-means atau algoritma berbasis jarak (misal euclidean), dapat dipengaruhi oleh perbedaan skala antar atribut. Dengan menormalkan rasio Min-Max, data akan memiliki rentang nilai yang seragam, sehingga algoritma ini dapat bekerja lebih baik.

2. Normalisasi Z-Score

Normalisasi Z-Score adalah metode transformasi data yang mengubah nilai-nilai dalam suatu kumpulan data menjadi rata-rata 0 dan standar deviasi 1. Metode ini berguna untuk memastikan bahwa data mengelompok di sekitar rata-rata dan memiliki variasi yang seragam. Rumus normalisasi z-score:

$$Zscore = \frac{X - \text{mean}(X)}{SD(X)} \quad (7)$$