BAB II
KAJIAN LITERATUR

2.1 Tinjauan Pustaka

Bagian ini berisi landasan terkait teori-teori dan metode yang digunakan serta pekerjaan yang sudah dilakukan oleh peneliti sebelumnya untuk mendukung penyelesaian penelitian yang akan dilakukan.

2.1.1 Segmentasi Pelanggan (*Customer Segmentation*)

Pemasaran (*Marketing*) STP (*Segmentation, Targeting, Positioning*) adalah pendekatan yang penting untuk memahami dan memenuhi kebutuhan pelanggan. Dalam proses ini, segmentasi pelanggan menjadi langkah awal yang krusial. Segmentasi pelanggan adalah mengelompokkan individu berdasarkan kesamaan dalam kebutuhan, profil, karakteristik, atau perilaku, yang memungkinkan bisnis untuk merumuskan strategi pemasaran yang lebih efektif. Dengan segmentasi yang tepat, perusahaan dapat meningkatkan hubungan dengan pelanggan, meningkatkan retensi dan loyalitas, serta mengidentifikasi nilai pelanggan dari setiap segmen. Hal ini juga membantu dalam menyesuaikan produk dan layanan agar sesuai dengan preferensi masing-masing segmen, sehingga meningkatkan kepuasan pelanggan dan keuntungan secara keseluruhan [2] [5] [14].

Ada beberapa metode segmentasi pelanggan berdasarkan karakteristik informasi yang serupa [12] [14], antara lain:

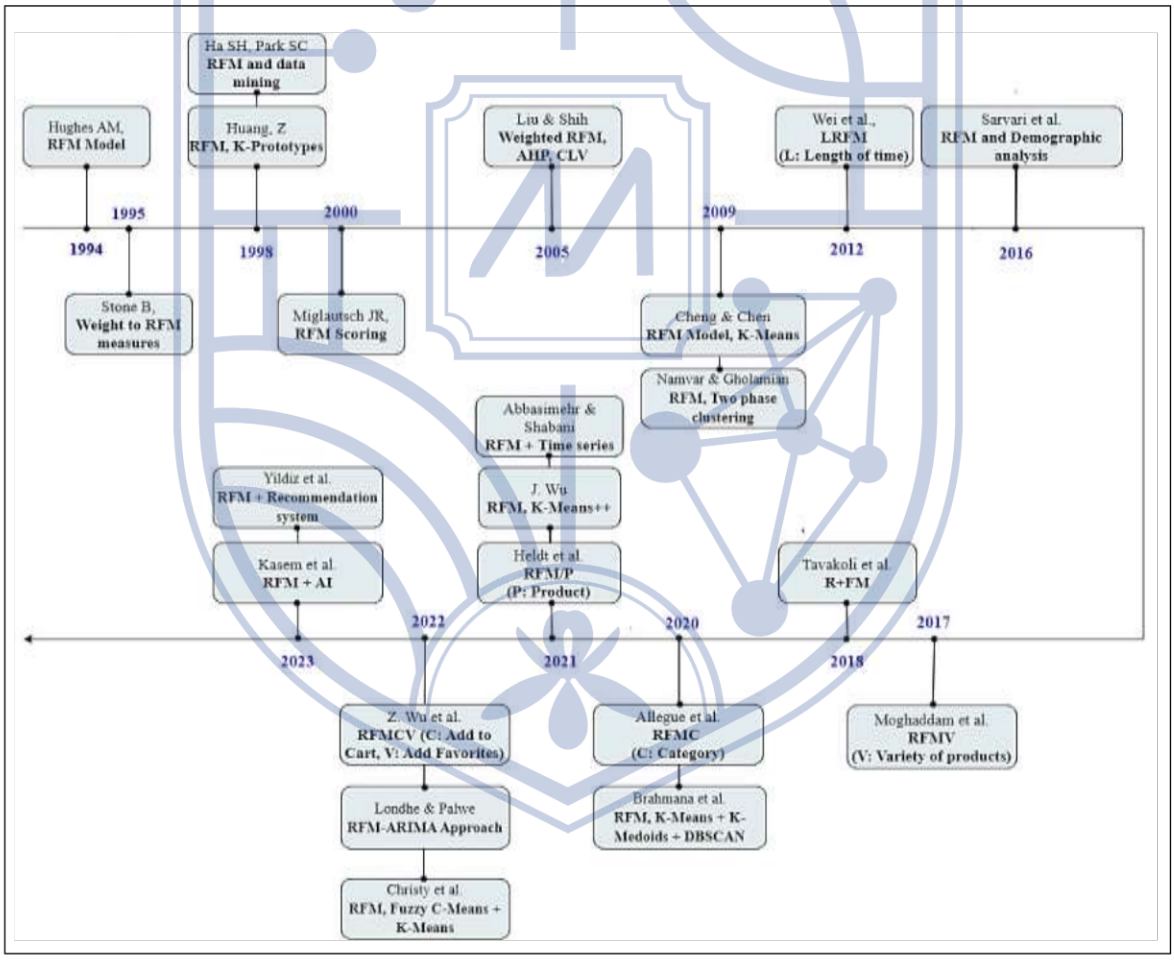
1. **Perilaku (*Behavioral*)**, mencakup manfaat produk atau jasa yang dicari, status pembelian, tingkat penggunaan produk atau jasa, serta frekuensi pembelian [12].
2. **Psikografis (*Psychographic*)**, mencakup tipe kepribadian, gaya hidup (*Lifestyle*) dan nilai-nilai moral [12].
3. **Geografis (*Geographic*)**, mencakup termasuk lokasi seperti negara, provinsi, kota, kabupaten, kode pos, dan iklim [12].
4. **Demografis (*Demographic*)**, mencakup usia, jenis kelamin, jumlah anggota keluarga, luas tempat tinggal, pendapatan, profesi, pendidikan terakhir, status kepemilikan rumah, status sosial (jabatan atau gelar), agama, dan kewarganegaraan [12].

Segmentasi pelanggan dapat dilakukan menggunakan beberapa algoritma, seperti algoritma asosiasi (*Association*), klaster (*Clustering*), klasifikasi (*Classification*), dan

regresi (*Regression*). Di antara beberapa algoritma ini, pengklasteran (*Clustering*) adalah algoritma yang paling tepat dan efektif untuk melakukan segmentasi pelanggan [14].

2.1.2 Model RFM (*Recency, Frequency, Monetary*)

Model RFM pertama kali diperkenalkan oleh Arthur Hughes pada tahun 1994. Dalam model ini terdapat 3 (tiga) variabel utama yaitu R (*Recency*), F (*Frequency*), dan M (*Monetary*), yang dianggap setara dan penting dalam menghitung skor untuk segmentasi pelanggan. Model RFM adalah salah satu segmentasi pelanggan berdasarkan perilaku (*Behavioral*) yang paling umum [14]. Sejak diperkenalkan, model RFM telah berkembang hingga tahun 2023 dan telah banyak diterapkan dalam penelitian terkait segmentasi pelanggan [2] [11].



Gambar 2.1 Sejarah Model RFM [11]

Analisis model RFM merupakan salah satu metode paling populer dan efektif untuk segmentasi pelanggan berdasarkan perilaku serta kebiasaan pembelian pelanggan [2] [3]. Model ini melibatkan 3 (tiga) variabel, yaitu Kekinian (*Recency*), Frekuensi (*Frequency*)

dan Moneter (*Monetary*). Dimana Kekinian (R) berdasarkan pada waktu terkini atau terakhir pelanggan melakukan pembelian hingga saat ini, sementara Frekuensi (F) mengacu pada seberapa banyak atau sering pelanggan melakukan pembelian dalam periode tertentu. Di sisi lain, moneter (M) menunjukkan total jumlah uang yang telah dibelanjakan atau dikeluarkan pelanggan pada pembelian dalam jangka waktu tertentu [2] [3] [4]. Tujuan dari analisis model RFM adalah mengevaluasi nilai pelanggan (*Customer Lifetime Value/CLV*), sehingga dapat diidentifikasi pelanggan yang paling berharga [2] [11] [13].

Gambar 2.2 Model RFM [11]

Proses perhitungan skor RFM untuk setiap pelanggan didasarkan pada 3 (tiga) variabel, yaitu Kekinian (R), Frekuensi (F), dan Moneter (M). Setiap variabel akan diberikan skor dalam rentang 1 hingga 5, dengan penentuan skor menggunakan metode kuintil, di mana setiap kelompok kuintil mencakup 20% dari total data transaksi masing-masing variabel. Pelanggan dikelompokkan ke dalam 5 (lima) kuintil (*Quintile*) yang setara, dengan jangkauan (*Range*) antar kuintil menggunakan *min*, *max*, persentil (*Percentile*) 25%, 50%, dan 75%. Dan setiap pelanggan diberikan skor berdasarkan kuintil sesuai variabel R, F dan M. Untuk kekinian (R), kuintil terkini dengan nilai tertinggi diberikan skor 5, sementara kuintil R lainnya secara berurutan diberi skor 4, 3, 2, dan 1. Proses yang sama diterapkan untuk frekuensi (F) dan moneter (M). Akhirnya, nilai dari ketiga variabel R, F, dan M digabungkan untuk memperoleh skor RFM keseluruhan pelanggan. [4] [11] [13] [14].

Berikut adalah cara untuk menentukan skor RFM yaitu:

1. **Kekinian (R)**, atribut ini menilai dari jarak antara tanggal pembelian terakhir pelanggan (*Last Payment/Buy Date*) dan tanggal saat ini (*Current Date*). Semakin dekat interval waktu tersebut dengan tanggal sekarang, semakin tinggi skor R yang diberikan [4] [12].
2. **Frekuensi (F)**, atribut ini menilai seberapa sering pelanggan melakukan pembelian (*Frequency of the Payment*) dalam periode tertentu. Misalnya, pelanggan A melakukan

pembelian 4 (empat) kali dalam sebulan, sementara pelanggan B hanya melakukannya sekali dalam sebulan, maka skor F untuk pelanggan A akan lebih tinggi, menunjukkan frekuensi pembelian yang lebih sering [4] [12].

3. **Moneter (M)**, atribut ini menilai jumlah uang dikeluarkan pelanggan dari transaksi pembelian (*Value of the Payment*) yang dilakukan dalam jangka waktu tertentu. Semakin besar nominal yang dibelanjakan, semakin tinggi pula skor M yang diberikan [12].

Tabel 2.1 Kuintil skor RFM [12]

Score	Kekinian (*Recency*)	Frekuensi (*Frequency*)	Moneter (*Monetary*)
5	Sangat terkini (*Very Recency*)	Sangat sering (*Very Frequent*)	Sangat tinggi (*Very High*)
4	Terkini (*Recent*)	Sering (*Frequent*)	Tinggi (*High*)
3	Standar (*Standard*)	Normal	Normal
2	Tidak baru-baru ini (*Not Recent*)	Jarang (*Rare*)	Rendah (*Low*)
1	Lama (*Long Ago*)	Langka (*Very Rare*)	Sangat rendah (*Very Low*)

Semua pelanggan dalam model RFM (*Recency, Frequency,* dan *Monetary*) akan diberikan skor yang mencerminkan ketiga dimensi tersebut, seperti 555, 554, 553, ..., 113, 112, 111. Kelompok pelanggan terbaik adalah 555 dan yang terburuk adalah 111 [13] [14] [27]. Masing-masing dimensi R, F, dan M memiliki satuan nilai yang berbeda, sehingga untuk memudahkan analisis dan memastikan akurasi, data RFM harus dinormalisasi terlebih dahulu. Normalisasi bertujuan untuk mengurangi perbedaan skala antara dimensi-dimensi ini, sehingga analisis dapat dilakukan dengan lebih efektif. Salah satu metode yang umum digunakan untuk menormalisasi data adalah metode *Min-Max*. Metode ini berfungsi untuk memetakan indikator ke dalam interval antara 0 dan 1, sehingga setiap nilai dalam dataset akan berada dalam rentang yang sama [2] [15].

Rumus normalisasi metode *Min-Max* pada RFM adalah sebagai berikut (1):

$$R' = \frac{R - Rmin}{Rmax - Rmin} \qquad F' = \frac{F - Fmin}{Fmax - Fmin} \qquad M' = \frac{M - Mmin}{Mmax - Mmin} \quad \text{..... (1)}$$

Keterangan:

R', F', M' = Nilai Normalisasi *Recency, Frequency, Monetary*

R, F, M = Nilai *Recency, Frequency, Monetary*

Rmin, Fmin, Mmin = Nilai Kuintil Paling Rendah dari *Recency, Frequency, Monetary*
Rmax, Fmax, Mmax = Nilai Kuintil Paling Tinggi dari *Recency, Frequency, Monetary*

Analisis RFM tidak hanya efektif dalam segmentasi pelanggan, tetapi juga berguna untuk mengembangkan profil pelanggan. Dengan pendekatan ini, perusahaan dapat menghasilkan definisi baru yang lebih akurat dalam menentukan target, memahami karakteristik setiap segmen, serta menemukan cara terbaik untuk memuaskan pelanggan. Melalui analisis dimensi *Recency, Frequency, dan Monetary*, perusahaan dapat mengidentifikasi kebutuhan dan perilaku pelanggan [2] [12].

2.1.3 Nilai Pelanggan (*Customer Lifetime Value/CLV*)

Customer Lifetime Value (CLV) adalah perkembangan dari manajemen hubungan pelanggan (*Customer Relationship Management* atau CRM), yang bertujuan untuk membangun dan memelihara hubungan yang kuat dengan pelanggan serta meningkatkan nilai pelanggan bagi perusahaan [15]. CLV didefinisikan sebagai "nilai sekarang atau semua keuntungan saat ini yang diperoleh perusahaan dari pelanggan selama hubungan transaksi". Perusahaan dapat menggunakan CLV untuk menghitung dan mengekspresikan profitabilitas pelanggan, yang membantu dalam mengembangkan strategi pemasaran untuk menentukan target pelanggan yang paling menguntungkan [15][16].

CLV ditentukan oleh nilai transaksi yang dilakukan oleh pelanggan selama siklus hidup (*Lifecycle*) hubungan dengan perusahaan atau organisasi. Dengan memahami CLV, perusahaan dapat mengalokasikan sumber daya (*Resources*) dengan efisien seperti dalam manajemen persediaan. Dengan begitu perusahaan dapat fokus dalam menjaga tingkat layanan (*Customer Service Level*) dan kepuasan pelanggan serta menghindari pemborosan sumber daya yang dapat berdampak negatif pada biaya dan persediaan yang tinggi [6] [8] [16].

Selain itu, dengan perhitungan CLV yang tepat, berbagai strategi pemasaran dapat diidentifikasi untuk setiap kategori pelanggan. Misalnya, perusahaan dapat mengelola persediaan produk dengan lebih efektif berdasarkan kategori CLV, pelanggan dengan CLV tinggi dapat mendapatkan akses ke produk atau layanan baru, sementara pelanggan dengan CLV lebih rendah mungkin hanya memiliki akses ke produk standar atau umum. Pendekatan ini memungkinkan perusahaan untuk lebih memahami dan memenuhi kebutuhan pelanggan, meningkatkan pengalaman mereka, dan pada akhirnya meningkatkan loyalitas serta retensi pelanggan [6] [7] [16].

Perhitungan CLV berdasarkan inputan skor dari variabel model RFM. Dimana setiap pelanggan berasal dari data transaksi yang sama dengan analisis RFM. Selain itu, tidak perlu membagi data menjadi (atau lebih) beberapa periode waktu. Keseluruhan data transaksi pelanggan menjadi sampel tunggal untuk memperkirakan CLV. Setelah melakukan perhitungan CLV, maka akan menghasilkan peringkat setiap pelanggan berdasarkan nilai RFM yang telah dinormalisasi sebelumnya. Dengan menghitung CLV setiap pelanggan, perusahaan dapat mengkategorikan pelanggan berdasarkan kontribusi masing-masing terhadap profit. CLV dapat membantu mengembangkan strategi yang lebih efektif untuk menangani setiap pelanggan secara berbeda, dibandingkan memperlakukan setiap pelanggan dengan cara yang sama dengan menggunakan pendekatan dan startegi pemasaran yang sama. Meskipun perusahaan atau organisasi tertarik untuk mengetahui CLV pelanggan saat ini, mereka juga perlu mengidentifikasi faktor-faktor yang dapat mereka kendalikan yang berpotensi meningkatkan CLV. Karena tidaklah cukup hanya mengetahui siapa pelanggan yang paling berharga atau paling menguntungkan, tetapi yang lebih penting adalah bagaimana cara mengkonversi pelanggan yang saat ini kurang menguntungkan menjadi pelanggan yang lebih menguntungkan [2] [3] [4] [15].

Rumus perhitungan nilai pelanggan (*Customer LifeTime Value*/*CLV*) (2):

$$CLV = \frac{(wR\ x\ nR\ +\ wF\ x\ nF\ +\ wM\ x\ nM)}{wR\ +\ wF\ +\ wM} \quad \text{..... (2)}$$

Keterangan:

CLV = Nilai Pelanggan (*Customer Lifetime Value*)

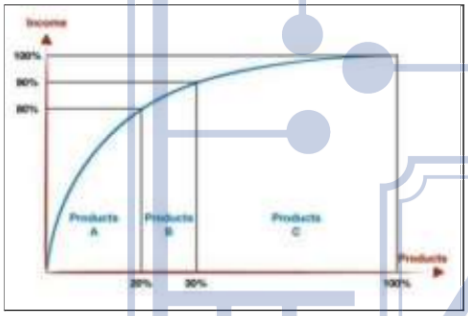
nR, nF, nM = Nilai Normalisasi *Recency, Frequency, Monetary*

wR, wF, wM = Nilai Pembobotan (*Weighted*) dari Variabel *Recency, Frequency, Monetary*

2.1.4 Klasifikasi Permintaan (*Demand Classification*)

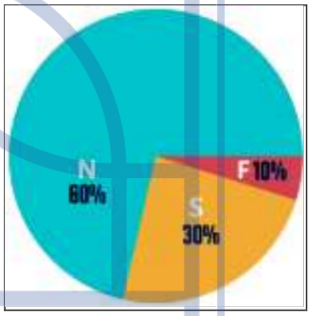
Banyak ahli meyakini bahwa klasifikasi permintaan yang tepat dapat memberikan wawasan dan pengetahuan baru untuk membantu perusahaan dalam memahami permintaan pelanggan terhadap persediaan. Salah satu klasifikasi yang paling sering diterapkan adalah klasifikasi ABC, yang juga dikenal sebagai klasifikasi Pareto (berasal dari Ekonom asal Italia Vilfredo Pareto). Metode ini menggunakan prinsip 80:20 yang berfungsi sebagai dasar untuk analisis [9] [28]. Biasanya digunakan dalam banyak industri untuk mengidentifikasi berbagai kategori atau pengelompokkan entitas (seperti produk, layanan, pelanggan,

penyuplai, dan lainnya) dengan tingkat kepentingan yang berbeda-beda [9] [10]. Dalam manajemen persediaan, metode klasifikasi ABC digunakan untuk mengelompokkan permintaan (*Demand*) ke dalam kategori berdasarkan kapasitas (*Volume*). Kategori A mencakup permintaan dengan kapasitas paling besar, kategori B permintaan dengan kapasitas sedang, dan kategori C permintaan dengan kapasitas rendah [17]. Selain klasifikasi ABC, terdapat juga pendekatan FSN (*Fast, Slow dan Non-Moving*), yang mengelompokkan produk berdasarkan kecepatan permintaan (*Demand Velocity*). Dalam pendekatan FSN, permintaan produk dikategorikan menjadi *Fast Moving* (bergerak cepat), *Slow Moving* (bergerak lambat) dan *Non-Moving* (tidak bergerak) [6] [10] [17] [28].



Gambar 2.3 Klasifikasi ABC

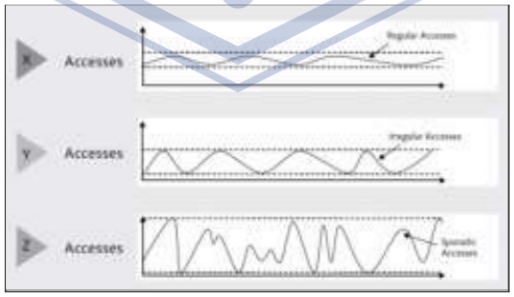
Sumber: Google.com



Gambar 2.4 Klasifikasi FSN

Sumber: Google.com

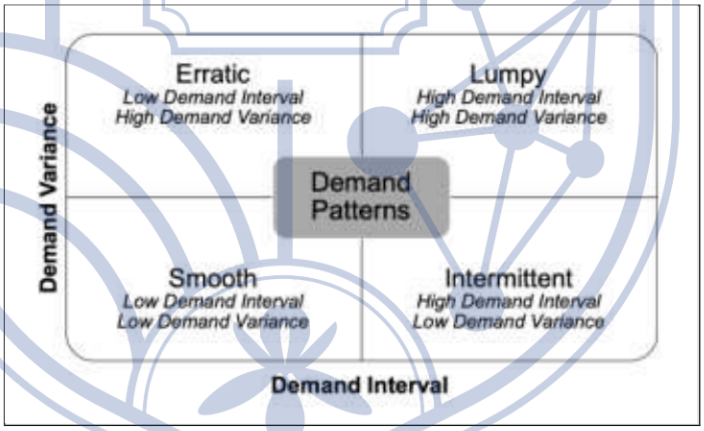
Klasifikasi lain yang ditemukan dalam literatur adalah klasifikasi XYZ, yang mengelompokkan persediaan berdasarkan fluktuasi kebutuhan (*Consumption*). Kategori X untuk produk dengan kebutuhan hampir konstan atau stabil (*Regular*), kategori Y untuk produk dengan tingkat kebutuhan fluktuasi akibat pengaruh tren atau musiman (*Irregular*), dan kategori Z untuk produk dengan kebutuhan permintaan tidak rutin (*Sporadic*) [17].



Gambar 2.5 Klasifikasi XYZ

Sumber: Google.com

Namun, baik metode klasifikasi ABC, FSN, dan XYZ hanya mempertimbangkan satu kriteria tunggal, sehingga banyak dimensi dan area manajemen terabaikan [6] [10] [17] [28]. Bagi suatu perusahaan atau organisasi, tujuan strategis manajemen persediaan adalah mencari keseimbangan tingkat stok (*Level of Stock*) sesuai dengan kebutuhan permintaan pelanggan, serta mengurangi dan menghindari biaya yang tidak perlu [8] [10]. Berbeda dengan klasifikasi sebelumnya, klasifikasi permintaan yang diajukan oleh Syntetos, Boylan dan Croston adalah dengan menggabungkan beberapa kriteria dalam analisis untuk mempertimbangkan karakteristik pola permintaan. Metode ini mengelompokkan permintaan berdasarkan variabilitas dan fluktuasi, sehingga memberikan pemahaman yang lebih baik tentang dinamika kebutuhan pelanggan. Melalui cara ini, manajemen persediaan dapat lebih responsif terhadap perubahan permintaan serta mendukung strategi bisnis dalam mempertahankan dan kepuasan pelanggan [5] [6] [8] [17]. Klasifikasi permintaan dari Syntetos, Boylan dan Croston membagi karakteristik pola permintaan menjadi 4 (empat) kuadran yaitu *Smooth, Erratic, Intermittent,* dan *Lumpy* [8] dengan menggunakan 2 (dua) parameter utama, koefisien variasi permintaan kuadrat (*Demand Variation Coefficient/CV*^{2}) pada posisi sumbu Y dan interval permintaan rata-rata (*Average Demand Interval/ADI*) pada posisi sumbu X [8] [9] [17].



Gambar 2.6 Klasifikasi Pola Permintaan [8]

CV^2 didefinisikan sebagai rasio kuadrat dari standar deviasi data permintaan dibagi dengan rata-rata permintaan, yang menunjukkan tinggi-rendah (*High-Low*) variabilitas (*Variance*) kuantitas permintaan. Sedangkan ADI didefinisikan sebagai pengukuran jumlah rata-rata periode waktu antar selang dua periode permintaan yang berhasil terjadi transaksi, yang menunjukkan tinggi-rendah (*High-Low*) intermittensi (*Interval*) atau fluktuasi periode permintaan [8] [18]. Berdasarkan Syntetos, Boylan dan Croston nilai ambang batas (*Threshold*) untuk CV^2 adalah 0.49 dan untuk ADI adalah 1.32 [6] [8] [9] [10] [18].

Dengan rumus perhitungan CV^2 dan ADI [17] [18] [20] adalah sebagai berikut (3) (4):

$$CV = \frac{Demand\ Standard\ Deviation}{Demand\ Mean} \qquad CV^2 = \left(\frac{\sqrt{\sum_{i=1}^{n} (\mathcal{E}_i - \mathcal{E}_a)^2 / n}}{\mathcal{E}_a} \right) 2 \quad \text{.... (3)}$$

Keterangan:

CV = Nilai Koefisien Variasi (*Coefficient of Variation*)

$\mathcal{E}_i$ = Nilai Permintaan (*Demand Product*) suatu Periode i

$\mathcal{E}_a$ = Nilai Rata-rata Permintaan (*Average Product Demand*)

n = Nilai Jumlah Periode (*Number of Time Periods*)

$$ADI = \frac{Total\ Periods}{Total\ Demand\ Buckets} \qquad ADI = \frac{\sum_{i=1}^{n} p_i}{n} \quad \text{..... (4)}$$

Keterangan:

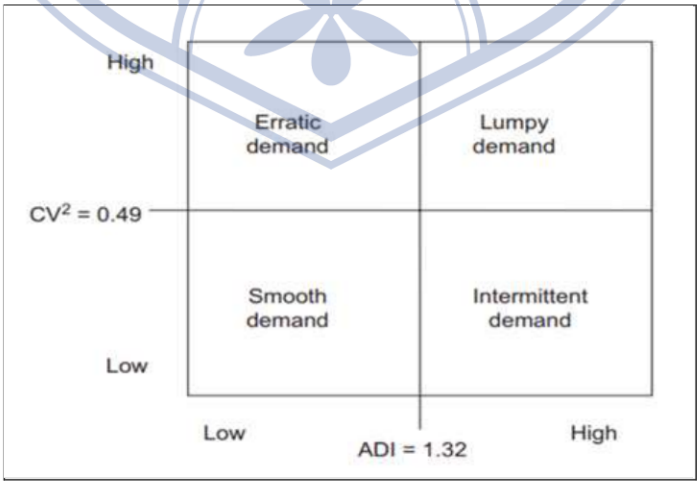
ADI = Nilai Interval Permintaan Rata-rata (*Average Demand Interval*)

p_i = Total Jumlah Periode

n = Nilai Jumlah Periode Permintaan bernilai bukan nol (Periode terjadi Transaksi)

i = Nilai Indeks Periode Permintaan

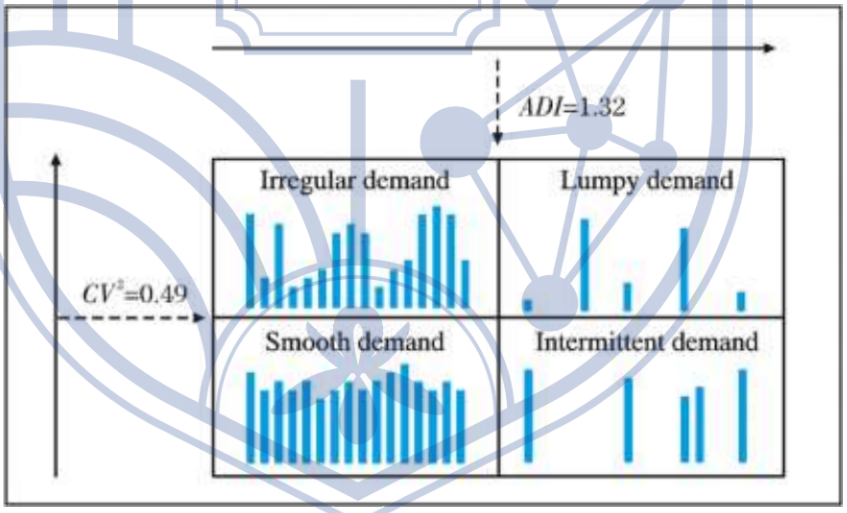
Titik batas untuk parameter CV^2 dan ADI yang digunakan dalam klasifikasi ini telah ditetapkan dan ditunjukkan pada Gambar 2.7 dibawah ini [10].



Gambar 2.7 Nilai Ambang Batas (*Threshold*) Klasifikasi Permintaan [10]

Berkat kedua parameter ini, klasifikasi dapat dibuat menjadi 4 (empat) kuadran:

1. **Permintaan Lancar (*Smooth/Soft Demand*)**, dengan karakteristik permintaan teratur dan rutin (*Regular*), hampir setiap periode terdapat permintaan, dengan koefisien CV^2 rendah dan ADI rendah (ADI < 1.32 and CV^2 < 0.49) [10] [18].
2. **Permintaan Tidak Menentu (*Erratic Demand*)**, dengan karakteristik permintaan teratur dari waktu ke waktu, namun variasi permintaan untuk jumlah kuantitas yang tinggi. Memiliki interval antar selang periode permintaan rendah (rutin) dan koefisien permintaan CV^2 tinggi dan ADI rendah (ADI < 1.32 and $CV^2 \geq 0.49$) [10] [18].
3. **Permintaan Terputus-putus / Tidak Rutin (*Intermittent Demand*)**, dengan karakteristik periode permintaan sedikit lebih acak, terdapat sedikit variasi permintaan untuk jumlah kuantitas dan variasi yang lebih besar dalam interval antar selang dua periode permintaan, dengan koefisien CV^2 rendah dan ADI tinggi (ADI ≥ 1.32 and CV^2 < 0.49) [10] [18].
4. **Permintaan Tidak Stabil (*Lumpy/High Demand*)**, dengan karakteristik permintaan variabilitas kuantitas yang diminta dan fluktuasi interval periode permintaan nya tinggi, untuk koefisien CV^2 tinggi dan ADI tinggi (ADI ≥ 1.32 and $CV^2 \geq 0.49$) [10] [18].



Gambar 2.8 Klasifikasi Permintaan (*Demand Classification*) [9]

2.1.5 Penggalian Data (*Data Mining*)

Penggalian data (*data mining*) merupakan suatu proses eksplorasi dan analisis data dalam jumlah besar untuk menemukan pola, hubungan kompleks, dan mengidentifikasi informasi yang berguna. Proses ini melibatkan penerapan algoritma dan teknik analisis untuk mengeksplorasi data, dengan tujuan menghasilkan pengetahuan baru [5]. Dalam industri

bisnis, penggalian data memegang peran penting bagi perusahaan dan organisasi, terutama dalam meningkatkan efektivitas proses pengambilan keputusan, membuka peluang baru maupun inovasi. Perusahaan yang berhasil memanfaatkan teknik dan alat *data mining* dengan baik akan dapat meningkatkan keunggulan strategis dan kompetitif mereka [5] [19]. Beberapa teknik rekayasa penggalian data yang memberikan keberhasilan signifikan dalam menghasilkan pengetahuan atau informasi baru adalah teknik klasifikasi (*Classification*), pengklasteran (*Clustering*) dan regresi (*Regression*) [12]. Salah satu penggalian data yang berharga adalah data pelanggan, dimana saat ini perusahaan atau organisasi modern tidak lagi berfokus pada kemudahan bertransaksi dan mengutamakan produk (*Product Centric*), tetapi juga perlu menerapkan startegi yang berfokus pada pelanggan (*Customer Centric*) [19].

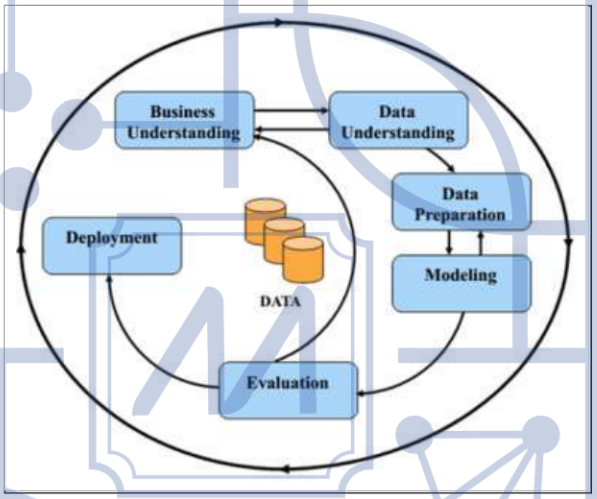
Penggalian data dapat digunakan untuk mengembangkan segmentasi pelanggan yang juga dapat mengidentifikasi perilaku pelanggan. Dengan cara ini, segmentasi akan membagi pelanggan sesuai karakteristik pelanggan yang memiliki kesamaan (seperti jumlah transaksi belanja, produk atau layanan yang digunakan, frekuensi transaksi, dan lainnya), dengan begitu perusahaan atau organisasi dapat mengetahui kelompok pelanggan yang berharga dan prioritas [5] [12].

Salah satu model yang banyak digunakan dalam teknik penggalian data yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*) [5] [19]. Menurut model CRISP-DM, terdapat 6 (enam) tahapan dalam proses penggalian data, antara lain:

1. **Pemahaman Bisnis (*Business Understanding*)**, pada tahap ini tujuannya untuk memahami gambaran secara keseluruhan dari proses penggalian data yang akan dilakukan berdasarkan startegi dan target dari bisnis [5].
2. **Pemahaman Data (*Data Understanding*)**, tahap kedua mencakup proses identifikasi, pengumpulan, deskripsi, seleksi, relevansi dan evaluasi kualitas data untuk memastikan kebutuhan penggalian data dapat mencapai tujuan yang diharapkan [5].
3. **Persiapan Data (*Data Preparation*)**, tahap ini mencakup proses pembersihan data dari *outlier* seperti data anomali (menyimpang), data redundansi (duplikasi), serta mendeteksi distorsi dan menghilangkan data yang tidak valid, tujuannya untuk kesesuaian dan kelancaran proses permodelan [5].
4. **Pemodelan (*Modeling*)**, tahap ini meliputi pemilihan teknik pemodelan, menjelaskan dan mengembangkan model yang diusulkan serta penerapannya sesuai dengan data yang sudah tersedia dan siap untuk memenuhi tujuan bisnis [5].

5. **Evaluasi (*Evaluation*)**, tahap pemodelan yang telah dikembangkan akan melalui proses penilaian, pengujian dan evaluasi terhadap tingkat akurasi dan generalitasnya sebelum dilakukan penerapan secara keseluruhan dalam lingkungan bisnis [5].
6. **Penerapan (*Deployment*)**, setelah melakukan evaluasi, jika hasilnya konsisten dengan target dan tujuan utama bisnis, maka selanjutnya tahap *deployment* siap diluncurkan (*Release*) dengan versi baru. Agar masalah segera teratasi dan kebutuhan bisnis dapat terpenuhi setelah diterapkan [5].

Gambar 2.9 dibawah menunjukkan tahapan-tahapan dari model CRISP-DM [19].



Gambar 2.9 Model CRISP-DM (*Data Mining)* [19]

Mengolah data mentah (*Raw Data*) menjadi data berkualitas dalam prosedur penggalian data sangat penting, karena kualitas data yang tinggi berkontribusi pada akurasi hasil analisis (menjadi sumber data input pemodelan). Pembersihan data (*Data Cleaning*) adalah langkah awal yang penting, meliputi penghapusan data atau informasi yang tidak relevan dan pengisian nilai yang hilang (imputasi) pada suatu kolom (*Field/Attribute*). Sebelum menangani data yang hilang (*Missing Value*), analisis dan identifikasi antar kolom harus dilakukan untuk memastikan pemilihan kolom dan baris (*Record*) sesuai dengan kebutuhan analisis agar mencegah pengurangan kualitas pengetahuan yang diperoleh dari suatu *dataset*. Sedangkan untuk nilai yang hilang dalam suatu data yang berisik (*Noise Data*), perlu diterapkan metode *forward fill* penting untuk mengisi baris atau kolom NaN (*Not a Number*) dan NaT (*Not a Time*) berdasarkan data terakhir yang valid. Dengan menjaga kualitas data melalui langkah-langkah ini, hasil dari penggalian data dapat memberikan

wawasan yang akurat dan pengetahuan yang bermanfaat pada perusahaan maupun organisasi dalam pengambilan keputusan [19].

2.1.6 Metode Analisis Data (*Data Analysis Method*)

Merupakan proses mengubah data yang dikumpulkan menjadi informasi yang berharga dengan menerapkan berbagai metode maupun teknik seperti pemodelan untuk menemukan pola, tren, hubungan dan kesimpulan guna mendukung pengambilan keputusan dan menyelesaikan masalah [24]. Analisis data merupakan suatu proses di mana data mentah (*Raw Data*) disusun dan dikelola untuk mendapatkan informasi yang berguna. Pengelolaan dan pemahaman terhadap data adalah langkah penting untuk mengetahui apa yang terkandung dan tidak terkandung di dalamnya. Ada berbagai cara untuk melakukan analisis data, namun data bisa dengan mudah dimanipulasi untuk mendukung kesimpulan atau tujuan tertentu. Oleh karena itu, sangat penting untuk memeriksa dengan cermat terhadap data yang disajikan dan berpikir kritis mengenai hasil yang di dapat dari analisis data [25].

Data mentah untuk analisis dapat berupa pengukuran, hasil survei, atau observasi. Walaupun data mentah tersebut sangat berguna, tetapi jumlahnya yang besar bisa membuatnya sulit untuk diproses dan diolah. Selama proses analisis, data mentah akan dikelola dengan cara yang lebih sistematis, misalnya menghitung hasil survei (lewat kuesioner) untuk melihat pola jawaban atau respons tertentu, sehingga pengguna dengan mudah mendapatkan dan memperhatikan tren, pola atau informasi yang muncul [23] [25].

Selain itu, penggunaan visualisasi data seperti bentuk tabel, grafik, bagan, atau teks dirancang untuk menyajikan informasi secara jelas dan mudah dipahami. Data mentah dapat disertakan dalam bentuk lampiran jika diperlukan untuk memperoleh informasi detail lebih lanjut. Namun, saat memeriksa data dan kesimpulan, penting untuk mengevaluasi asal-usul data, metode pengumpulan, serta jumlah atau ukuran sampel yang digunakan. Apabila terdapat potensi konflik kepentingan (*Conflict of Interest*) atau masalah dengan sampel yang digunakan, hasil analisis tersebut mungkin diragukan. Seorang analis data (*Data Analyst*) yang berintegritas biasanya akan mengungkapkan secara transparan metode pengumpulan data, sumber data, dan tujuan analisis agar pengguna dapat melakukan evaluasi dan pengambilan keputusan dengan tepat [25]. Berikut ada 2 (dua) komponen penting dalam proses analisis data:

1. **Analisis dan Statistik Deskriptif (*Descriptive Analysis and Statistics*)**

   Analisis data yang pertama dikenal dan digunakan adalah analisis deskriptif, dan dianggap sebagai metode analisis data yang paling sederhana, efisien dan mudah diterapkan. Hal ini membuat analisis deskriptif sangat cocok digunakan untuk

menangani data (*Dataset*) dalam jumlah besar. Proses analisis deskriptif ini, dapat dilakukan pada keseluruhan data atau hanya pada beberapa sampel data. Metode ini membantu mendeskripsikan dan meringkas data menjadi bermakna dalam bentuk suatu presentasi atau kesimpulan sederhana sebagai hasilnya (*Output*), yang berupa statistik deskriptif. Dimana hasilnya disajikan dalam bentuk tabel dengan data angka/numerik (*Numeric*). Mengenai analisis deskriptif untuk data yang berkelanjutan (*Continuous*), dapat ditampilkan dengan nilai rata-rata (*Mean*) dan deviasi (*Deviation*), yang memberikan gambaran mengenai variasi data. Sedangkan untuk pengkategorian data, dapat ditampilkan melalui nilai persentase (*Precentage*) dan frekuensi (*Frequency*) [24] [25]. Statistik deskriptif tidak memungkinkan kita untuk menarik kesimpulan apa pun di luar data yang tersedia, tetapi membantu dalam menafsirkan data yang ada [21]. Dalam statistik deskriptif menunjukkan hubungan antara variabel dalam sampel tertentu, dan sering digunakan untuk merapikan dan meringkas sebaran data, yang sangat penting untuk membuat perbandingan statistik inferensial dan penelitian. Statistik deskriptif sering digunakan dalam pemodelan statistik. Model statistik akan menghasilkan suatu probabilitas dan informasi baru melalui eksperimen yang dilakukan menggunakan data pengetahuan sebelumnya. Ada beberapa pengukuran dalam statistik deskriptif, antara lain pengukuran tendensi sentral (*Measures of Central Tendency*) dan pengukuran penyebaran (*Measures of Spread*) [21] [23]. Beberapa statistik deskriptif umum yang sering digunakan untuk pengukuran tendensi sentral (*Measures of Central Tendency*) adalah sebagai berikut:

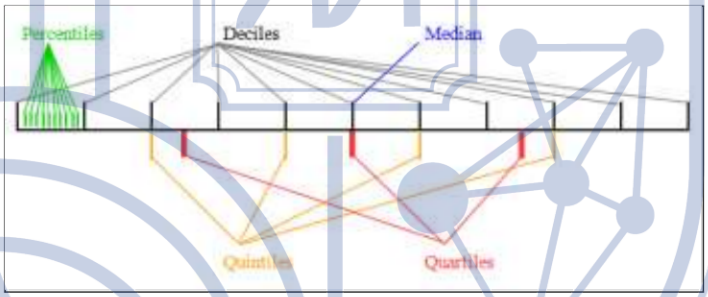
- **Rata-rata (*Mean*)**, merupakan nilai rata-rata aritmatika dari sekumpulan data dibagi dengan jumlah observasi. Contoh *mean*, terdapat sampel data usia siswa dengan jumlah observasi sebanyak 5 yaitu 15, 17, 19, 14, 13, maka hasil *mean* adalah 15,6. Dengan rumus perhitungan *mean* (15+17+19+14+13) / 5 = 15,6 [23]. *Mean* dapat disebut sebagai pengukuran tendensi sentral yang paling intuitif [21].
- **Nilai Tengah (*Median*)**, adalah nilai di tengah statistik yang disusun dari terendah ke tertinggi atau sebaliknya. Jika jumlah data nilai genap, *median* adalah rata-rata dari dua nilai tengah. Namun, jika jumlah data nilai ganjil, *median* hanyalah nilai tengah. Contoh *median* dengan jumlah data nilai ganjil menggunakan contoh data *mean* di atas, dengan menyusun ulang secara berurutan: 13, 14, 15, 17, 19, maka hasil *median* adalah 15. Contoh *median* dengan jumlah data nilai genap, yaitu 4, 7, 10, 13, 15, 18, maka hasil *median* adalah (10 + 13) / 2 = 11,5 [23]. Ketika distribusi

data tidak merata atau terdapat nilai ekstrem, *median* mungkin merupakan ukuran tendensi sentral yang lebih baik [21].

- **Modus (*Mode*)**, merupakan nilai yang paling sering muncul dari sejumlah atau sekumpulan data [21] [23]. Contoh *mode*, menggunakan sejumlah data berikut 3, 5, 7, 7, 8, 9, 10, maka hasil *mode* adalah 7 [21] [23].
- **Jangkauan (*Range*)**, adalah selisih antara nilai maksimum (terbesar) dan minimum (terkecil) dalam sekumpulan data 9, 3, 5, 8, 4, 6, sehingga di dapat *max* adalah 9 dan *min* adalah 3, maka *range* adalah 9 - 3 = 6 [21].

Sedangkan untuk statistik deskriptif umum yang sering digunakan untuk pengukuran penyebaran (*Measures of Spread*) adalah sebagai berikut:

- **Persentil, Desil, Kuintil dan Kuartil (*Percentile, Deciles, Quintile and Quartile*)**, dengan menggunakan persentil, desil, dan kuartil, kita dapat membagi kumpulan data menjadi 100 (seratus) bagian yang sama (*Percentile*), 10 (sepuluh) bagian yang sama (*Deciles*), dan 5 (lima) bagian yang sama (*Quintile*), dan 4 (empat) bagian yang sama (*Quartile*) untuk kumpulan data yang telah diurutkan [14] [23].



Gambar 2.10 Persentil, Desil, Kuintil, dan Kuartil
Sumber: Google.com

- **Varians (*Variance*)**, merupakan ukuran dari variabilitas (keberagaman) dalam data. Varians selalu bernilai bukan negatif (*Non-Negative*). *Variance* digunakan untuk menggambarkan sebaran atau distribusi data, dengan kata lain *variance* menunjukkan seberapa dekat data setiap nilai yang diamati dengan nilai rata-rata (*Mean*) [21]. Berikut rumus untuk perhitungan *variance* populasi dan *variance* sampel (5):

$$\sigma^2 = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \mu)^2}}{N}$$

Rumus *Variance* Populasi

$$s^2 = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}{n}$$ Rumus *Variance* Sampel ..... (5)

- **Standar Deviasi (*Standard Deviation*)**, secara sederhana adalah akar kuadrat dari *variance*. Standar Deviasi menyatakan variabilitas (tingkat fluktuasi) di sekitar nilai rata-rata (*Mean*) dari nilai awal suatu unit atau variabel (*Variable's Original Units*) [21]. Berikut rumus untuk perhitungan Standar Deviasi populasi dan sampel (6):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}} \qquad s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$ ..... (6)

Rumus Standar Deviasi Populasi Rumus Standar Deviasi Sampel

2. **Visualisasi Data (*Data Visualization*)**

Merupakan seni analisis data dalam menampilkan suatu informasi, visualisasi data sangat bergantung pada jenis data yang disajikan agar pesan-pesan utama yang ingin disampaikan dapat tersampaikan dengan jelas. Karakteristik kebutuhan informasi dari pengguna sangat penting, karena merupakan sasaran dan target utama dalam visualisasi data. Oleh karena itu, pemilihan fitur kontekstual yang tepat sangat berpengaruh, agar interpretasi dan pemahaman materi dapat dilakukan dengan lebih tepat dan akurat. Semua visualisasi data memiliki serangkaian parameter umum untuk estetika tampilan desain, konsisten dengan tema yang digunakan, sehingga meningkatkan desain bagan (*Chart Design*) yang baik [21]. Jenis-jenis visualisasi data yang sering digunakan sebagai berikut:

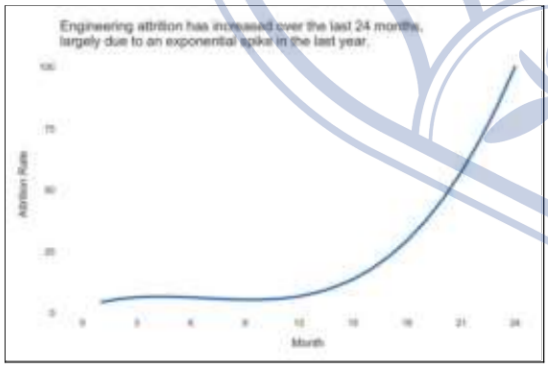
- **Tabel (*Tables*)**, merupakan cara paling dasar untuk mengelola data. Karena tabel umumnya berisi banyak data metrik, tabel biasanya lebih baik ditempatkan sebagai materi referensi di lampiran dokumen atau di dalam suatu dashboard (bentuk *drill-through*) daripada menempati lokasi utama yang seharusnya dapat dimanfaatkan secara startegis untuk memfokuskan perhatian pengguna pada pesan-pesan utama dan penting [21].

Category	Status	Avg Price	Last Year	This Year	Goal
100-Groceries		$1.36	$810,176	$829,778	$810,176
090-Home		$3.28	$2,913,647	$3,053,328	$2,913,647
080-Accessories		$4.22	$1,273,096	$1,370,239	$1,273,096
070-Hosiery		$3.37	$573,604	$486,106	$573,604
060-Intimate		$4.02	$955,370	$852,329	$955,370
050-Shoes		$13.73	$3,640,471	$3,574,900	$3,640,471
040-Juniors		$7.06	$3,105,550	$2,930,283	$3,105,550
030-Kids		$5.20	$2,726,882	$2,703,490	$2,726,882
020-Mens		$6.89	$4,453,133	$4,452,421	$4,453,133
010-Womens		$6.70	$2,940,662	$1,787,938	$2,940,662
Total		**$5.19**	**$23,132,601**	**$22,051,952**	**$23,132,601**

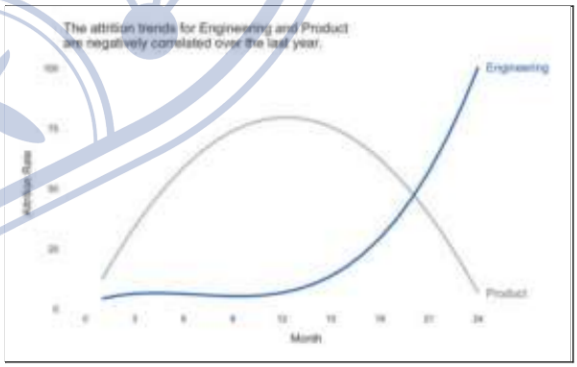
Gambar 2.11 Contoh Visualisasi Data Tabel

Sumber: Google.com

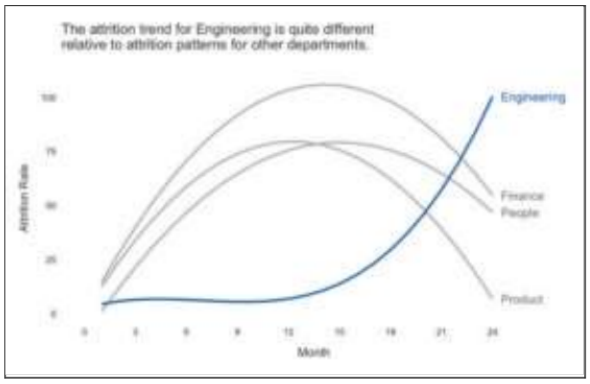
- **Grafik Garis (*Line Graphs*)**, Grafik garis digunakan untuk memvisualisasikan data berkelanjutan dari waktu ke waktu. Saat menggambarkan tren data (contoh data penjualan), penting untuk menghindari penggunaan data kumulatif, karena grafik tersebut akan menunjukkan tren yang positif meskipun secara data non kumulatif menunjukkan tren yang menurun. Pada grafik garis terdapat 3 (tiga) jenis yaitu grafik garis seri tunggal (*Single Series*), dua seri (*Two Series*), dan multipel seri (*Multiple Series*). Grafik garis seri tunggal (*Single Series*) adalah jenis grafik garis yang paling dasar, yang menggambarkan tren untuk data 1 (satu) kelompok atau kategori saja. Grafik garis dua seri (*Two Series*) menggambarkan tren antar data 2 (dua) kelompok, digunakan untuk mendefinisikan kelompok yang dibedakan dengan 2 (dua) garis. Grafik garis multipel seri (*Multiple Series*) menggambarkan tren untuk 2 (tiga) kelompok atau lebih. Jika fokusnya adalah membandingkan satu kelompok (contoh *Engineering*) dengan dua kelompok lainnya, tidak perlu membedakan kelompok lain berdasarkan warna, cukup menggunakan label [21].



Gambar 2.12 Contoh Visualisasi Data Grafik Garis Satu Seri (*Single Series*) [21]

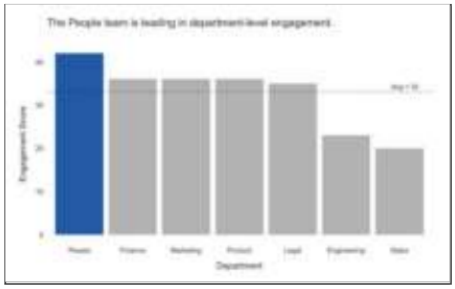


Gambar 2.13 Contoh Visualisasi Data Grafik Garis Dua Seri (*Two* Series) [21]

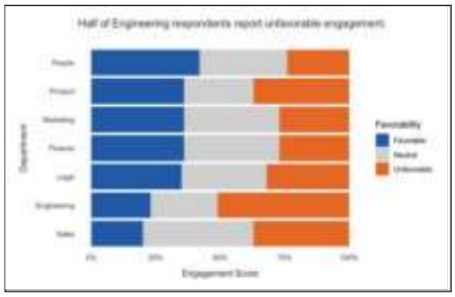


Gambar 2.14 Contoh Visualisasi Data Grafik Garis Multipel Seri (*Multiple* Series) [21]

- **Bagan Batang (*Bar Charts*)**, digunakan untuk menampilkan pengkategorian data. Terdapat 4 (empat) jenis bagan batang yang umum adalah vertikal, horizontal, bertumpuk (*Stacked*), dan dua arah (*Bidirectional*). Seperti grafik garis, bagan batang dapat berupa seri tunggal, 2 (dua) seri, atau beberapa seri sesuai dengan data yang perlu ditampilkan. Bagan batang vertikal dan horizontal adalah metode paling dasar dan umum untuk memvisualisasikan kategori data. Bagan batang bertumpuk 100% (*Stacked Bar Chart*) berguna untuk mengilustrasikan kontribusi relatif subkomponen suatu kategori. Bagan batang bertumpuk (*Stacked*) merupakan alat yang efektif untuk memvisualisasikan distribusi keseluruhan item serta berbagai dimensi kategoris, contoh departemen, lokasi, atau profil pekerjaan. Penyesuaian yang diperlukan untuk membuat bagan area bertumpuk (*Stacked*) adalah dengan menentukan variabel kategori untuk pengurutan terhadap parameter yang digunakan. Diagram batang dua arah (*Bidirectional*) merupakan visual yang efektif untuk membandingkan dua metrik secara berdampingan pada keseluruhan nilai variabel kategoris. Diagram batang dua arah terkadang disebut sebagai diagram batang divergen (*Divergent Bar Chart*), diagram batang *back-to-back*, atau diagram batang cermin (*Mirror Bar Chart*) [21].

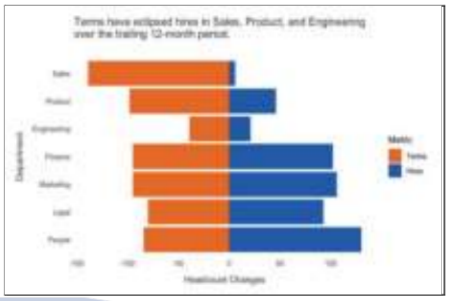


Gambar 2.15 Contoh Visualisasi Data *Vertical Bar Chart* [21]

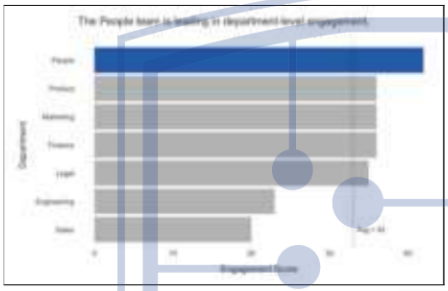


Gambar 2.16 Contoh Visualisasi Data *Horizontal Bar Chart* [21]

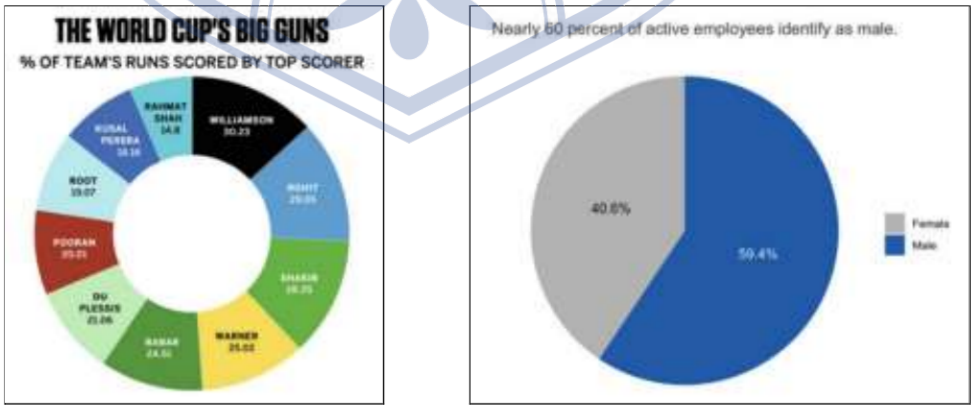
Gambar 2.17 Contoh Visualisasi Data *100% Stacked Bar Chart* [21]



Gambar 2.18 Contoh Visualisasi Data *Bidirectional Bar Chart* [21]

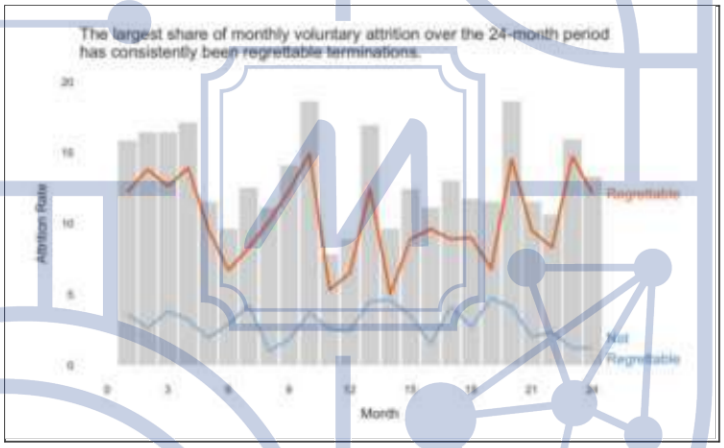


- **Diagram Lingkaran/Bagan Pai (*Pie Charts*)**, Diagram lingkaran (*Pie Chart*) kurang efektif untuk memvisualisasikan data. Karena sulit untuk menentukan ukuran relatif yang sesuai, apabila terdapat banyak atau lebih beberapa kelompok atau kategori. Diagram lingkaran dapat sesuai digunakan jika ada 2 (dua) atau 3 (tiga) kelompok yang saling eksklusif dan menyeluruh, dan kita perlu memvisualisasikan kontribusi relatif masing-masing kelompok atau kategori terhadap keseluruhan. Selain itu, proporsi nilai harus selalu berjumlah 1 atau 100% (secara total), jadi jika data bukan merupakan bagian yang saling eksklusif dari suatu keseluruhan, penggunaan diagram lingkaran untuk analisis data menjadi tidak tepat [21].



Gambar 2.19 Contoh Visualisasi Data *Pie Chart* [21]

- **Bagan Kombinasi** (***Combination Charts***), menampilkan data menggunakan berbagai jenis visualisasi ke dalam bagan yang sama. Contoh dalam konteks analisis data, kita mungkin dapat menyoroti perbedaan antara tren *turnover* yang disesalkan (buruk) dan yang tidak disesalkan (baik) yang relatif terhadap total tingkat *turnover* yang terjadi. Biasanya lebih sulit untuk membandingkan besarnya tingkat yang disesalkan (*Regrettable*) dan yang tidak disesalkan (*Not-Regrettable*) dari waktu ke waktu hanya menggunakan bagan batang (*Bar Charts*), dan bagan kombinasi sering kali merupakan metode yang lebih intuitif untuk menyajikan informasi ini. Seperti yang diilustrasikan pada Gambar 2.20, kita dapat memanfaatkan bagan garis 2 (dua) seri untuk tingkat yang disesalkan (*Regrettable*) dan yang tidak disesalkan (*Not-Regrettable*) bulanan yang relatif terhadap total tingkat *turnover* yang divisualisasikan dengan latar belakang bagan batang vertikal abu-abu muda [21].



Gambar 2.20 Contoh Visualisasi Data *Combination Chart* [21]

2.1.7 Analisis Klaster (*Cluster Analysis*)

Merupakan metode statistik yang digunakan untuk mengelompokkan item data, baik secara fisik maupun abstrak, ke dalam segmen-segmen homogen berdasarkan karakteristik tertentu. Item-item data tersebut akan dikelompokkan dalam segmen yang sama berdasarkan kesamaan karakteristik [12] [15]. Dalam proses klasterisasi, item atau objek data yang homogen akan dikelompokkan dalam klaster yang sama, sementara item atau objek data lain yang lebih heterogen akan dimasukkan ke dalam klaster yang berbeda [12] [15] [26]. Melalui pengklasteran, kita dapat menganalisis data dengan mengidentifikasi kesamaan atau kemiripan (*Related*) antar item data. Setiap item data mewakili objek atau individu dari suatu *dataset*. Oleh karena itu, pengklasteran berfungsi untuk mengelompokkan item data

berdasarkan kesamaan atributnya, dengan tujuan untuk menemukan pola, tren, hubungan, informasi dan pengetahuan baru (*New Knowledge*) dari sekumpulan data [11] [12] [26].

Pengklasteran adalah alat penting dan populer dalam melakukan teknik penggalian data (*Data Mining*) untuk segmentasi pelanggan. Secara umum, pengklasteran dapat dibagi menjadi 2 (dua) jenis berdasarkan metode pengelompokkannya, yaitu metode pengelompokkan hierarki (*Hierarchical Clustering*) dan metode partisi (*Partitioning Methods*) [11] [19]. Metode pengelompokkan hierarki, adalah teknik membentuk klaster menjadi suatu struktur hierarki atau pohon klaster (*Dendogram*) menggunakan pendekatan algoritma aglomeratif (*Agglomerative*), yang melibatkan penggabungan (*Merging*) atau pemisahan (*Splitting*) kelompok data. Keuntungan dari pengelompokkan hierarki adalah tidak perlu menentukan jumlah klaster di awal [11] [12]. Sementara itu, metode pengelompokkan partisi (*Partitioning Methods*) adalah teknik yang membagi data ke dalam sejumlah klaster yang telah ditentukan sebelumnya. Metode ini lebih efisien dibandingkan dengan metode hierarki dalam menangani *dataset* yang besar [11]. Salah satu algoritma yang digunakan dalam metode pengelompokkan partisi adalah algoritma *K-Means (Non-Hierarchical)*. *K-Means* merupakan metode pengklasteran yang sangat populer karena algoritmanya yang sederhana serta kecepatan dalam menentukan pusat klaster (*Centroid*). Dalam prosesnya, Metode *K-Means* menggunakan metrik jarak *Euclidean* untuk mengukur kesamaan atau kemiripan antar item data (*Similarity*) dalam klaster secara iteratif [4] [12]. Berikut merupakan penjelasan analisis klaster, beserta 2 (dua) komponen utamanya, yaitu *K-Means* dan Indeks Validitas Klaster.

1. **Algoritma *K-Means* (*K-Means Algorithm*)**

   Merupakan salah satu metode pengklasteran (*Clustering*) dalam pembelajaran mesin (*Machine Learning*) yang menggunakan algoritma *Unsupervised Learning* untuk mengelompokkan item atau objek data berdasarkan metrik jarak terpendek antar titik data [11] [26]. Algoritma *K-Means* sangat populer dan masuk 10 (sepuluh) besar untuk metode klaster yang paling banyak digunakan dalam penggalian data (*Data Mining*) untuk penemuan wawasan dan pengetahuan baru (*Insight*) [11]. *K-Means* menggunakan pendekatan sederhana untuk mengelompokkan item atau objek data yang diobservasi ke dalam sejumlah *k* klaster yang berbeda, dengan menetapkan observasi klaster pada suatu pusat klaster (*Centroid*) terdekat. Suatu metrik jarak perlu dipilih untuk mengukur jarak antar item data setiap observasi dan *Centroid* klaster. Meskipun saat ini terdapat banyak metrik jarak, seperti *Manhattan, Jaccard, Minkowski, Cosine*. Namun metrik jarak yang paling umum digunakan adalah metrik jarak *Euclidean*. Pengukuran metrik

jarak ini, mengukur antara 2 (dua) titik item data dengan jarak garis lurus berdasarkan koordinat pengamatan menggunakan rumus teorema Pythagoras (*Pythagorean theorem*) yaitu [21]:

$$a^2 + b^2 = c^2$$

Rumus jarak *Euclidean* untuk menghitung jarak antar item atau objek data terhadap pusat klaster (*Centroid*) [26] adalah sebagai berikut (7):

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2} \quad \ldots\ldots (7)$$

Keterangan:

d (x, y) = Jarak antar kedua titik item atau objek data (d = *Distance*)

$X = (x_1, x_2, \ldots x_n)$ adalah koordinat titik item atau objek data pertama

$Y = (y_1, y_2, \ldots y_n)$ adalah koordinat titik item atau objek data kedua

Langkah-langkah pengklasteran data menggunakan metode *K-Means,* seperti yang diilustrasikan pada Gambar 2.21, adalah sebagai berikut:

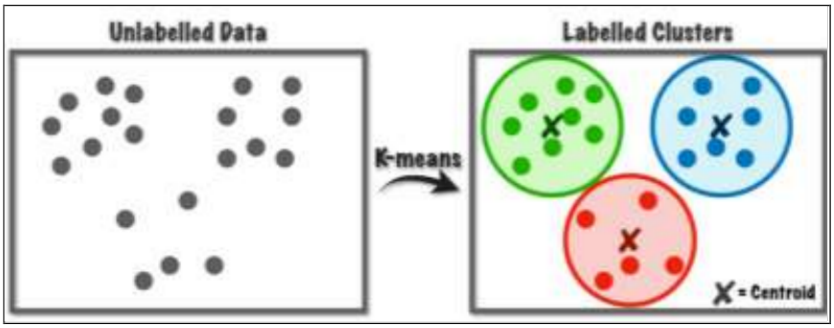
A. Menentukan jumlah *k* klaster [4] [21] [26], kemudian inisialisasi atau pilih suatu item data untuk dijadikan nilai k sebagai pusat klaster (*Centroid*) sementara secara acak sebagai observasi awal [4] [21] [26].

B. Kelompokkan setiap item data ke dalam klaster terdekat terhadap pusat klaster (*Centroid*). Kedekatan antar 2 (dua) item atau objek data dihitung menggunakan jarak *Euclidean* [4] [26].

C. Hitung ulang setiap pusat klaster (*Centroid*) untuk mendapatkan nilai *Centroid* berikutnya, dengan menghitung rata-rata (*Average*) semua data *Centroid* berdasarkan data yang telah diperoleh [4] [21] [26].

D. Lakukan klaster ulang setiap item atau objek data menggunakan semua data *Centroid* baru sampai *Centroid* tidak berubah lagi [4] [26].

E. Apabila *Centroid* tidak ada perubahan lagi maka proses pengklasteran dianggap selesai, selanjutnya tetapkan setiap item atau objek data observasi ke klaster dengan *Centroid* terdekat [4] [21] [26].



Gambar 2.21 Pengklasteran *K-Means* (*K-Means Clustering*)

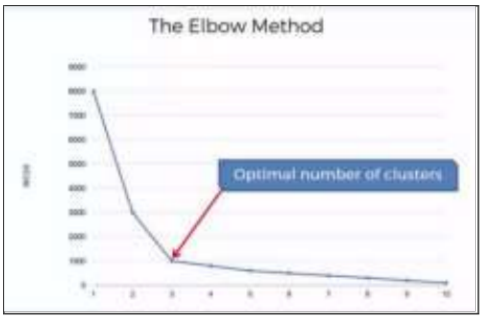
Sumber: Google.com

Algoritma *K-Means* memiliki beberapa kelemahan, terutama terkait dengan kompleksitasnya dalam mendeteksi klaster "alami" ("*natural*"), yaitu klaster yang memiliki ukuran, ketebalan, atau bentuk yang bervariasi dan tidak berbentuk oval [12]. Selain itu, tantangan utama dalam metode *K-Means* adalah menentukan jumlah klaster *k* yang optimal. Beberapa literatur menunjukkan bahwa akurasi metode *K-Means* dapat meningkat, jika nilai awal pusat klaster (*Centroid*) dan jumlah *k* klaster ditentukan dengan tepat. Terdapat berbagai cara untuk memperkirakan dan mengukur jumlah klaster *k* yang optimal, yaitu Metode Siku (*Elbow*), Indeks *Silhouette*, Indeks *Calinski Harabasz*, Indeks *Davies-Bouldin*, Indeks *Ratkowski*, Indeks *Hubert*, Indeks *Ball-Hall*, dan Indeks *Krzanowski Lai* [4] [26].

2. **Indeks Validitas Klaster (*Cluster Validity Index*)**

   Digunakan untuk menilai dan mengevaluasi kualitas klaster yang terbentuk dalam proses pengklasteran [2] [4] [26]. Terdapat 2 (dua) metode yang umum digunakan untuk menentukan jumlah klaster optimal, yaitu Metode *Elbow* dan Indeks *Silhouette*.

   - **Metode Siku (*Elbow Method*)**, merupakan metode untuk menentukan jumlah klaster optimal, dengan melihat persentase perbandingan antara jumlah klaster yang akan membentuk sudut pada kurva atau perubahan bentuk kurva. Jika nilai klaster pertama dengan nilai klaster kedua membentuk sudut siku (*Elbow*) pada kurva, dan terjadi penurunan nilai paling besar, maka jumlah klaster tersebut dianggap yang terbaik (optimal) seperti yang diilustrasikan pada Gambar 2.22. Metode ini bersifat visual dengan mengukur variasi intra-klaster (*Intra-Cluster*) atau yang dikenal sebagai total *Within-Clusters Sum of Squares* (WCSS) sebagai fungsi dari jumlah klaster optimal. Jumlah klaster *k* paling optimal, ditunjukkan dari nilai WCSS yang semakin kecil, atau sebaliknya [2] [4].



Gambar 2.22 Metode Siku (*Elbow Method*)

Sumber: Google.com

Berikut rumus WCSS adalah sebagai berikut (8):

$$WCSS = \sum_{j=1}^{k} \sum_{i=1}^{n} || x_i{}^{(j)} - c_j ||^2 \quad \text{..... (8)}$$

Keterangan:

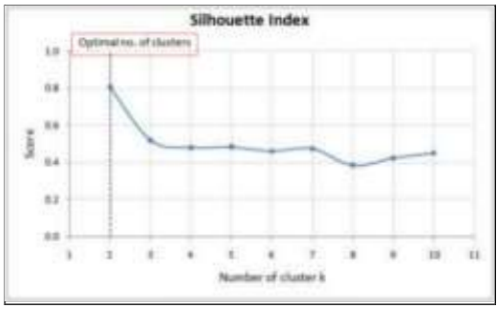
k = Jumlah klaster (sebanyak j)

n = Jumlah item atau objek data (sebanyak i)

x_i = Nilai item atau objek data (Elemen) ke i dalam klaster (j)

c_i = Titik pusat klaster (*Centroid*) ke j

- **Indeks *Silhouette* (*Silhouette Index/SI)***, adalah metode untuk mengevaluasi kualitas klaster dari jumlah klaster yang terbentuk dengan mengukur seberapa baik dan tepat untuk setiap item atau objek data tersebut sudah berada pada klaster yang diberikan, dibandingkan dengan klaster lain [26]. Pengukuran indeks *Silhouette* menggunakan skor, dimana kualitas klaster berkisar antara -1,0 dan 1. Jika skor menunjukkan nilai positif maka pengklasteran memiliki kohesi dan pemisahan yang baik, sebaliknya skor menunjukkan nilai negatif maka pengklasteran kurang tepat. Berdasarkan literatur, ambang batas (*Threshold*) skor *Silhouette* yang diterapkan adalah 0.5 sampai 0.7 sebagai kisaran skor pengelompokkan yang wajar/normal [2] [4] [26].



Gambar 2.23 Indeks Silhouette (*Silhouette Index*)

Sumber: Google.com

Skor *Silhouette* dapat dihitung dengan persamaan sebagai berikut (9):

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad \text{..... (9)}$$

Keterangan:

a(i) = Rata-rata jarak antar titik data *i* dalam klaster yang sama (*Intra-Cluster Distance*)

b(i) = Rata-rata jarak klaster berbeda terdekat untuk titik data *i* (*Nearest-Cluster Distance*)

Silhouette mengacu pada metode interpretasi dan validasi konsistensi dalam klaster data, sehingga nilai skor koefisien *Silhouette* yang lebih besar akan menunjukkan klaster yang lebih baik. Berikut intepretasi dari nilai skor *Silhouette* [4] [26]

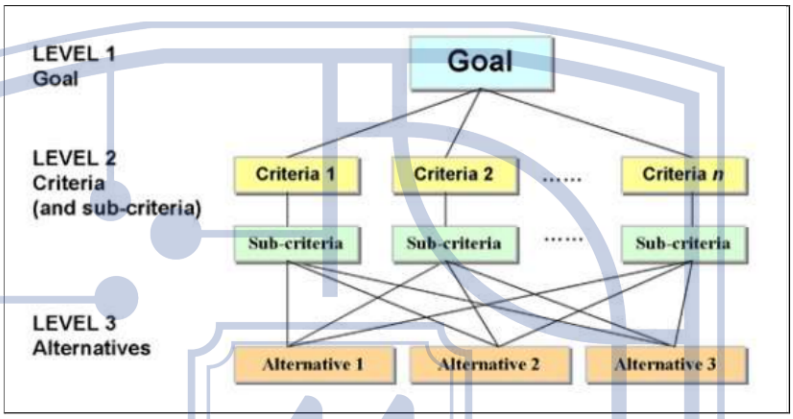
- ***S(i)* mendekati 1 (Positif)**, menunjukkan titik (item atau objek) data tersebut sangat tepat dengan klasternya, dan cukup jauh dari klaster lain.
- ***S(i)* mendekati 0 (Normal)**, menunjukkan titik (item atau objek) data berada di antara 2 (dua) klaster, dan tidak jelas klaster mana yang paling cocok.
- ***S(i)* mendekati -1 (Negatif)**, menunjukkan titik (item atau objek) data lebih cocok dengan klaster lain daripada klaster yang sekarang, yang menunjukkan bahwa klaster sekarang tersebut tidak sesuai.

2.1.8 Proses Hierarki Analitik (*Analytical Hierarchy Process/AHP*)

Proses hierarki analitik (AHP) adalah salah satu metode yang berfungsi menentukan pemilihan, keputusan atau prioritas dari beberapa kriteria. AHP juga dapat digunakan sebagai penimbang nilai kepentingan dari variabel model RFM. AHP dikembangkan oleh Thomas L. Saaty dari Universitas Pittsburgh pada tahun 1970-an dan sejak itu telah digunakan secara luas dalam pengambilan keputusan dengan skor tertimbang untuk berbagai masalah rumit, yang terdapat banyak kriteria atau variabel yang dipertimbangkan untuk penetapan suatu perioritas [15] [22].

AHP merupakan teknik terstruktur untuk mengatur dan menganalisis keputusan kompleks berdasarkan aspek matematik dan psikologi. Proses ini memanfaatkan hierarki dalam melakukan evaluasi secara menyeluruh dan memilih dari salah satu solusi alternatif

untuk masalah tertentu [16]. Teknik ini mengabungkan metode kualitatif dan kuantitatif dalam mempresentasikan permasalahan yang kompleks secara teratur dengan struktur hierarki [22], menyediakan pendekatan matematis yang objektif untuk memproses preferensi subjektif dan pribadi yang berasal dari seorang individu atau kelompok dalam pengambilan keputusan. Teknik AHP dirancang untuk membantu dan mendukung proses pengambilan keputusan multi-kriteria (*Multi-Criteria Decision Making*/MCDM) [16].



Gambar 2.24 Proses Hierarki Analitik (*Analytical Hierarchy Process/AHP*)

Sumber: Google.com

AHP didasarkan pada pengalaman subjektif dan pengetahuan para pengambil keputusan untuk menentukan prioritas kriteria. Para pengambil keputusan dapat menilai beberapa kriteria yang berbeda melalui perbandingan pasangan untuk menentukan solusi terbaik. Perbandingan ini dilakukan dengan menggunakan skala kepentingan relatif yang disarankan oleh Saaty (2008) [15]. Untuk menentukan kepentingan relatif suatu kriteria, pengambil keputusan atau pengguna mengaitkan nilai yang bervariasi dari 1 hingga 9 antar kriteria [15] [22]. Proses ini memastikan bahwa setiap kriteria dianalisis secara mendalam, sehingga menghasilkan keputusan yang lebih akurat dan relevan dalam konteks yang kompleks. Selain itu, AHP memungkinkan kolaborasi antar anggota tim, sehingga perspektif yang berbeda dapat diintegrasikan, meningkatkan kualitas keputusan yang diambil secara keseluruhan [15].

Tabel 2.2 AHP Penilaian Skala Kepentingan [15]

Skala	Keterangan
1	Sama Penting (*Equally Important Preferred*)
2	Sama sampai Cukup Penting (*Equally to Moderately Important Preferred*)

3	Cukup Penting (*Moderately Important Preferred*)
4	Cukup Penting sampai Penting (*Moderately to Strongly Important Preferred*)
5	Penting (*Strongly Important Preferred*)
6	Penting sampai Sangat Penting (*Strongly to Very Strongly Important Preferred*)
7	Sangat Penting (*Very Strongly Important Preferred*)
8	Sangat Penting sampai Paling Penting (*Very Strongly to Extremely Important Preferred*)
9	Paling Penting (*Extremely Important Preferred*)

AHP terdiri dari 3 (tiga) komponen penting yaitu:

1. Dekomposisi (*Decomposition*), yaitu menyusun keputusan ke dalam hierarki yang terdiri dari tujuan atau sasaran (*Objective/Goal*) dan turunan seperti kriteria (*Criteria*), sub kriteria (*Sub-Criteria*) serta solusi pilihan alternatif (*Alternative*) [16].
2. Evaluasi (*Evaluation*), dengan melakukan perbandingan penilaian berpasangan antar kriteria mencakup indikator atau variabel dan memberikan nilai 1 hingga 9 untuk setiap perbandingan oleh para pengambil keputusan (*Evaluator*). Para *Evaluator* dapat membuat penilaian yang tidak konsisten antar elemen dalam tabel matriks penilaian, sehingga sebelum mendapatkan nilai bobot (*weighted*) dihitung berdasarkan penilaian berpasangan, tingkat ketidakkonsistenan diukur dengan indeks ketidakkonsistenan (*Consistent Index/CI*) yang harus kurang dari 0,1. Jika tidak, penilaian berpasangan perlu direvisi [15] [16].
3. Sintesis (*Synthesis*), setelah mendapatkan nilai prioritas atau bobot untuk setiap kriteria, nilai prioritas lokal (*Local Optimum*) dapat ditetapkan untuk setiap elemen (misalnya, *Recency*: 0.20, *Frequency*: 0.30, *Monetary*: 0.50). Selanjutnya, proses sintesis dilakukan untuk mempropagasi keseluruhan prioritas atau solusi alternatif yang tersedia ke tingkat prioritas global (*Global Optimum*). Hal ini melibatkan penggabungan prioritas lokal untuk menghasilkan suatu nilai prioritas global yang mencerminkan kontribusi relatif setiap elemen dalam pengambilan keputusan. Sintesis memastikan bahwa keputusan yang diambil mencerminkan keseluruhan nilai skala kepentingan dan prioritas secara keseluruhan hierarki [16].

Tabel 2.3 AHP Nilai Skala Kepentingan dengan Kriteria Model RFM [15]

Kriteria	**Nilai Skala Kepentingan**																	**Kriteria**
Recency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Frequency

Frequency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Monetary
Monetary	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Recency

2.2 Tinjauan Objek Penelitian

Penelitian ini akan dilaksanakan pada PT Karya Logistik, yang merupakan bagian dari Cahaya Matahari Group yang didirikan pada tahun 1984, adalah perusahaan yang berfokus pada impor dan distribusi mesin-mesin penggerak berkualitas tinggi. Menyediakan berbagai jenis mesin, mulai dari mesin *diesel*, mesin bensin, hingga *gearbox* kapal laut, serta berbagai peralatan yang dibutuhkan oleh petani dan nelayan. Selain itu, perusahaan ini juga mendistribusikan mesin untuk sektor perikanan, pertanian dan perkebunan, termasuk produk-produk dari merek terkenal seperti ASAHI, TIANLI, dan MATARI dengan komitmen untuk memberikan produk berkualitas dan layanan terbaik termasuk menyediakan berbagai kebutuhan suku cadang mesin.

Berkantor pusat di Jakarta, dengan lokasi alamat di Jl. Sukarjo Wiryopranoto No.5, RT.11/RW.3, Maphar, Kec. Taman Sari, Kota Jakarta Barat, DKI Jakarta KodePos 11160 dan memiliki beberapa cabang perusahaan yang tersebar di beberapa Pulau Indonesia khususnya di Pulau Sumatera (Kota Medan), Pulau Jawa (Kota Jakarta dan Surabaya) dan Pulau Sulawesi (Kota Makassar).

2.2.1 Visi dan Misi Perusahaan

Visi dari PT Karya Logistik adalah "Menjadi pemimpin pasar mesin penggerak dalam distribusi, penjualan dan pelayanan purna jual secara berkesinambungan". Terdapat misi dari PT Karya Logistik adalah sebagai berikut:

1. Menyediakan produk mesin penggerak berkualitas tinggi yang memenuhi kebutuhan pelanggan di berbagai sektor, termasuk industri perikanan, pertanian, dan perkebunan.
2. Memberikan layanan distribusi yang efisien dan tepat waktu serta memastikan ketersediaan produk di seluruh cabang.
3. Meningkatkan kepuasan pelanggan melalui penjualan dan pelayanan purna jual yang profesional.

Gambar 2.25 Logo Perusahaan PT Karya Logistik

Visi dan misi, sejalan dengan makna yang terkandung dalam logo perusahaan, yaitu:

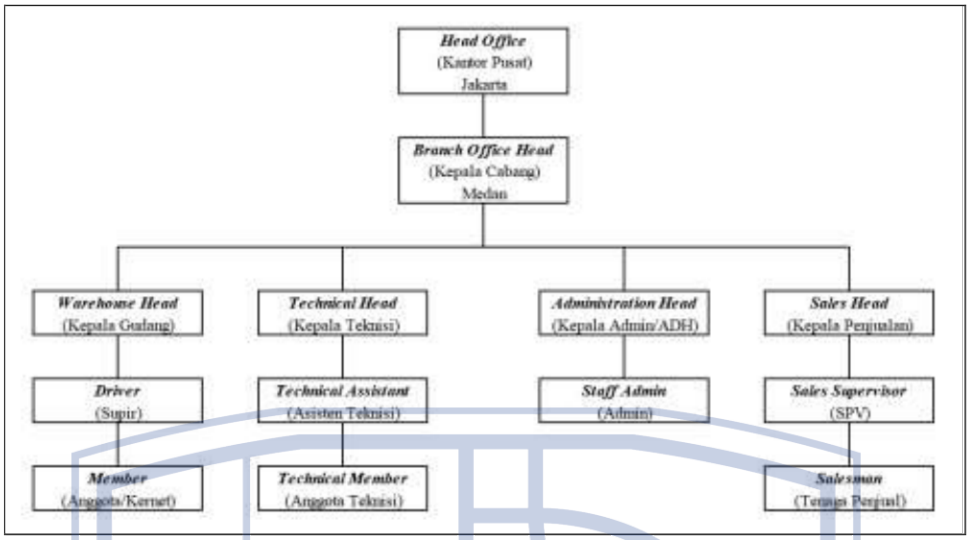
1. Huruf K, berasal dari huruf K (Ka pa) Yunani, dengan mengambil filosofi tangan terbuka yang artinya perusahaan senantiasa terbuka, transparan dan terpercaya mengenai produk yang dijual dan mengutamakan kepuasan pelanggan.
2. Lingkaran, melambangkan keberlanjutan dan kesinambungan, seperti Matahari yang memancarkan sinergi yang terus menerus tanpa putus. Dimana dapat diartikan sebagai pertahanan solid, integritas dan kesempurnaan.

2.2.2 Fokus Perusahaan

Dalam mendukung komitmen perusahaan untuk menjaga kesinambungan bisnis serta memberikan layanan terbaik kepada pelanggan, perusahaan perlu melakukan analisis segmentasi pelanggan untuk memahami lebih mendalam kebutuhan setiap segmen pelanggan. Penelitian ini bertujuan untuk memberikan solusi yang tepat dalam menjaga kepuasan dan loyalitas pelanggan, dengan meningkatkan manajemen persediaan yang efisien. Fokus penelitian ini dilaksanakan pada PT Karya Logistik Cabang Sumatera, dengan lokasi di Kawasan Industri Medan (KIM) 3 (tiga) Jl. Pulau Nias Selatan IV, Tangkahan, Kec. Medan Labuhan, Kota Medan, Sumatera Utara 20242 (Lokasi Kantor https://maps.app.goo.gl/9Gw9e31sjEaS6cB69). Melalui penelitian ini, perusahaan diharapkan dapat memperkuat posisi di pasar serta mempertahankan dan mengembangkan bisnis secara berkelanjutan.

2.2.3 Struktur Organisasi

Struktur organisasi pada PT Karya Logistik Cabang Sumatera dapat dilihat pada gambar berikut ini:



Gambar 2.26 Struktur Organisasi PT Karya Logistik Cabang Sumatera (Medan)

Dari setiap bagian dalam struktur organisasi pada gambar 2.26 memiliki tugas dan fungsi masing-masing, pada bagian Gudang (*Warehouse*) memiliki tanggung jawab pada distribusi dan persediaan barang, bagian Teknisi (*Technical*) memiliki tanggung jawab pada layanan purna jual seperti pergantian, perbaikan, dan perawatan mesin, bagian Administrasi (*Administration*) memiliki tanggung jawab pada kegiatan tata usaha operasional sehari-hari, dan bagian Penjualan (*Sales*) memiliki tanggung jawab terhadap penjualan dan layanan pelanggan.

2.3 Penelitian Terdahulu

Berdasarkan kajian literatur dari beberapa penelitian terdahulu terkait segmentasi pelanggan dan pengukuran nilai pelanggan (*Customer Lifetime Value/CLV*), sebagian besar studi telah menggunakan model RFM (*Recency*, *Frequency*, *Monetary*) dengan berbagai variasinya untuk memahami perilaku, kebiasaan serta karakteristik pelanggan.

Penelitian oleh Aslantas [2] menerapkan model RFM dengan algoritma *K-Means* untuk melakukan segmentasi pelanggan. Meskipun pendekatan ini efektif secara teknis, penggunaannya masih terbatas pada variabel tradisional RFM, tanpa mempertimbangkan variabel eksternal lain yang dapat mempengaruhi segmentasi pelanggan. Selanjutnya, penelitian dengan model RFM-D (*Demography*) [11], telah mengintegrasikan variabel demografis seperti usia, wilayah geografis atau regional, dan jenis kelamin. Namun cakupan variabel ini masih bersifat dasar dan statis, sehingga kurang menggambarkan kompleksitas perubahan perilaku dan kebutuhan pembelian pelanggan. Sementara itu, model RFM-V (*Variety*) [13] menekankan pada keberagaman produk yang dibeli pelanggan melalui

pendekatan analisis keranjang belanja. Meskipun model ini menambah dimensi preferensi produk, namun belum sepenuhnya menangkap dinamika perubahan kebutuhan pelanggan yang dinamis.

Sedangkan penelitian dengan model RFM-AR (*Age & Return*) [19] telah memberikan wawasan tambahan dalam segmentasi pelanggan, namun hanya terbatas aktivitas pengembalian barang dan lama hubungan pelanggan terkini pelanggan (retensi dan potensi *churn*) dan model RFM-DP (*Discount Proportion*) [26] hanya fokus pada sensitivitas harga pembelian pelanggan, belum mempertimbangkan perilaku pembelian aktual dan loyalitas melalui pengukuran nilai pelanggan (CLV). Adapun model RFM-T (*Time*), meskipun berhasil memperhitungkan variabel waktu pembelian secara detail (intensitas), tetapi belum mampu mengidentifikasi perubahan permintaan (*Demand*) pelanggan yang bersifat dinamis seperti fluktuasi preferensi produk, penyesuaian tren pasar, urgensi kebutuhan pelanggan [30]. Beberapa penelitian yang menggunakan metode pembobotan *Analytic Hierarchy Process* (AHP) dengan variabel RFM [15] dan LRFM [16] telah memberikan kontribusi dalam menyediakan pendekatan evaluasi yang lebih sistematis terhadap nilai pelanggan (CLV). Namun, pendekatan ini memiliki keterbatasan karena hanya berfokus nilai pelanggan bukan segmentasi pelanggan. Berikut adalah tabel penelitian terdahulu, yang menggunakan model RFM dan beberapa variasinya.

Tabel 2.4 Penelitian Terdahulu (Periode 2021 - 2024)

Study	***Writers & Year***	***Methods***	***Result***
Customer Segmentation Using K-Means Clustering Algorithm and RFM Model	Gozde Aslantas, Mustafacan Gencgul, Merve Rumelli, Mustafa Ozsarac, Gozde Bakirli (2022)	➢ Ekstraksi fitur RFM dari data transaksi ➢ Normalisasi data ➢ *Clustering* menggunakan algoritma *K-Means* (*Elbow Method*) ➢ Evaluasi dengan *Silhouette Score* ➢ Perhitungan nilai pelanggan (CLV)	Terbentuk 4 (empat) segmen pelanggan yang optimal, yaitu: *Best, Good, Average,* dan *Worst Customers*. Dan nilai pelanggan (CLV) untuk masing-masing segmen

An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry	Thanh Ho, Suong Nguyen, Huong Nguyen, Ngoc Nguyen, Dac-Sang Man, ThaoGiang Le (2023)	➢ Pengembangan model RFM-D (*Demographic - Age, Region, Gender*) ➢ Penerapan algoritma *K-Means* (*Elbow Method*) dan *K-Prototypes* ➢ Evaluasi dengan *Adjusted Rand Index* (ARI) dan *Adjusted Mutual Information* (AMI) ➢ Pengkodean variabel kategorikal dengan *One-Hot Encoding* ➢ Analisis *Cohort* untuk tingkat retensi pelanggan	Terbentuk 5 (lima) segmen pelanggan yang optimal dengan karakteristik berbeda, yaitu *Loyal* (0), *New* (1), *Need Attention* (2), *At-Risk* (3), *Can't Lose Them* (4) *Customers.* Model menunjukkan hasil yang stabil dengan 2 (dua) algoritma *Clustering*.
A Customer Segmentation Model Proposal for Retailers: RFM-V	Pinar Ozkan, Ipek Deveci Kocakoc (2021)	➢ Pengembangan model RFM-V (*Variety*) untuk mengukur kedalaman hubungan pelanggan ➢ Penyusunan *Customer-Product Depth Matrix* fokus parameter M (*Monetary*) dan V (*Variety*) ➢ *Clustering* menggunakan algoritma *K-Means* (*Elbow Method*) ➢ Analisis segmentasi pelanggan berdasarkan kedalaman hubungan serta integrasi dengan analisis keranjang belanja.	Terbentuk 4 (empat) kuadran pelanggan berdasarkan kedalaman hubungan: *Lo-Spender Broad* (*Frequent*), *Lo-Spender Narrow* (*Uncertain*), *Hi-Spender Narrow* (*Spender*), *Hi-Spender Broad* (*Best*)

Customer Value Analysis Using Weighted RFM model: Empirical Case Study	Tarek BELHADJ (2021)	➢ Penggunaan model RFM berbobot untuk menghitung nilai pelanggan (CLV) ➢ Penentuan nilai bobot relatif dari variabel RFM menggunakan metode *Analytic Hierarchy Process* (AHP) ➢ Normalisasi data ➢ Segmentasi pelanggan menggunakan algoritma *clustering K-Means*	Teridentifikasi segmen pelanggan dengan nilai pelanggan (CLV) tertinggi. Fokus strategi perusahaan pada kelompok pelanggan (segmen) yang paling berharga (*Top, Big, Medium, Small, Inactive*)
Measuring Customer Lifetime Value: Application of Analytic Hierarchy Process in Determining Relative Weights of LRFM	Saurabh Pradhan, Gokulananda Patel, Pankaj Priya (2021)	➢ Penambahan variabel *Length of Relationship* (L) pada model RFM menjadi LRFM ➢ Penentuan nilai bobot relatif dari variabel L, R, F, dan M menggunakan metode AHP	Model LRFM berbobot memberikan hasil nilai pelanggan (CLV) yang lebih akurat dan dapat diandalkan dibandingkan dengan pendekatan RFM tradisional
RFM-AR Model for Customer Segmentation using K-Means Algorithm	Ali Khumaidi, Herry Wahyono, Risanto Darmawan, Harry Dwiyana Kartika, Nuke L Chusna, Muhammad Kaisar Fauzy (2023)	➢ Pengembangan Model RFM-AR yang menggabungkan variabel *Recency, Frequency, Monetary*, *Age*, dan *Return*. ➢ Penerapan algoritma *K-Means* untuk klasterisasi pelanggan.	Terbentuk 3 (tiga) segmen pelanggan A (*High*), B (*Middle*) dan C (*Low*). Namun penentuan jumlah klaster (segmen) optimal menggunakan metode Elbow

		➢ Penggunaan *Elbow Method* untuk menentukan jumlah klaster (segmen) optimal ➢ Implementasi proses CRISP-DM (*Cross-Industry Standard Process for Data Mining*)	menghasilkan nilai klaster optimal (K) adalah 2 untuk setiap variabel RFM-AR
Enhancing Customer Segmentation Insights by using RFM + Discount Proportion Model with Clustering Algorithms	Victor Hugo Antonius, Devi Fitrianah (2024)	➢ Model RFM + DP yang menggabungkan variabel *Recency, Frequency, Monetary*, dan *Discount Proportion* ➢ Penerapan algoritma *Clustering*: *K-Means, K-Medoids, Fuzzy C-Means*, dan *Mini-Batch K-Means* ➢ Penentuan jumlah klaster (segmen) optimal menggunakan metode *Elbow*. ➢ Evaluasi performa algoritma menggunakan *Silhouette Score* dan *Calinski-Harabasz Index*	Terbentuk 4 (empat) segmen pelanggan, yaitu *Platinum*, *Gold*, *Silver*, dan *Bronze* berdasarkan model RFM + DP. *Mini-Batch K-Means* menunjukkan *Silhouette Score* tertinggi sebesar 0,50, sementara *K-Means* memiliki nilai *Calinski-Harabasz* (CH) *Index* tertinggi sebesar 1056
Customer Analysis Using Machine Learning-Based Classification Algorithms for Effective Segmentation	Asmat Ullah, Muhammad Ismail Mohmand, Hameed Hussain, Sumaira Johar, Inayat Khan, Shafiq Ahmad,	➢ Penerapan model RFMT (*Time*) dengan algoritma *Agglomerative*, *K-Means*, *Gaussian*, dan DBSCAN. ➢ Evaluasi klaster menggunakan metode *Elbow*, *Dendrogram*,	Terbentuk 3 (tiga) segmen pelanggan yang berbeda berdasarkan analisis RFMT, yaitu C0, C1, dan C2.

Using Recency, Frequency, Monetary, and Time	Haitham A. Mahmoud, Shamsul Huda (2023)	*Silhouette*, *Calinski–Harabasz*, *Davies–Bouldin*, dan *Dunn Index* ➢ Pemilihan klaster stabil menggunakan teknik *majority voting* (*mode version*)	Segmentasi dilakukan berdasarkan kategori produk, tahun, tahun fiskal, bulan, status transaksi, dan musim

Meskipun berbagai penelitian sebelumnya telah mengembangkan model RFM (*Recency*, *Frequency*, *Monetary*) dengan penambahan sejumlah variabel seperti demografis (*RFM-Demography*), keberagaman pembelian produk (*RFM-Variety*), proporsi diskon (*RFM-Discount Proportion*), aspek waktu (*RFM-Time*), serta lama hubungan pelanggan (*Length-RFM*), namun hingga saat ini belum terdapat penelitian yang secara eksplisit mengintegrasikan variabel permintaan (*Demand*) ke dalam model RFM. Padahal, variabel permintaan (*Demand*) memiliki peran penting dalam memahami perilaku dan kebiasaan pembelian serta kecenderungan kebutuhan pelanggan yang dinamis. Dengan adanya variabel permintaan (*Demand*), perusahaan dapat lebih akurat menangkap perubahan preferensi pelanggan, preferensi produk (pergeseran minat atau kebutuhan terhadap produk tertentu), dan kebutuhan pasar yang fluktuatif. Hal ini menjadi penting dalam menyusun startegi pemasaran dan manajemen persediaan yang sesuai dengan kondisi serta permintaan pasar. Selain itu, pemahaman terhadap permintaan (*Demand*) turut berkontribusi dalam pengukuran nilai pelanggan (CLV), khususnya dalam konteks loyalitas dan retensi pelanggan. Dengan demikian, perusahaan dapat meningkatkan daya saing di pasar yang semakin kompetitif sekaligus mempertahankan pelanggan.

Tabel 2.5 Perbandingan Model RFM

Model	**Variabel Tambahan**	**Kelebihan**	**Keterbatasan**
RFM (*Standard*)	-	Dasar segmentasi pelanggan, sederhana dan mudah diimplementasikan	Tidak terdapat variabel tambahan berupa faktor eksternal atau preferensi lain

			yang dapat mempengaruhi segmentasi pelanggan
RFM-D (*Demography*)	Usia, Wilayah Geografis atau Regional, dan Jenis Kelamin	Tersedia informasi profil pelanggan dan karakteristik demografis pelanggan secara umum	Variabel demografis bersifat statis dan kurang mencerminkan perubahan perilaku dan kebutuhan pembelian pelanggan
RFM-V (*Variety*)	Keberagaman Produk	Mengukur preferensi dan minta produk, cocok untuk analisis keranjang belanja	Tidak fokus pada loyalitas atau perubahan permintaan pelanggan
LRFM (*Length*)	Lama Hubungan Pelanggan	Memberikan gambaran loyalitas berdasarkan lamanya hubungan	Tidak mempertimbangkan perubahan preferensi produk selama periode lama hubungan pelanggan
RFM-AR (*Age* & *Return*)	*Age* (Lama menjadi Pelanggan) & *Return* (Pengembalian Barang)	Menilai loyalitas dan kepuasan pelanggan berdasarkan durasi dan aktivitas pengembalian barang	Terlalu fokus pada hubungan dan aktivitas pengembalian barang, tidak memperhitungkan dimensi kebutuhan pelanggan yang dinamis
RFM-DP (*Discount Proportion*)	Proporsi diskon yang digunakan pelanggan	Memahami sensitivitas harga pelanggan dan membantu startegi promosi	Kurang mempertimbangkan perilaku pembelian yang lebih dalam (aktual) dan loyalitas jangka panjang karena fokus pada harga (diskon)
RFM-T (*Time*)	Intensitas Waktu antar Pembelian (Durasi, Interval)	Mendeteksi ritme pembelian dan intensitas pelanggan	Tidak menggambarkan fluktuasi atau perubahan preferensi produk dan permintaan pelanggan dari waktu ke waktu
RFM + AHP	Pembobotan atau Nilai	Memberikan pendekatan evaluasi nilai pelanggan	Hanya fokus pada pembobotan RFM, tidak

	Prioritas *Recency*, *Frequency*, *Monetary*	(CLV) yang lebih sistematis dan terukur	mempertimbangkan variabel eksternal lainnya
RFM + Permintaan (*Demand*) + AHP	**Klasifikasi Permintaan (*Smooth*, *Erratic*, *Intermittent*, *Lumpy*)**	**Mencerminkan kebutuhan pelanggan dan pasar yang dinamis, mendukung manajemen persediaan, segmentasi pelanggan dan nilai pelanggan (CLV)**	**Pendekatan holistik dengan solusi yang lebih komprehensif dan adaptif**

Oleh karena itu, pengembangan model RFM + Permintaan (*Demand*) + AHP diusulkan sebagai solusi yang lebih komprehensif dan adaptif dalam melakukan segmentasi pelanggan serta pengukuran nilai pelanggan (*Customer Lifetime Value/CLV*). Integrasi variabel permintaan (*Demand*), khususnya melalui pendekatan klasifikasi permintaan (*Demand Classification*) seperti *Smooth*, *Erratic*, *Intermittent*, dan *Lumpy*, memungkinkan perusahaan untuk tidak hanya memahami nilai pelanggan berdasarkan histori transaksi pembelian, tetapi juga pola permintaan (*Demand Pattern*) produk aktual yang bersifat dinamis dan fluktuatif. Dengan mempertimbangkan permintaan (*Demand*) pelanggan, perusahaan dapat mengidentifikasi kebutuhan penyediaan produk yang konsisten, dan membantu perencanaan persediaan yang lebih fleksibel dan responsif terhadap variasi permintaan pasar.

Model RFM + Permintaan (*Demand*) + AHP tidak hanya mendukung segmentasi pelanggan dan memberikan pengukuran nilai pelanggan (CLV) yang lebih relevan (melalui metode AHP), tetapi juga berkontribusi langsung terhadap efisiensi operasional perusahaan, khususnya dalam hal perencanaan permintaan (*Market*), pengendalian persediaan (*Stock-Level*), serta mempertahankan loyalitas pelanggan dalam jangka panjang bisnis.

2.4 Kerangka Berpikir

Penelitian ini bertujuan untuk mempertahankan loyalitas pelanggan dan mengelola persediaan melalui segmentasi pelanggan, yaitu melakukan analisis RFM (*Recency, Frequency, Monetary*), perhitungan nilai pelanggan (*Customer Lifetime Value/CLV*), dan pembobotan Proses Hierarki Analitik (*Analytical Hierarchy Process/AHP*) dengan mempertimbangkan faktor klasifikasi permintaan (*Demand Classification*). Tahapan

penelitian dimulai dengan pengumpulan *dataset* transaksi, pelanggan dan produk selama periode 2022-2024, yang kemudian akan menjadi *dataset* primer. Selanjutnya, dilakukan pra-pemrosesan dan transformasi data yang mencakup pembersihan data, pemilihan data, kontruksi fitur, dan normalisasi data untuk memperoleh *dataset* sekunder yang relevan untuk analisis segmentasi RFM dan klasifikasi permintaan (*Smooth Erratic/SE*). Setelah itu, dilakukan analisis klaster untuk mendapatkan hasil klasterisasi RFMSE. Segmentasi pelanggan dilakukan dengan algoritma *K-Means,* diikuti validasi dengan melakukan evaluasi perhitungan optimal klaster melalui metode *Elbow* (WCSS) dan Indeks *Silhouette.* Pada tahap terakhir, dilakukan penentuan peringkat (*Ranking*) pelanggan berdasarkan perhitungan nilai pelanggan (CLV) yang dihitung melalui pembobotan (*Weighted*) menggunakan metode proses hierarki analitik (AHP). Pembobotan ini melibatkan variabel RFM + SE dan disertai dengan pengujian *Consistency Index* (*CI*) dan *Consistency Ratio* (*CR*). Untuk lebih jelasnya, dapat dilihat pada gambar 2.27 dibawah berikut.

Gambar 2.27 Kerangka Berpikir