

BAB II

KAJIAN LITERATUR

2.1 Penjualan dan Pembelian

Penjualan dan pembelian merupakan dua aktivitas utama dalam operasional perusahaan, khususnya pada perusahaan dagang maupun ritel. Kedua aktivitas ini saling berkaitan dan berperan penting dalam menjaga kelangsungan bisnis serta pencapaian keuntungan perusahaan. Pengelolaan yang baik terhadap kedua kegiatan ini akan membantu perusahaan dalam memenuhi kebutuhan pasar secara optimal [10].

Penjualan adalah proses pertukaran barang atau jasa antara penjual dan pembeli yang disertai dengan imbalan berupa uang atau nilai tertentu. Dalam konsep pemasaran, penjualan tidak hanya berfokus pada transaksi semata, tetapi juga pada upaya memenuhi kebutuhan dan keinginan konsumen. Aktivitas penjualan menjadi sumber utama pendapatan perusahaan, sehingga harus dikelola secara efektif. Penjualan dapat dilakukan secara tunai maupun kredit sesuai dengan kebijakan perusahaan [10].

Secara umum, proses penjualan dalam sistem dapat dijelaskan sebagai berikut [10]:

1. Identifikasi kebutuhan pelanggan.
Perusahaan mengidentifikasi kebutuhan pelanggan melalui analisis pasar atau interaksi langsung, sehingga produk yang ditawarkan sesuai dengan keinginan konsumen.
2. Penawaran produk.
Produk ditawarkan kepada pelanggan dengan informasi yang lengkap, seperti nama produk, harga, deskripsi, dan ketersediaan stok.
3. Pemesanan produk.
Pelanggan memilih produk dan menentukan jumlah (kuantitas) yang akan dibeli.
4. Pencatatan transaksi.
Sistem mencatat data transaksi, meliputi identitas pelanggan, alamat pengiriman, metode pembayaran, serta detail produk yang dibeli.
5. Proses pembayaran.
Pembayaran dilakukan baik secara tunai maupun non-tunai sesuai metode yang dipilih pelanggan.
6. Pengiriman barang.
Barang yang telah dibeli akan diproses dan dikirim ke alamat pelanggan.

7. Penyelesaian transaksi.

Transaksi dinyatakan selesai setelah barang diterima oleh pelanggan.

Pembelian merupakan kegiatan memperoleh barang atau jasa dari pihak lain untuk mendukung operasional perusahaan atau untuk dijual kembali. Pada perusahaan dagang, pembelian bertujuan untuk menjaga ketersediaan stok agar dapat memenuhi permintaan pelanggan. Proses pembelian meliputi identifikasi kebutuhan, pemilihan pemasok, pemesanan, hingga penerimaan barang. Pembelian yang baik harus mempertimbangkan kualitas, harga, dan ketepatan waktu pengiriman [10].

Proses pembelian dalam perusahaan umumnya meliputi beberapa tahapan berikut [10]:

1. Identifikasi kebutuhan barang.
Perusahaan menentukan barang yang perlu dibeli berdasarkan data stok dan tingkat penjualan.
2. Pemilihan pemasok (*supplier*).
Pemilihan pemasok dilakukan dengan mempertimbangkan harga, kualitas barang, dan ketepatan waktu pengiriman.
3. Pemesanan barang.
Perusahaan melakukan pemesanan kepada pemasok sesuai kebutuhan.
4. Penerimaan barang.
Barang yang diterima akan diperiksa kesesuaiannya dengan pesanan.
5. Pencatatan pembelian.
Data pembelian dicatat dalam sistem untuk keperluan administrasi dan pelaporan.
6. Pembayaran kepada pemasok.
Pembayaran dilakukan sesuai kesepakatan, baik secara tunai maupun kredit.

Hubungan antara penjualan dan pembelian sangat erat karena keduanya saling mempengaruhi. Pembelian yang tepat akan memastikan ketersediaan barang sehingga penjualan dapat berjalan lancar. Sebaliknya, tingkat penjualan yang tinggi akan mempengaruhi jumlah dan frekuensi pembelian. Maka diperlukan pengelolaan yang terintegrasi agar tercapai keseimbangan antara persediaan dan permintaan pasar [10].

Dalam kegiatan penjualan dan pembelian, terdapat konsep penting yaitu utang piutang. Utang merupakan kewajiban yang harus dibayar oleh suatu pihak kepada pihak lain, sedangkan piutang adalah hak untuk menerima pembayaran dari pihak lain. Dalam penjualan kredit, pelanggan memiliki utang kepada perusahaan, sementara perusahaan memiliki

piutang yang harus diterima. Sebaliknya, dalam pembelian kredit, perusahaan memiliki utang kepada pemasok [10].

Konsep utang piutang ini memiliki beberapa karakteristik, yaitu [10]:

1. Adanya pihak yang berutang dan berpiutang
2. Adanya jumlah kewajiban yang harus dibayar
3. Adanya jangka waktu pembayaran (jatuh tempo)
4. Adanya bukti transaksi atau perjanjian

Seiring dengan perkembangan teknologi, sistem informasi penjualan dan pembelian menjadi sangat penting dalam mendukung operasional perusahaan. Sistem ini membantu dalam pencatatan transaksi, pengelolaan stok, serta penyusunan laporan secara akurat dan real-time. Dengan adanya sistem yang terintegrasi, perusahaan dapat meningkatkan efisiensi serta mendukung pengambilan keputusan yang lebih tepat [10].

Sistem penjualan dan pembelian dalam praktiknya memiliki beberapa bentuk yang berkembang seiring dengan kemajuan teknologi dan perubahan perilaku konsumen. Secara umum, terdapat beberapa jenis sistem penjualan yang banyak digunakan, yaitu sebagai berikut [11]:

1. Penjualan Ritel di Toko Fisik (*Offline*).

Penjualan ritel di toko fisik merupakan metode penjualan tradisional yang dilakukan secara langsung antara penjual dan pembeli di suatu lokasi tertentu. Pada sistem ini, konsumen dapat melihat, memilih, dan membeli produk secara langsung. Keunggulan dari sistem ini adalah adanya interaksi langsung serta kepercayaan konsumen yang lebih tinggi terhadap produk yang dibeli.

2. Penjualan Melalui *Online* (*Website*/Media Sosial).

Penjualan *online* dilakukan dengan memanfaatkan media internet seperti website maupun platform media sosial. Dalam sistem ini, transaksi tidak dilakukan secara tatap muka, melainkan melalui komunikasi digital. Penjualan online memberikan kemudahan bagi konsumen dalam mengakses produk kapan saja dan di mana saja, serta memperluas jangkauan pasar bagi perusahaan.

3. Penjualan Melalui *E-Commerce*.

E-commerce merupakan bentuk penjualan yang menggunakan platform khusus seperti marketplace untuk melakukan transaksi jual beli secara digital. Sistem ini biasanya sudah dilengkapi dengan fitur pembayaran, pengiriman, serta manajemen pesanan. Penggunaan *e-commerce* mempermudah proses transaksi serta meningkatkan keamanan dan kepercayaan konsumen.

4. Pemanfaatan *Digital Catalog*.

Digital catalog adalah media digital yang digunakan untuk menampilkan daftar produk secara sistematis, lengkap dengan deskripsi, harga, dan gambar produk. Katalog digital memudahkan konsumen dalam melihat informasi produk tanpa harus datang langsung ke toko. Selain itu, digital catalog juga dapat terintegrasi dengan sistem penjualan untuk mempercepat proses pemesanan.

2.2 Toko Ritel Minuman Beralkohol

Toko ritel minuman beralkohol merupakan usaha yang bergerak di bidang penjualan minuman beralkohol secara langsung kepada konsumen akhir dalam jumlah eceran. Minuman beralkohol yang dijual biasanya meliputi berbagai jenis seperti *bir*, *wine*, dan *spirit* (misalnya *whiskey*, *vodka*, dan *rum*). Toko ritel ini berperan sebagai perantara antara distributor atau produsen dengan konsumen, sehingga memiliki peran penting dalam rantai distribusi produk minuman beralkohol [12].

Dalam operasionalnya, toko ritel minuman beralkohol memiliki karakteristik khusus dibandingkan dengan toko ritel pada umumnya. Hal ini disebabkan oleh adanya regulasi yang ketat terkait penjualan dan distribusi minuman beralkohol. Di Indonesia, penjualan minuman beralkohol harus memenuhi persyaratan perizinan tertentu, seperti Nomor Induk Berusaha (NIB), serta izin khusus seperti Surat Izin Usaha Perdagangan Minuman Beralkohol (SIUP-MB). Regulasi ini bertujuan untuk mengontrol peredaran minuman beralkohol serta mengurangi dampak negatif yang dapat ditimbulkan [12].

Selain aspek legalitas, pengelolaan toko ritel minuman beralkohol juga harus memperhatikan pengendalian stok dan variasi produk. Setiap jenis minuman memiliki karakteristik berbeda, baik dari segi kadar alkohol, harga, maupun segmentasi pasar. Dengan memahami preferensi konsumen, pengelola toko dapat menyediakan produk yang lebih sesuai dengan kebutuhan pasar. Ketersediaan produk yang tepat akan berpengaruh terhadap tingkat penjualan dan kepuasan pelanggan [12].

Tantangan dalam aktivitas pemasaran juga dialami oleh toko ritel minuman beralkohol, khususnya saat merancang dan menjalankan program promosi. Hal ini dikarenakan adanya pembatasan dalam iklan dan promosi minuman beralkohol. Strategi pemasaran yang digunakan biasanya lebih menekankan pada penataan produk (*display*), pelayanan pelanggan, serta pemanfaatan teknologi seperti sistem informasi penjualan dan rekomendasi produk untuk meningkatkan pengalaman konsumen [12].

Selain itu, perkembangan teknologi informasi telah mendorong toko ritel untuk mengadopsi sistem berbasis digital dalam operasionalnya. Sistem ini digunakan untuk mengelola data penjualan, stok barang, serta analisis perilaku konsumen. Dengan adanya sistem yang terintegrasi, pengelolaan toko menjadi lebih efisien dan akurat, sehingga dapat mendukung pengambilan keputusan yang lebih baik [13].

Toko ritel minuman beralkohol tidak hanya berfokus pada aktivitas penjualan, tetapi juga melibatkan aspek regulasi, manajemen persediaan, strategi pemasaran, serta pemanfaatan teknologi. Pengelolaan yang baik terhadap aspek-aspek tersebut akan membantu toko dalam meningkatkan kinerja operasional dan daya saing di pasar [13].

2.2.1 Jenis-Jenis Minuman Beralkohol

Toko ritel minuman beralkohol menyediakan berbagai jenis produk dengan karakteristik yang berbeda-beda. Jenis-jenis minuman beralkohol yang umum dijual antara lain [14]:

1. *Beer*.

Beer merupakan minuman beralkohol yang dihasilkan melalui proses fermentasi bahan dasar seperti malt (biji-bijian), hop, ragi, dan air. *Beer* memiliki kadar alkohol yang relatif rendah, biasanya berkisar antara 3% hingga 8%. Minuman ini termasuk yang paling populer dan banyak dikonsumsi secara global karena rasanya yang ringan dan mudah diterima.

2. *Wine*.

Wine adalah minuman beralkohol yang dibuat dari fermentasi buah anggur. Kadar alkohol *wine* umumnya berkisar antara 8% hingga 15%. *Wine* memiliki berbagai jenis seperti red wine, white wine, dan sparkling wine, yang masing-masing memiliki karakteristik rasa yang berbeda tergantung pada jenis anggur dan proses pembuatannya.

3. *Whiskey*.

Whiskey merupakan minuman beralkohol hasil distilasi dari biji-bijian yang difermentasi, seperti gandum, jagung atau barley. *Whiskey* biasanya disimpan dalam tong kayu untuk proses pematangan, sehingga menghasilkan aroma dan rasa yang khas. Kadar alkohol *whiskey* umumnya berkisar antara 40% hingga 50%.

4. *Vodka*.

Vodka adalah minuman beralkohol hasil distilasi yang umumnya dibuat dari kentang atau biji-bijian. *Vodka* memiliki karakteristik rasa yang netral dan jernih, sehingga sering

digunakan sebagai bahan dasar dalam berbagai jenis koktail. Kadar alkohol *vodka* biasanya sekitar 40%.

5. *Rum*.

Rum merupakan minuman beralkohol yang dihasilkan dari fermentasi dan distilasi gula tebu atau *molase*. *Rum* memiliki rasa yang cenderung manis dan sering digunakan dalam minuman campuran (*cocktail*). Kadar alkohol rum biasanya berkisar antara 37% hingga 50%.

6. *Gin*.

Gin adalah minuman beralkohol yang dihasilkan melalui distilasi dengan tambahan bahan botani, terutama buah juniper yang memberikan aroma khas. *Gin* sering digunakan sebagai bahan utama dalam berbagai koktail. Kadar alkohol gin umumnya sekitar 37% hingga 45%.

7. *Tequila*.

Tequila merupakan minuman beralkohol khas Meksiko yang dibuat dari tanaman *agave* biru. Proses pembuatannya melibatkan fermentasi dan distilasi. *Tequila* memiliki rasa yang khas dan biasanya dikonsumsi langsung atau sebagai campuran dalam koktail. Kadar alkohol *tequila* umumnya sekitar 35% hingga 50%.

8. *Sake*.

Sake adalah minuman beralkohol tradisional Jepang yang dibuat melalui fermentasi beras. Berbeda dengan minuman lain, *sake* tidak melalui proses distilasi. Kadar alkohol *sake* biasanya berkisar antara 12% hingga 16%, dengan rasa yang cenderung ringan dan sedikit manis.

9. *Soju*.

Soju merupakan minuman beralkohol tradisional Korea yang biasanya dibuat dari beras atau bahan lain seperti ubi. *Soju* memiliki rasa yang ringan dan mudah diminum, dengan kadar alkohol yang bervariasi antara 16% hingga 25%. Minuman ini sangat populer di Korea Selatan.

10. *Cognac*.

Cognac adalah jenis *brandy* yang berasal dari wilayah *Cognac* di Prancis dan dibuat dari anggur melalui proses distilasi serta pematangan dalam tong kayu. *Cognac* memiliki rasa yang kompleks dan biasanya dikonsumsi sebagai minuman premium. Kadar alkoholnya sekitar 40%.

11. *Liquor (Liqueur).*

Liqueur merupakan minuman beralkohol yang telah ditambahkan rasa manis serta bahan tambahan seperti buah, rempah atau krim. *Liqueur* biasanya memiliki kadar alkohol yang lebih rendah dibandingkan minuman hasil distilasi murni, yaitu sekitar 15% hingga 30%. Minuman ini sering digunakan sebagai campuran dalam koktail atau diminum sebagai pencuci mulut.

2.2.2 Klasifikasi Minuman Beralkohol

Minuman beralkohol merupakan minuman yang mengandung etanol (alkohol) yang dihasilkan melalui proses fermentasi maupun distilasi bahan yang mengandung gula atau pati. Klasifikasi minuman beralkohol umumnya didasarkan pada kadar alkohol yang terkandung di dalamnya. Pengelompokan ini penting karena berkaitan dengan pengawasan, distribusi, serta ketentuan penjualannya [15].

Berdasarkan kadar alkohol, minuman beralkohol diklasifikasikan menjadi tiga golongan, yaitu sebagai berikut [15]:

1. Golongan A.

Minuman beralkohol golongan A memiliki kadar alkohol rendah, yaitu antara 1% sampai dengan 5%. Minuman dalam golongan ini umumnya dihasilkan melalui proses fermentasi. Contohnya adalah bir dan minuman sejenisnya.

2. Golongan B

Minuman beralkohol golongan B memiliki kadar alkohol sedang, yaitu lebih dari 5% sampai dengan 20%. Contoh minuman pada golongan ini antara lain anggur (wine) dan minuman fermentasi lainnya dengan kadar alkohol yang lebih tinggi dibandingkan golongan A.

3. Golongan C

Minuman beralkohol golongan C merupakan minuman dengan kadar alkohol tinggi, yaitu lebih dari 20% sampai dengan 55%. Minuman ini biasanya dihasilkan melalui proses distilasi, seperti whiskey, vodka, rum, dan sejenisnya.

Selain berdasarkan kadar alkohol, klasifikasi juga dapat dilihat dari proses pembuatannya, yaitu minuman hasil fermentasi dan minuman hasil distilasi. Klasifikasi ini membantu dalam menentukan pengaturan distribusi serta pengendalian konsumsi oleh pemerintah [15].

2.3 Artificial Intelligence (AI)

Artificial Intelligence (AI) merupakan sistem yang dirancang untuk meniru kemampuan kognitif manusia dalam memecahkan masalah. AI berperan sebagai otak sistem yang belajar mengenali pola defisiensi nutrisi melalui data historis (dataset) tanpa perlu diprogram secara eksplisit untuk setiap aturan visualnya [8]. Tujuan utama AI adalah untuk mengembangkan sistem yang dapat melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia [16].

Beberapa tujuan utama AI meliputi [17]:

1. Pengolahan Bahasa Alami (*Natural Language Processing*)
AI bertujuan untuk mengembangkan sistem yang mampu memahami, menghasilkan, dan berinteraksi dengan bahasa manusia secara alami. Hal ini melibatkan kemampuan sistem untuk memahami konteks, sintaksis, semantik, dan pragmatik dalam bahasa manusia.
2. Pembelajaran Mesin (*Machine Learning*)
AI bertujuan untuk mengembangkan algoritma dan teknik yang memungkinkan mesin untuk belajar dari data dan pengalaman, tanpa diprogram secara eksplisit. Pembelajaran mesin melibatkan identifikasi pola, membuat prediksi, dan mengoptimalkan kinerja sistem berdasarkan data yang diberikan.
3. Pengenalan Pola (*Pattern Recognition*)
AI bertujuan untuk mengembangkan sistem yang dapat mengenali pola dan fitur dalam data, seperti pengenalan wajah, pengenalan suara, dan pengenalan tulisan tangan. Ini melibatkan penggunaan algoritma dan teknik untuk mengidentifikasi dan mengklasifikasikan pola-pola yang ada dalam data.
4. Penalaran dan Pengambilan Keputusan (*Reasoning and Decision Making*)
AI bertujuan untuk mengembangkan sistem yang dapat menganalisis informasi, melakukan penalaran, dan membuat keputusan berdasarkan pemahaman konteks dan aturan yang ditentukan. Ini melibatkan penggunaan logika formal, inferensi, dan pemodelan pengetahuan untuk memungkinkan sistem AI untuk mengambil keputusan yang masuk akal.

Komponen utama dalam sistem AI meliputi [17]:

1. Basis Pengetahuan (*Knowledge Base*)
Tempat penyimpanan informasi dan pengetahuan yang digunakan oleh sistem AI. Basis pengetahuan dapat berisi fakta, aturan, konsep, dan model yang relevan dengan domain atau tugas yang sedang dihadapi.

2. Mesin Inferensi (*Inference Engine*)

Mesin inferensi bertanggung jawab untuk melakukan proses logika dan inferensi berdasarkan informasi yang ada dalam basis pengetahuan. Ini melibatkan penerapan aturan, penalaran, dan manipulasi pengetahuan untuk mencapai kesimpulan atau membuat keputusan.

3. Antarmuka Pengguna (*User Interface*).

Antarmuka pengguna memungkinkan interaksi antara pengguna manusia dan sistem AI. Ini dapat berupa antarmuka grafis, perintah suara atau interaksi berbasis teks yang memungkinkan pengguna untuk memberikan input, menerima *output*, dan berkomunikasi dengan sistem AI.

4. Algoritma dan Teknik Kecerdasan Buatan (*AI Algorithms and Techniques*).

Ada berbagai algoritma dan teknik yang digunakan dalam AI, termasuk pembelajaran mesin, jaringan saraf tiruan, logika proposisional, pemrosesan bahasa alami, dan banyak lagi. Algoritma ini memungkinkan sistem AI untuk memproses data, mengenali pola, belajar dari pengalaman, dan membuat keputusan.

Ada beberapa jenis AI yang digunakan dalam berbagai konteks dan aplikasi. Dalam hal ini, kita akan menjelajahi beberapa jenis utama AI, termasuk sistem berbasis aturan, sistem pakar, pembelajaran mesin, jaringan saraf tiruan, dan AI yang terkait dengan robotika.

Berikut adalah beberapa jenis AI yang umum [17]:

1. Sistem Berbasis Aturan.

Jenis AI ini menggunakan aturan-aturan yang ditentukan sebelumnya untuk menghasilkan output berdasarkan input yang diberikan. Sistem berbasis aturan terdiri dari himpunan aturan dan basis pengetahuan yang menggambarkan hubungan antara *input* dan *output*. Sistem ini bekerja dengan mencocokkan input dengan aturan yang sesuai untuk menghasilkan *output* yang diinginkan.

2. Sistem Pakar.

Sistem pakar adalah jenis AI yang menggabungkan pengetahuan ahli manusia dengan kemampuan komputasi untuk memecahkan masalah yang kompleks dalam domain tertentu. Sistem ini beroperasi dengan mengumpulkan pengetahuan dari para ahli manusia dan memanfaatkannya dalam membuat keputusan atau memberikan nasihat yang sesuai. Sistem pakar digunakan dalam bidang-bidang seperti kedokteran, keuangan, dan teknik.

3. Pembelajaran Mesin.

Pembelajaran mesin adalah cabang AI yang melibatkan pengembangan algoritma dan model statistik yang memungkinkan sistem untuk belajar dari data dan pengalaman. Dalam pembelajaran mesin, sistem menganalisis data untuk mengidentifikasi pola dan membuat prediksi atau keputusan berdasarkan informasi yang diberikan. Algoritma pembelajaran mesin dapat diklasifikasikan menjadi beberapa kategori, termasuk pembelajaran terawasi, pembelajaran tak terawasi, dan pembelajaran penguatan.

4. Jaringan Saraf Tiruan.

Jaringan saraf tiruan (*artificial neural networks*) adalah jenis AI yang terinspirasi oleh struktur dan fungsi jaringan saraf dalam otak manusia. Jaringan saraf tiruan terdiri dari banyak unit pemrosesan sederhana yang disebut neuron buatan, yang saling terhubung dan berinteraksi untuk memproses informasi. Jaringan saraf tiruan dapat digunakan untuk tugas-tugas seperti pengenalan pola, pengenalan wajah, dan pemrosesan bahasa alami.

5. AI dalam Robotika.

AI juga digunakan dalam pengembangan robotika, di mana sistem AI dikombinasikan dengan fisik robot untuk menghasilkan perilaku yang cerdas dan adaptif. Robotika AI melibatkan penggunaan algoritma pengolahan sensorik, perencanaan gerakan, dan pembelajaran untuk memungkinkan robot berinteraksi dengan lingkungan sekitarnya dan melakukan tugas-tugas tertentu.

AI merupakan sistem yang dirancang untuk meniru kemampuan kognitif manusia dalam memecahkan masalah kompleks. Salah satu cabang utama dari Artificial Intelligence yang digunakan adalah *Machine Learning*, yang berfokus pada kemampuan sistem untuk belajar dari data dan meningkatkan kinerjanya secara otomatis. Sub bab berikutnya akan dibahas lebih lanjut mengenai konsep dan penerapan *Machine Learning* dalam sistem rekomendasi yang dikembangkan [18][19].

2.4 Machine Learning

Machine Learning (ML) adalah cabang ilmu komputer yang berkembang dari bidang pengenalan pola dan teori pembelajaran komputer dalam kecerdasan buatan. ML mempelajari konstruksi dan analisis algoritma pembelajaran dan prediksi data. Algoritma ML membangun model berdasarkan *input* model untuk mendapatkan sebuah prediksi atau keputusan yang didasarkan dari data daripada mengikuti instruksi program statis secara ketat [20].

ML dikategorikan dalam 3 kategori besar yaitu [21]:

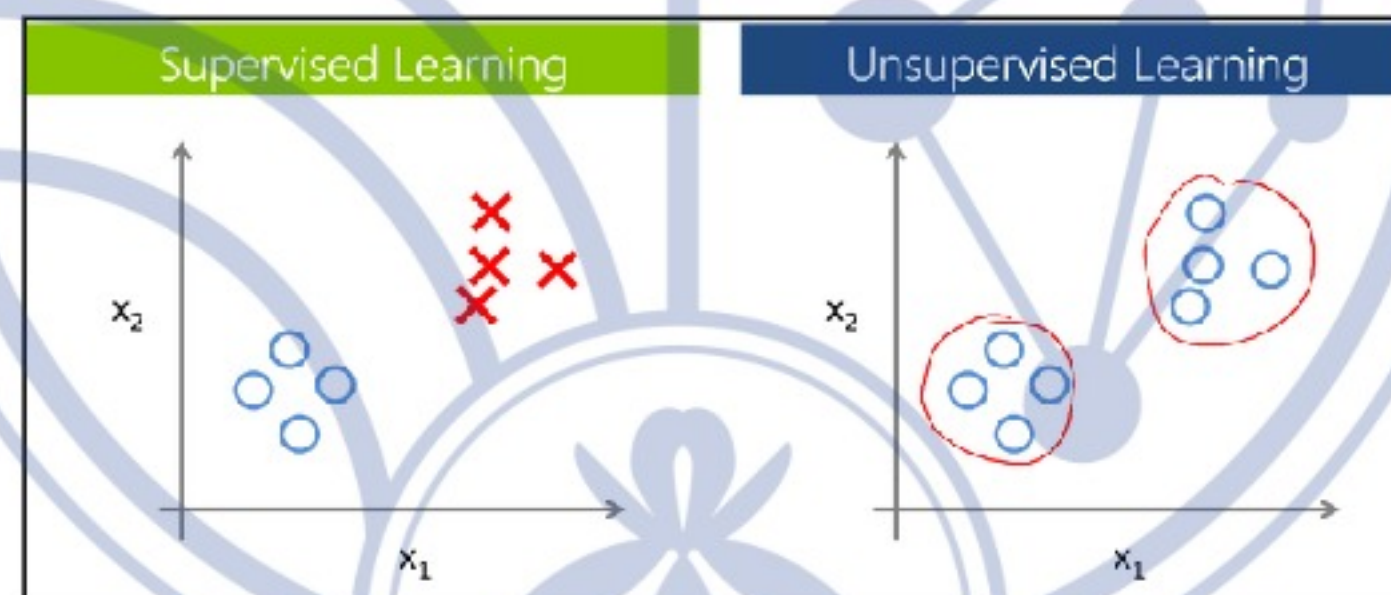
1. *Supervised Learning*.

Model atau algoritma disajikan dengan contoh masukan dan keluaran yang diinginkan kemudian dicari pola dan hubungan antara masukan dan keluaran tersebut. Tujuannya adalah untuk mempelajari aturan umum yang memetakan *input* ke *output*. Proses pelatihan berlanjut hingga model mencapai tingkat akurasi yang diinginkan pada data pelatihan. Beberapa contoh implementasi dari *Supervised Learning* di kehidupan nyata adalah:

- a. Klasifikasi gambar dimana dilakukan pelatihan dengan menggunakan gambar beserta label yang sesuai. Tujuannya adalah agar komputer dapat mengenali objek pada gambar yang baru di masa yang akan datang.
- b. Prediksi dimana komputer dilatih menggunakan data harga saham sebelumnya untuk menghasilkan prediksi harga di masa depan.

2. *Unsupervised Learning*.

Pada *Unsupervised Learning* data tidak diberikan label pada algoritma pembelajaran, dimana algoritma tersebut berusaha menemukan pola sebagai masukannya. Implementasi *Unsupervised Learning* dapat dimanfaatkan dalam mengelompokkan populasi yang memiliki perbedaan. *Unsupervised Learning* dapat mengelompokkan data tersendiri dan menemukan pola tersembunyi yang berada di dalam data.



Gambar 2.1 Perbedaan *Supervised Learning* dan *Unsupervised Learning*

Seperti yang dapat dilihat pada Gambar 2.1, data dalam *Supervised Learning* diberi label, sedangkan data dalam *Unsupervised Learning* tidak diberi label.

3. *Reinforcement Learning*.

Suatu bidang interdisipliner dalam pemelajaran mesin dan kontrol optimal yang berkaitan dengan bagaimana suatu agen cerdas dapat mengambil aksi di lingkungan yang dinamis dalam rangka untuk memaksimalkan penghargaan kumulatif. Dalam *Reinforcement Learning*, pengembang merancang metode untuk menghargai perilaku yang diinginkan dan menghukum perilaku negatif. Metode ini memberikan nilai positif

pada tindakan yang diinginkan untuk mendorong agen dan nilai negatif pada perilaku yang tidak diinginkan. Metode ini memprogram agen untuk mencari imbalan keseluruhan jangka panjang dan maksimum untuk mencapai solusi yang optimal.

Berdasarkan kategori pembelajaran dalam *Machine Learning* tersebut, termasuk dalam pendekatan *Unsupervised Learning*. Hal ini karena metode yang digunakan adalah algoritma *K-Means Clustering*, yang bekerja dengan mengelompokkan data produk minuman beralkohol berdasarkan kemiripan karakteristik tanpa menggunakan label data sebelumnya.

2.5 *Unsupervised Learning*

Unsupervised Learning adalah jenis pembelajaran mesin (*machine learning*) di mana algoritma dilatih menggunakan data tanpa label atau supervisi eksplisit. Berbeda dengan supervised learning, yang memiliki input-output yang diketahui, *Unsupervised Learning* hanya menggunakan data mentah dan bertujuan untuk menemukan pola atau struktur tersembunyi dalam data. Metode ini banyak digunakan dalam berbagai aplikasi seperti clustering, yang mengelompokkan data berdasarkan kesamaan fitur, serta dimensionality reduction, yang menyederhanakan jumlah fitur dalam data untuk analisis lebih lanjut [22].

Terdapat beberapa cara kerja dari *Unsupervised Learning* dalam melakukan klasifikasi terhadap data, antara lain [22]:

1. Mengumpulkan data tanpa label atau kategori yang diketahui.
2. Mengekstraksi fitur atau representasi dari data (misal menggunakan vektor numerik).
3. Menggunakan algoritma *clustering*, seperti *K-Means*, *Hierarchical Clustering* atau DBSCAN, untuk mengelompokkan data berdasarkan kemiripan.
4. Mengevaluasi hasil *cluster* dengan metrik internal, seperti *Silhouette Score* atau *Within-Cluster Sum of Squares* (WCSS).

Adapun kelemahan dan kelebihan dari teknik *Unsupervised Learning*, antara lain [22]:

1. Kelebihan:
 - a. Tidak memerlukan data berlabel, sehingga dapat diterapkan pada berbagai jenis data baru.
 - b. Mampu menemukan pola, struktur atau hubungan tersembunyi dalam data.
 - c. Berguna untuk segmentasi, reduksi dimensi, dan eksplorasi data.
2. Kekurangan:
 - a. Hasil pengelompokan sering sulit dievaluasi karena tidak ada *ground truth*.

- b. Memerlukan pemilihan parameter, misal jumlah cluster, yang bisa memengaruhi hasil akhir.
- c. Sensitif terhadap outlier dan data yang tidak representatif.

Contoh penerapan teknik *Unsupervised Learning* mencakup segmentasi pelanggan dalam bisnis, deteksi anomali dalam keamanan siber, klasifikasi terhadap data, serta kompresi gambar dalam pengolahan citra. Dengan kemampuannya mengungkap pola tersembunyi, *Unsupervised Learning* menjadi alat penting dalam eksplorasi data dan pengambilan keputusan berbasis data seperti dalam sistem rekomendasi.

2.6 Sistem Rekomendasi

Sistem rekomendasi merupakan bagian dari pengembangan sistem informasi yang berfungsi untuk memberikan saran atau rekomendasi kepada pengguna berdasarkan data, preferensi, maupun pola perilaku pengguna sebelumnya. Di era digital saat ini, perkembangan teknologi informasi yang sangat pesat menyebabkan terjadinya ledakan informasi (*information overload*), sehingga pengguna sering mengalami kesulitan dalam menentukan pilihan yang sesuai dengan kebutuhan mereka. Sebagai solusi untuk membantu proses pengambilan keputusan secara lebih efektif dan efisien hadir sistem rekomendasi [23].

Sistem rekomendasi adalah sistem yang dirancang untuk memprediksi preferensi atau penilaian (*rating*) seorang pengguna terhadap suatu item, kemudian memberikan saran item yang paling relevan. *Item* yang dimaksud dapat berupa produk, buku, film, layanan, maupun informasi lainnya. Sistem ini bekerja dengan memanfaatkan data interaksi antara pengguna dan item, seperti rating, riwayat pembelian, klik atau pencarian [23].

Dalam implementasinya, sistem rekomendasi memiliki beberapa komponen utama, yaitu [23]:

1. Matriks Interaksi.

Matriks interaksi merupakan representasi hubungan antara pengguna dan item dalam bentuk matriks berukuran $M \times N$, di mana M adalah jumlah pengguna dan N adalah jumlah item. Matriks ini berisi nilai rating atau interaksi pengguna terhadap item tertentu.

2. Fitur Pengguna dan Item.

Fitur pengguna dapat berupa data indikator maupun metadata yang menggambarkan karakteristik pengguna, sedangkan fitur item mencerminkan atribut yang dimiliki oleh suatu produk atau layanan.

3. Fungsi Representasi.

Fungsi ini bertugas merepresentasikan pengguna dan item ke dalam bentuk tertentu (misalnya vektor) agar dapat dihitung tingkat kesamaannya.

4. Fungsi Prediksi

Fungsi prediksi digunakan untuk memperkirakan rating atau tingkat kesukaan pengguna terhadap item yang belum pernah dinilai sebelumnya.

Pendekatan dalam sistem rekomendasi secara umum terbagi menjadi beberapa metode, antara lain [23]:

1. *Content-Based Filtering*, yaitu metode yang merekomendasikan item berdasarkan kesamaan atribut dengan item yang sebelumnya disukai oleh pengguna. Sistem membangun profil pengguna berdasarkan preferensi historisnya.
2. *Collaborative Filtering*, yaitu metode yang memberikan rekomendasi berdasarkan kesamaan perilaku atau preferensi antar pengguna. Jika dua pengguna memiliki pola rating yang mirip, maka item yang disukai oleh salah satu pengguna dapat direkomendasikan kepada pengguna lainnya.
3. *Knowledge-Based Filtering*, yaitu metode yang mengumpulkan preferensi pengguna kemudian merekomendasikan item baik berdasarkan sekumpulan aturan rekomendasi yang telah ditentukan sebelumnya, maupun dengan menggunakan metrik kesamaan untuk membantu mengidentifikasi item yang serupa dengan preferensi pengguna.
4. *Hybrid Approach*, yaitu kombinasi dari beberapa metode untuk meningkatkan akurasi dan mengurangi kelemahan masing-masing pendekatan.

Dengan adanya sistem informasi rekomendasi, organisasi maupun perusahaan dapat meningkatkan kualitas layanan, membantu pengguna menemukan informasi yang relevan, serta mendukung pengambilan keputusan berbasis data. Sistem rekomendasi menjadi salah satu pilar penting dalam pengembangan sistem informasi modern, khususnya dalam bidang *e-commerce*, pendidikan, hiburan, dan layanan digital lainnya. Untuk membangun sistem rekomendasi yang efektif, diperlukan teknik pengolahan data yang mampu mengekstraksi informasi penting dari berbagai sumber, khususnya data teks seperti ulasan, deskripsi produk atau komentar pengguna. Teknik inilah yang dikenal dengan *Text Mining*, yang akan dibahas pada sub-bab berikutnya.

2.7 Text Mining

Text Mining dapat didefinisikan secara luas sebagai proses intensif pengetahuan di mana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan

menggunakan seperangkat alat analisis. Penambangan teks berusaha untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. Dalam kasus penambangan teks, bagaimanapun, sumber data adalah kumpulan dokumen, dan pola yang menarik ditemukan bukan di antara catatan *database* yang diformalkan tetapi dalam data tekstual yang tidak terstruktur dalam koleksi dokumen. Contohnya, kedua jenis sistem ini bergantung pada rutinitas *preprocessing*, algoritme penemuan pola, dan elemen lapisan presentasi seperti alat visualisasi untuk meningkatkan penelusuran set jawaban [24].

Text Mining memiliki beberapa tipe antara lain [24]:

1. *Search and Information Retrieval.*
Menyimpan dan menemukan kembali dokumen teks, termasuk mesin pencari dan kata kunci pencarian.
2. *Document Clustering.*
Pengelompokan dan kategorisasi istilah, potongan, paragraf atau dokumen menggunakan metode mining.
3. *Document Classification.*
Pengelompokan dan kategorisasi istilah, potongan, paragraf atau dokumen menggunakan metode *document classification*.
4. *Web Mining Data.*
Teknik ini diimplementasikan pada internet yang berfokus pada skala dan antar hubungan pada *website*.
5. *Information Extraction.*
Identifikasi dan ekstraksi fakta yang relevan.
6. *Natural Language Processing.*
Pemrosesan bahasa tingkat rendah yang biasanya digunakan untuk bahasa komputasi.
7. *Concept Extraction.*
Pengelompokan kata dan frase dalam grup yang sama tahap *preprocessing text*.

Implementasi *Text Mining* pada kasus nyata adalah sebagai berikut [25]:

1. Klasifikasi *email* untuk menentukan email *spam* atau bukan sehingga digunakan oleh pemilik layanan *email* untuk memasukkan email yang bukan spam ke folder inbox dan email spam dikirim ke folder spam atau junk. Metode yang umum digunakan adalah *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Logistic Regression*. Teknik ini membantu penyedia layanan *email* memisahkan *email spam* ke *folder spam* atau *junk*, sementara email non-*spam* masuk ke *folder inbox*.

2. Mengetahui sentimen terhadap tokoh publik dengan mengklasifikasikan komentar masyarakat di media sosial. Dengan pengetahuan ini dapat diimplementasikan dalam sebuah aplikasi yang memungkinkan penggunanya mengetik nama tokoh publik, kemudian aplikasi akan menarik data dari media sosial dengan kata kunci nama tokoh tersebut. Selanjutnya aplikasi akan otomatis untuk menampilkan berapa persen sentimen positif dan negatif. Untuk menganalisis sentimen terhadap tokoh publik, metode seperti *Naïve Bayes*, SVM, dan pendekatan berbasis leksikon sering digunakan. Aplikasi dapat menarik data dari media sosial dengan kata kunci nama tokoh tersebut dan menampilkan persentase sentimen positif dan negatif..
3. Aplikasi pemantauan keadaan fasilitas publik dengan mengklasifikasikan komentar penduduk suatu kota dari media sosial. Aplikasi secara otomatis menampilkan fasilitas publik yang rusak atau yang perlu perhatian. Metode seperti *Naïve Bayes* dan SVM dapat digunakan untuk mengklasifikasikan komentar terkait kondisi fasilitas publik. Aplikasi secara otomatis menampilkan fasilitas publik yang rusak atau memerlukan perhatian berdasarkan analisis komentar masyarakat.
4. Mengetahui kinerja pelayanan publik suatu lembaga pemerintahan dengan melakukan klasifikasi data komentar masyarakat baik dari survei atau media sosial. Dengan pengetahuan ini dapat dibuat aplikasi infografik yang menampilkan kinerja lembaga-lembaga pemerintahan berdasarkan komentar masyarakat tersebut. Analisis sentimen dengan SVM adalah salah satu teknik untuk memprediksi sentimen pada sebuah teks, baik itu positif, negatif atau netral. Dengan pengetahuan ini, dapat dibuat aplikasi infografik yang menampilkan kinerja lembaga-lembaga pemerintahan berdasarkan komentar masyarakat tersebut.
5. Mengetahui sentimen terhadap produk atau jasa dengan mengklasifikasikan komentar netizen di media sosial. Metode seperti *Naïve Bayes*, SVM, dan pendekatan berbasis leksikon dapat digunakan untuk menganalisis sentimen terhadap produk atau jasa. Implementasinya dapat berupa hasil klasifikasi sentimen yang menampilkan sentimen positif dan negatif terhadap produk atau jasa tertentu.
6. Aplikasi pemantauan keadaan saat terjadi bencana alam dengan mengklasifikasikan komentar masyarakat untuk membedakan pesan bencana yang ditulis oleh masyarakat yang merasakan bencana alam, pesan bencana yang ditulis masyarakat namun bukan saksi mata langsung, dan pesan yang tidak berhubungan dengan bencana. Dengan membedakan ketiga kategori ini maka aplikasi dapat menentukan lokasi bencana dengan menggunakan geolokasi dari komentar yang ditulis oleh masyarakat yang

merasakan bencana alam tersebut. Metode seperti *K-Nearest Neighbor* (K-NN) dapat digunakan untuk mengklasifikasikan komentar masyarakat terkait bencana alam.

Dalam proses *text mining*, tahap awal yang dilakukan adalah *preprocessing data*. Tahap ini bertujuan untuk membersihkan dan mempersiapkan data teks agar dapat diolah lebih lanjut. Tahapan *preprocessing* yang dilakukan adalah sebagai berikut [25]:

1. *Lowercase* yaitu proses mengubah seluruh huruf dalam teks menjadi huruf kecil agar semua kata memiliki format yang seragam dan tidak membedakan antara huruf besar dan huruf kecil.
2. *Cleansing* yaitu proses menghapus karakter simbol, tanda baca, angka, atau karakter alfabet seperti titik, koma, tanda seru, dan lainnya agar teks menjadi lebih bersih dan hanya berisi kata-kata yang relevan.
3. *Tokenization* yaitu proses memecah kalimat atau teks menjadi unit-unit kata (*token*) sehingga setiap kata dapat dianalisis secara terpisah dalam proses pengolahan data.
4. *Stopword removal* yaitu proses menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti kata sambung atau kata penghubung (contoh: “dan”, “yang”, “di”, “a”, “the”) agar fokus analisis hanya pada kata bermakna.
5. *Lemmatization* yaitu proses mengubah kata ke bentuk dasarnya (kata dasar) berdasarkan makna linguistik, sehingga variasi kata dengan arti yang sama dapat disatukan menjadi satu bentuk standar.

Selain *preprocessing* pada data teks, *text mining* juga dapat melibatkan data numerik atau atribut tambahan yang digunakan dalam analisis. Pada tahap *preprocessing* juga dilakukan penyesuaian terhadap data numerik agar memiliki skala yang seragam. Salah satu metode yang digunakan adalah *Min-Max Scaling* yang akan dijelaskan pada sub bab berikutnya [25].

2.8 *Min-Max Scaling*

Min-Max Scaling merupakan salah satu metode normalisasi data yang digunakan untuk mengubah nilai atribut numerik ke dalam rentang tertentu, umumnya antara 0 hingga 1. Metode ini bertujuan untuk menyamakan skala setiap fitur sehingga tidak terdapat perbedaan rentang nilai yang terlalu jauh antar variabel, yang dapat memengaruhi hasil perhitungan dalam proses analisis data atau algoritma machine learning. Proses *Min-Max Scaling* dilakukan dengan cara mentransformasikan setiap nilai data berdasarkan nilai minimum dan maksimum dari atribut tersebut. Rumus yang digunakan adalah sebagai berikut [26]:

$$x^i = \frac{x - x_{min}}{x_{max} - x_{min}}$$

(1)

Keterangan:

x = nilai asli

x' = nilai hasil normalisasi

x_{min} = nilai minimum pada data

x_{max} = nilai maksimum pada data

Setelah melalui tahap *preprocessing*, data teks yang telah bersih dan terstruktur belum dapat langsung digunakan dalam proses analisis, sehingga diperlukan tahap transformasi data menjadi bentuk numerik. Proses transformasi dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) yang berfungsi untuk memberikan bobot pada setiap kata berdasarkan tingkat kemunculan kata dalam suatu dokumen serta tingkat kepentingannya terhadap keseluruhan dokumen. Hasil dari proses TF-IDF ini akan digunakan sebagai representasi fitur utama dalam proses pengelompokan data menggunakan algoritma.

2.9 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode klasik dalam mengekstrak kata kunci yang digunakan untuk menilai kebermaknaan kata atau frasa dalam teks. Terdapat dua faktor penting pada TF-IDF yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). TF mengacu pada frekuensi kemunculan suatu kata pada dokumen, di mana semakin tinggi frekuensinya, semakin penting kata tersebut, sementara IDF mengacu pada dokumen di mana suatu kata berada dalam sebuah korpus. Dalam korpus ini, jika frekuensi kata tersebut tinggi, maka kata tersebut dianggap kurang representatif dan kurang penting, sebaliknya, jika frekuensi kata tersebut rendah, maka kata tersebut dianggap sangat representatif dan penting [27].

Metode TF-IDF termasuk salah satu metode populer dalam menentukan bobot setiap kata. TF-IDF memberikan bobot pada kata-kata dalam dokumen teks, mencerminkan tingkat kepentingan kata-kata tersebut dalam dokumen. Tujuan dari metode TF-IDF adalah mendapatkan label atau sentimen dari setiap kata yang terdapat dalam dokumen. Data yang telah melewati tahap *pre-processing* kemudian diproses dengan penggunaan pembobotan kata dengan metode TF-IDF. Tahap penentuan nilai bobot pada metode TF-IDF yaitu [28].

1. Perhitungan *Term Frequency* (TF) dalam dokumen teks yang diasumsikan memiliki tingkat kepentingan setara pada tiap dokumen. Rumus TF adalah:

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \dots\dots\dots$$

(2)

Dimana:

$f_{t,d}$ adalah frekuensi kata t dalam dokumen d

$\sum_k f_{k,d}$ adalah total semua kata dalam dokumen d

2. Selanjutnya, semakin jarang *term* tersebut muncul dalam dokumen, nilai *Inverse Document Frequency* (IDF) akan meningkat. Rumus IDF adalah:

$$IDF(t) = \log \left(\frac{N}{1+n_t} \right)$$

(3)

Dimana:

N adalah total jumlah dokumen

n_t = jumlah dokumen yang mengandung kata t

Ditambah 1 di penyebut untuk menghindari pembagian dengan nol

3. TF-IDF merupakan hasil perkalian antara TF dan IDF.

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \dots\dots\dots$$

(4)

Dengan nilai bobot kata yang diperoleh melalui TF-IDF, setiap dokumen atau teks dapat direpresentasikan dalam bentuk vektor numerik yang mencerminkan tingkat kepentingan kata-kata di dalamnya. Representasi vektor ini memudahkan penerapan metode analisis data lebih lanjut, seperti *clustering*, untuk mengelompokkan dokumen atau teks berdasarkan kemiripan kontennya. TF-IDF menjadi tahap awal yang penting sebelum proses *clustering* dilakukan, sehingga kelompok yang terbentuk dapat mencerminkan kesamaan tema atau karakteristik antara dokumen.

2.10 Clustering

Clustering atau klasterisasi adalah suatu teknik atau metode untuk mengelompokkan data. *Clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa

cluster atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum. *Clustering* merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma *clustering* [29].

Clustering adalah teknik penting dalam data *mining* dan analisis data yang berfokus pada pengelompokan objek serupa ke dalam klaster. Setiap jenis clustering memiliki keunggulan dan keterbatasannya masing-masing. Pemilihan metode clustering yang tepat bergantung pada jenis data, tujuan analisis, dan sumber daya yang tersedia. Memahami perbedaan antara berbagai jenis clustering ini penting dalam menentukan pendekatan yang paling efektif untuk analisis data tertentu. Ada berbagai jenis clustering, masing-masing dengan karakteristik dan aplikasi yang berbeda. Berikut adalah penjelasan tentang beberapa jenis *clustering* utama [30]:

1. *Clustering Partitional*.

- a. *Clustering partitional* mengelompokkan data menjadi sejumlah klaster tanpa adanya tumpang tindih antar klaster.
- b. Algoritma paling populer dalam kategori ini adalah *K-Means*, yang bekerja dengan menentukan *centroid* secara acak dan kemudian mengelompokkan objek berdasarkan kedekatan mereka dengan *centroid* tersebut.
- c. *K-Means* efektif untuk dataset besar, tetapi sensitif terhadap pemilihan awal *centroid* dan cenderung terpengaruh oleh outlier.
- d. Algoritma lainnya termasuk *K-Medoids* dan *K-Medians*, yang serupa dengan *K-Means* tetapi kurang sensitif terhadap *outlier*.

2. *Clustering Hierarkis*.

- a. *Clustering* hierarkis mengelompokkan data dengan cara membangun hierarki klaster.
- b. Metode ini dapat dibagi menjadi dua jenis: *agglomerative (bottom-up)* dan *divisive (top-down)*.
- c. Dalam pendekatan *agglomerative*, setiap objek awalnya dianggap sebagai klaster sendiri dan klaster digabungkan secara bertahap berdasarkan kedekatan mereka. Proses ini berlanjut sampai semua objek berada dalam satu klaster atau kriteria tertentu terpenuhi.

- d. Pendekatan *divisive* bekerja dengan cara yang berlawanan, di mana semua objek awalnya berada dalam satu kluster besar yang kemudian dibagi menjadi kluster yang lebih kecil.
 - e. *Clustering* hierarkis sering kali divisualisasikan menggunakan dendrogram, yang menunjukkan bagaimana kluster dibentuk.
3. *Clustering* Berbasis Densitas.
- a. *Clustering* ini fokus pada identifikasi area dengan densitas tinggi yang dipisahkan oleh area dengan densitas rendah.
 - b. *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) adalah algoritma populer dalam jenis *clustering* ini. DBSCAN tidak memerlukan penentuan jumlah kluster terlebih dahulu dan dapat mengidentifikasi outlier sebagai noise.
 - c. Algoritma ini cocok untuk data dengan bentuk kluster yang tidak teratur dan beragam
4. *Clustering* Berbasis Model.
- a. *Clustering* ini didasarkan pada pemodelan matematika dari data dan kluster.
 - b. *Gaussian Mixture Models* (GMM) adalah contoh populer, di mana setiap kluster dimodelkan sebagai distribusi Gaussian.
 - c. Kelebihan metode ini adalah fleksibilitas dalam bentuk kluster, tetapi memerlukan pemahaman yang baik tentang model statistik
5. *Clustering* Berbasis Grid.
- a. Data dibagi menjadi sejumlah grid dan *clustering* dilakukan di dalam grid tersebut.
 - b. Algoritma seperti STING dan CLIQUE adalah contoh dari metode ini.
 - c. Cocok untuk data spasial dan dapat sangat efisien dari segi komputasi.

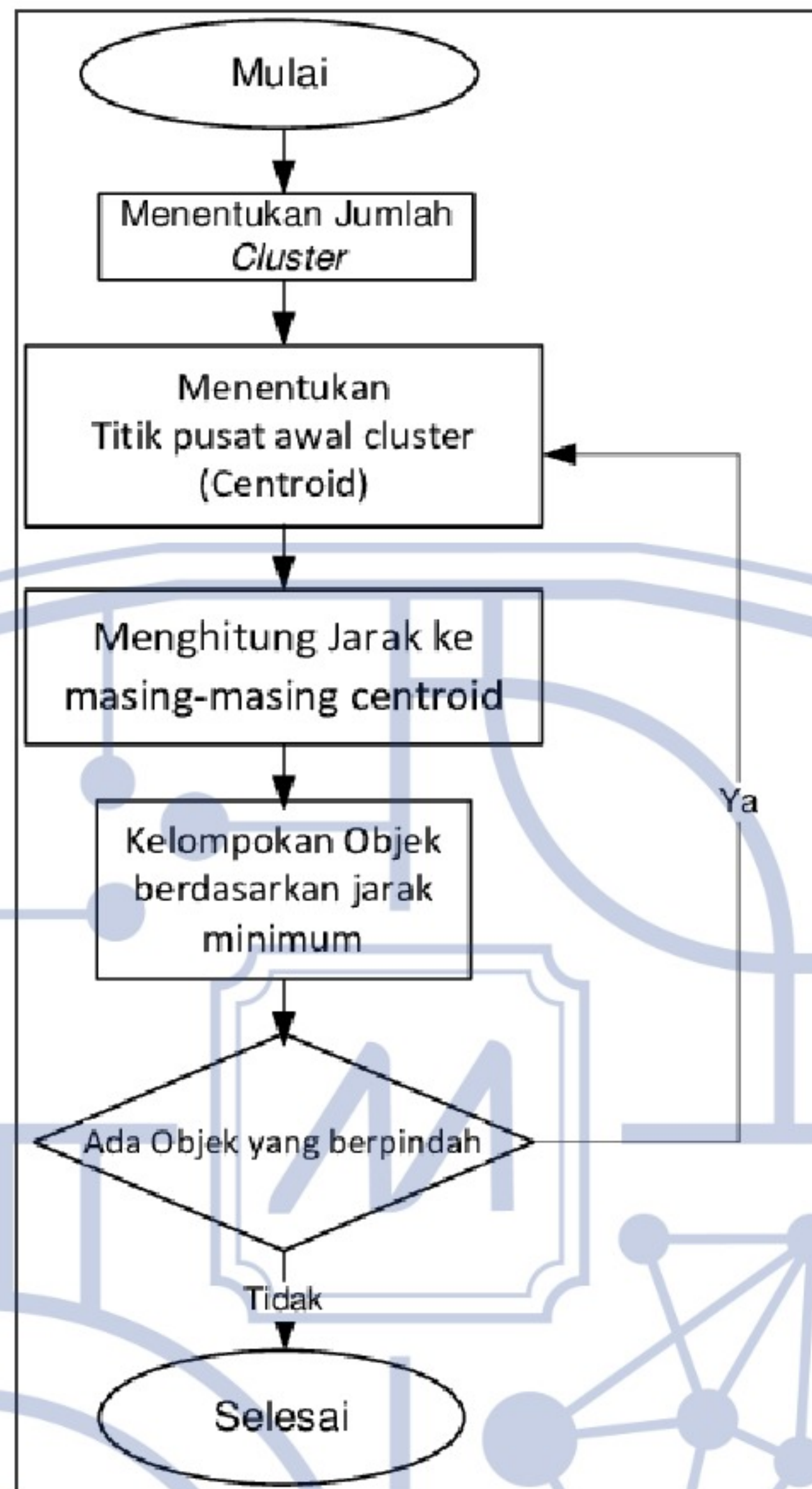
Algoritma *clustering* yang digunakan adalah *K-Means* dikarenakan algoritma ini sederhana, efisien, dan mampu mengelompokkan data dalam jumlah besar dengan cepat. Selain itu, *K-Means* memungkinkan sistem rekomendasi untuk mengelompokkan produk minuman beralkohol berdasarkan kesamaan deskripsi atau karakteristik, sehingga setiap cluster mewakili kelompok produk dengan profil rasa atau atribut yang serupa. Dengan pengelompokan ini, sistem dapat memberikan saran produk yang lebih relevan kepada pengguna, meningkatkan personalisasi rekomendasi, dan mempermudah pengguna dalam menemukan produk yang sesuai dengan preferensi mereka.

2.11 Algoritma *K-Means Clustering*

Algoritma *K-Means* merupakan salah satu metode *clustering* yang paling populer dalam teknik *Unsupervised Learning*. Algoritma ini termasuk dalam kategori *partitioning clustering*, yaitu metode pengelompokan data dengan cara membagi sekumpulan data ke dalam sejumlah kelompok (*cluster*) yang telah ditentukan sebelumnya. Tujuan utama dari algoritma *K-Means* adalah untuk meminimalkan jarak antara objek data dengan pusat *cluster* (*centroid*), sehingga objek yang berada dalam satu cluster memiliki tingkat kemiripan yang tinggi dibandingkan dengan objek pada *cluster* lainnya [32].

Prinsip dasar dari algoritma *K-Means* adalah mengelompokkan data berdasarkan kedekatan jarak terhadap titik pusat cluster (*centroid*). Setiap data akan ditempatkan pada *cluster* yang memiliki jarak paling dekat dengan *centroid*. Proses pengelompokan ini dilakukan secara iteratif hingga posisi *centroid* tidak lagi mengalami perubahan atau hingga mencapai kondisi konvergen. Algoritma *K-Means* mampu menghasilkan kelompok data yang memiliki karakteristik serupa dalam satu *cluster* dan perbedaan yang jelas dengan cluster lainnya [33].

Dengan proses tersebut, algoritma *K-Means* mampu menghasilkan kelompok data yang memiliki karakteristik serupa dalam satu cluster dan perbedaan yang jelas dengan cluster lainnya. Selain itu, metode ini banyak digunakan dalam berbagai bidang seperti pengelompokan pelanggan dan analisis data penjualan karena memiliki keunggulan dalam kesederhanaan, efisiensi, serta kemampuannya dalam mengolah data dalam jumlah besar. Tahapan proses kerja algoritma *K-Means* secara umum ditunjukkan pada Gambar 2.2 yang menggambarkan alur iterasi proses pengelompokan data [34].



Gambar 2.2 *Flowchart* Cara Kerja Algoritma *K-Means* [34]

Berdasarkan *flowchart* pada Gambar 2.2 diuraikan cara kerja dari algoritma *K-Means*, antara lain [34]:

1. Tentukan jumlah kelompok yang diinginkan:

$$k = \text{jumlah cluster}$$

(5)

2. Menentukan titik pusat awal (*centroid*):

$$C_j = (X_1, X_2, \dots, X_n)$$

(6)

Dengan:

$$C_j = \text{centroid ke-}j$$

n = jumlah fitur

3. Menghitung jarak ke masing-masing *centroid* dengan *Euclidean Distance*:

$$d(X_i, C_j) = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2}$$

.....
(7)

Keterangan:

X_i = data ke-i

C_j = *centroid* ke-j

n = jumlah atribut

4. Mengelompokkan objek berdasarkan jarak minimum:

Setelah jarak antara setiap data dengan *centroid* dihitung, maka setiap objek akan dimasukkan ke dalam *cluster* yang memiliki jarak paling kecil (minimum). Tahap ini menghasilkan pembagian data ke dalam beberapa cluster sesuai dengan kedekatan masing-masing objek terhadap *centroid*.

5. Mengecek apakah ada objek yang berpindah:

Setelah proses pengelompokan dilakukan, langkah selanjutnya adalah mengecek apakah terdapat objek yang berpindah dari satu *cluster* ke *cluster* lainnya dibandingkan dengan iterasi sebelumnya. Jika masih terdapat objek yang berpindah (jawaban “Ya”), maka proses akan kembali ke tahap penentuan *centroid* dan perhitungan jarak.

6. Proses iterasi hingga kondisi stabil:

Proses penghitungan *centroid*, pengukuran jarak, dan pengelompokan objek akan terus diulang hingga tidak ada lagi objek yang berpindah *cluster* (jawaban “Tidak”). Kondisi ini menunjukkan bahwa algoritma telah mencapai konvergensi.

7. Jika sudah tidak ada objek yang berpindah, maka proses clustering dinyatakan selesai dan hasil akhir berupa kelompok data (*cluster*) yang stabil telah diperoleh.

2.12 Metode *Elbow*

Metode *Elbow* merupakan salah satu metode yang digunakan untuk menentukan jumlah *cluster* yang optimal dalam proses pengelompokan data. Metode ini dilakukan dengan mengevaluasi nilai *Sum of Squared Errors* (SSE), yaitu jumlah total jarak kuadrat antara setiap data dengan pusat cluster yang terbentuk. Semakin kecil nilai SSE, maka semakin baik tingkat keseragaman data dalam suatu *cluster*. Secara matematis, SSE dapat dirumuskan sebagai berikut [35]:

$$SSE(k) = \sum_{i=1}^N \sum_{j=1}^N 1(x_i \in C_j) \cdot ||x_i - c_j||^2$$

(8)

Keterangan:

$SSE(k)$ = jumlah *error kuadrat* pada jumlah *cluster* ke- k

N = jumlah seluruh data

k = jumlah *cluster*

x_i = data ke- i

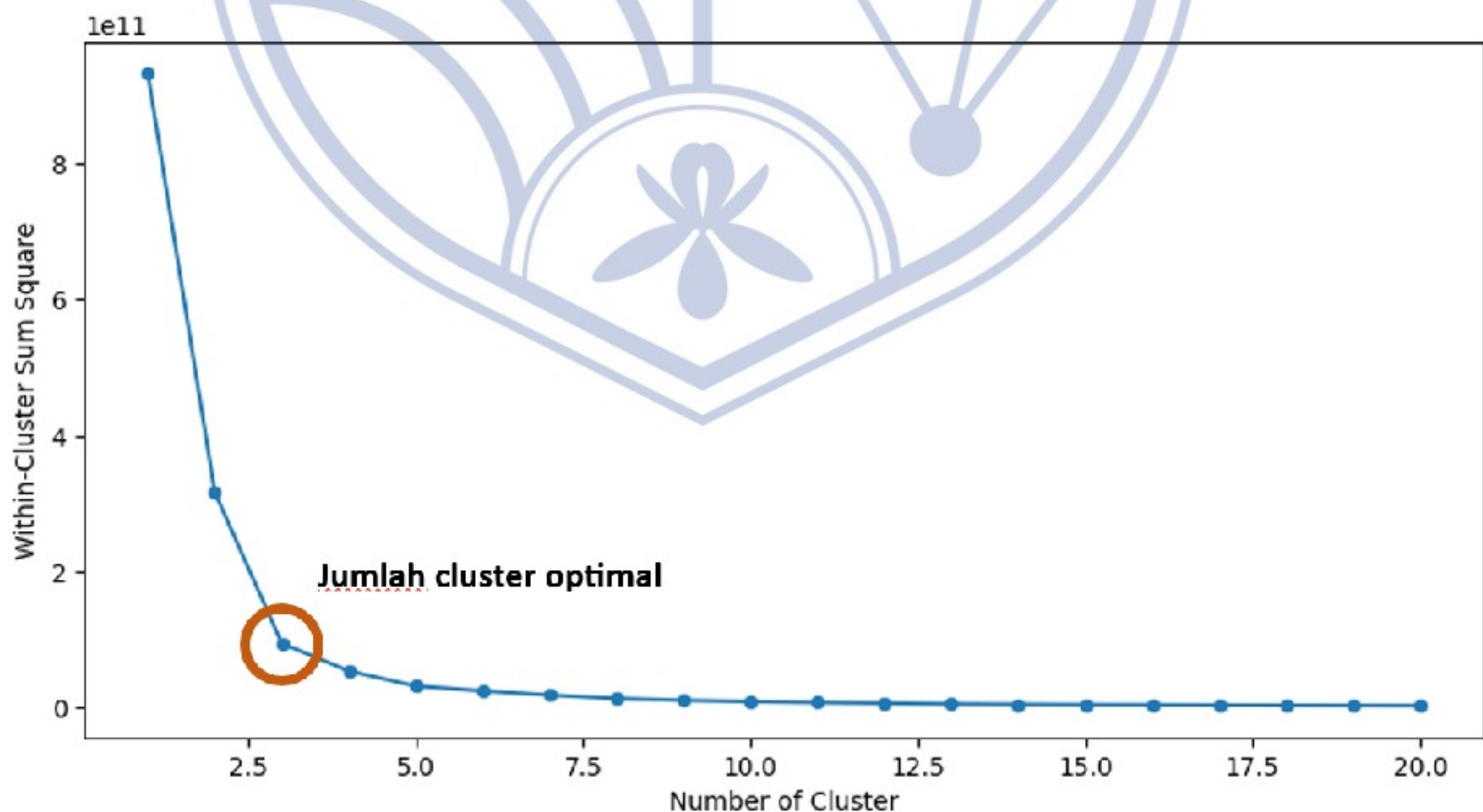
C_j = *cluster* ke- j

c_j = *centroid* (pusat *cluster*) pada *cluster* ke- j

$1(x_i \in C_j)$ = fungsi indikator (1 jika data termasuk *cluster* tersebut, 0 jika tidak)

$||x_i - c_j||^2$ = jarak kuadrat antara data dan *centroid*

Dalam penerapannya, nilai SSE dihitung untuk beberapa variasi jumlah *cluster* (k), kemudian setiap nilai tersebut diplot ke dalam bentuk grafik. Grafik ini menunjukkan hubungan antara jumlah *cluster* dengan nilai SSE yang dihasilkan. Penentuan jumlah *cluster* optimal dilakukan dengan melihat titik pada grafik yang membentuk pola seperti siku (*elbow point*), yaitu titik di mana penurunan nilai SSE mulai melambat secara signifikan. Setelah titik tersebut, penambahan jumlah *cluster* tidak lagi memberikan penurunan SSE yang signifikan, sehingga dianggap tidak efektif untuk meningkatkan kualitas pengelompokan data. Berikut ini, Gambar 2.3 merupakan ilustrasi grafik metode *Elbow* untuk menentukan jumlah *cluster* optimal [34].



Gambar 2.3 Grafik Metode *Elbow*

Untuk menentukan jumlah *cluster* yang optimal sebelum proses pengelompokan dilakukan, diperlukan suatu pendekatan yang dapat membantu memilih nilai k yang paling sesuai dengan karakteristik data. Pemilihan jumlah *cluster* yang tepat sangat berpengaruh terhadap kualitas hasil clustering, karena nilai k yang terlalu kecil atau terlalu besar dapat menurunkan akurasi pengelompokan [36]. Maka dari itu, penelitian menggunakan metode *Elbow* untuk membantu menentukan jumlah *cluster* optimal yang akan digunakan pada algoritma *K-Means*.

Seluruh konsep yang diuraikan diimplementasikan dalam sistem rekomendasi produk berbasis *website*, sehingga kualitas *website* menjadi aspek penting yang menentukan efektivitas sistem. Kualitas *website* tidak hanya mencakup tampilan visual dan navigasi, tetapi juga mencakup kecepatan akses, kemudahan penggunaan, keamanan, serta kemampuan sistem dalam menyajikan rekomendasi yang akurat dan responsif terhadap kebutuhan pengguna. Dengan memperhatikan kualitas *website*, sistem rekomendasi dapat memberikan pengalaman pengguna yang optimal, meningkatkan kepuasan pelanggan, serta mendorong penggunaan berulang dan loyalitas terhadap platform.

2.13 Kualitas Website

Kualitas *website* merupakan salah satu faktor penting dalam menentukan keberhasilan suatu sistem berbasis *web*. *Website* yang berkualitas tidak hanya mampu menyajikan informasi, tetapi juga memberikan pengalaman yang baik bagi pengguna. Kualitas *website* mencerminkan kemampuan suatu sistem dalam memenuhi kebutuhan pengguna secara efektif, efisien, dan memuaskan [37].

Selain itu, *website* yang baik adalah *website* yang mampu diakses dengan mudah, memiliki tampilan yang menarik, serta dapat berjalan dengan optimal di berbagai perangkat. Dalam era digital saat ini, pengguna mengakses *website* tidak hanya melalui komputer, tetapi juga melalui *smartphone* dan *tablet*. Maka *website* yang baik harus bersifat responsif, yaitu mampu menyesuaikan tampilan secara otomatis sesuai dengan ukuran layar perangkat pengguna. Hal ini bertujuan untuk meningkatkan kenyamanan dan kemudahan dalam penggunaan [37].

Dalam mengukur kualitas *website*, terdapat beberapa aspek utama yang umum digunakan, yaitu sebagai berikut [37]:

1. Kemudahan Penggunaan (*Usability*).

Usability merupakan tingkat kemudahan suatu *website* dalam digunakan oleh pengguna.

Website yang baik harus memiliki tampilan yang sederhana, mudah dipahami, serta

navigasi yang jelas. Pengguna harus dapat menemukan informasi yang dibutuhkan dengan cepat tanpa mengalami kebingungan. Selain itu, konsistensi desain dan struktur halaman juga menjadi bagian penting dalam meningkatkan *usability*.

2. Kualitas Informasi (*Information Quality*).

Kualitas informasi berkaitan dengan isi atau konten yang disajikan dalam website. Informasi yang ditampilkan harus akurat, relevan, lengkap, dan mudah dipahami. Selain itu, informasi juga harus diperbarui secara berkala agar tetap sesuai dengan kondisi terbaru. Penyajian informasi yang baik akan meningkatkan kepercayaan pengguna terhadap website.

3. Kinerja Sistem (*System Performance*).

Kinerja sistem mencakup kecepatan akses, responsivitas, dan stabilitas *website*. *Website* yang berkualitas harus memiliki waktu *loading* yang cepat serta mampu merespons permintaan pengguna dengan baik. Selain itu, performa *website* harus tetap optimal meskipun diakses oleh banyak pengguna secara bersamaan.

4. Keamanan (*Security*).

Keamanan merupakan aspek penting terutama bagi *website* yang mengelola data pengguna. *Website* harus mampu melindungi data dari ancaman seperti peretasan, pencurian data, dan penyalahgunaan informasi. Dengan adanya sistem keamanan yang baik, tingkat kepercayaan pengguna terhadap website akan meningkat.

5. Tampilan Visual (*Interface Design*).

Tampilan visual berkaitan dengan desain antarmuka *website* yang menarik dan nyaman dilihat. Penggunaan warna, font, gambar, serta tata letak yang tepat dapat meningkatkan pengalaman pengguna. Desain yang baik tidak hanya menarik secara visual, tetapi juga mendukung kemudahan penggunaan dan tidak membingungkan pengguna.

6. *Responsiveness* (Kesesuaian Perangkat).

Responsiveness adalah kemampuan *website* untuk menyesuaikan tampilan dan fungsi pada berbagai ukuran layar dan perangkat. *Website* yang responsif akan tetap nyaman digunakan baik pada *desktop* maupun perangkat *mobile*. Hal ini sangat penting karena mayoritas pengguna saat ini mengakses website melalui perangkat *mobile*.

Dengan memperhatikan aspek-aspek tersebut, kualitas *website* dapat ditingkatkan secara menyeluruh sehingga mampu memberikan pengalaman terbaik bagi pengguna. *Website* yang baik akan lebih efektif dalam mencapai tujuan, baik sebagai media informasi, sistem transaksi, maupun sebagai sarana pendukung pengambilan keputusan [37]. Setelah

website berhasil dibangun maka model dan sistem yang dibangun perlu diuji menggunakan beberapa teknik pengujian.

2.14 Teknik Pengujian

Terdapat beberapa teknik pengujian yang digunakan dan diuraikan pada sub bab berikutnya.

2.14.1 *Black Box Testing*

Black Box Testing atau sering disebut dengan pengujian fungsional adalah metode pengujian perangkat lunak yang digunakan untuk menguji perangkat lunak tanpa mengetahui struktur internal kode atau program. Tujuan utama pengujian ini adalah untuk memastikan bahwa perangkat lunak telah berfungsi sesuai dengan kebutuhan dan spesifikasi yang telah ditentukan. Dalam metode ini, penguji hanya mengetahui *input* dan *output* dari sistem, tanpa perlu mengetahui bagaimana proses di dalam sistem tersebut dijalankan [38].

Berikut beberapa metode yang digunakan untuk melakukan pengujian *Black Box*, yaitu [39]:

1. *Equivalence Partitioning*.

Equivalence Partitioning adalah teknik pengujian yang membagi himpunan *input* yang mungkin menjadi beberapa kelas ekuivalen (*equivalence class*), sehingga setiap anggota kelas dianggap memberikan hasil yang sama. Tujuannya adalah mengurangi jumlah kasus uji yang diperlukan tanpa mengurangi cakupan pengujian. Misalnya, jika sistem menerima *input* angka 1–100, maka rentang ini dapat dibagi menjadi kelas valid (1–100) dan kelas tidak valid (<1 atau >100). Dari setiap kelas, cukup diambil satu contoh sebagai *test case*.

2. *Boundary Value Analysis (BVA)*.

Boundary Value Analysis merupakan pengembangan dari *Equivalence Partitioning* dengan fokus pada nilai batas (*boundary values*) dari suatu *input*. Hal ini karena kesalahan pada program cenderung terjadi di sekitar batas minimal, maksimal atau transisi antar kelas *input*. Contohnya, jika *input* yang valid adalah 1–100, maka *test case* akan difokuskan pada nilai 1, 2, 99, 100, serta nilai di luar batas 0 dan 101.

3. *Syntax Testing*.

Syntax Testing adalah metode yang digunakan untuk menghasilkan kasus uji dari domain *input* yang telah didefinisikan secara formal, biasanya menggunakan *meta-language* seperti *Backus–Naur Form (BNF)*. Metode ini memanfaatkan *Abstrak Syntax*

Tree (AST) untuk memetakan struktur *input*. *Test case* yang valid dirancang untuk menutupi seluruh cabang AST, sedangkan *test case* yang tidak valid sengaja dibuat dengan kesalahan pada terminal atau *node* non terminal AST. Pendekatan ini sangat efektif untuk menguji sistem yang menerima *input* terstruktur, misalnya *compiler* atau *parser*.

4. *Random Testing*.

Random Testing adalah metode pengujian di mana *input* dihasilkan secara acak dari domain *input* yang mungkin. Tujuannya adalah menemukan kasus-kasus anomali yang tidak terpikirkan oleh penguji secara manual. Kelebihan metode ini adalah dapat menemukan *bug* tersembunyi dalam kondisi ekstrem atau tidak biasa, namun kekurangannya adalah tidak menjamin cakupan pengujian yang sistematis.

5. *Non-prescriptive Approaches to Blackbox Testing*.

Metode *non-prescriptive* ini bersifat lebih fleksibel dan tidak mengikuti aturan atau teknik formal tertentu. Pengujian dilakukan berdasarkan intuisi penguji, pengalaman sebelumnya, dan pemahaman terhadap sistem. Pendekatan ini bermanfaat ketika spesifikasi formal tidak tersedia atau sistem sangat kompleks, tetapi hasilnya cenderung lebih sulit diulang secara konsisten.

6. *Test Catalogues, Test Categories, dan Test Matrices*.

Metode ini melibatkan pendokumentasian kasus uji secara sistematis dalam bentuk katalog, kategori atau matriks. Dengan cara ini, setiap fungsi sistem, kombinasi *input*, dan jalur logika dapat dicatat, sehingga memudahkan penguji dalam menelusuri, menjalankan, dan melacak *test case*. Matriks pengujian biasanya memetakan *input* terhadap *output* yang diharapkan, sehingga memudahkan evaluasi hasil pengujian.

7. *Combinatorial Test Methods*.

Metode ini fokus pada pengujian kombinasi *input* yang berbeda-beda untuk memastikan bahwa sistem tetap berjalan dengan benar dalam berbagai kondisi *input* yang mungkin. Teknik ini sangat berguna ketika suatu sistem memiliki banyak parameter, karena jumlah kombinasi *input* bisa sangat besar, sehingga *combinatorial testing* membantu memprioritaskan kombinasi yang paling kritis untuk diuji.

8. *Category Partition Method* (CPM).

Category Partition Method digunakan untuk memformalkan dokumentasi kasus uji *Blackbox* dalam bahasa yang disebut *Test Specification Language* (TSL). Metode ini membagi *input* ke dalam kategori dan sub kategori, lalu menetapkan *test case* berdasarkan kombinasi kategori tersebut. CPM memberikan cara yang terstruktur untuk

menentukan cakupan pengujian dan menghasilkan dokumen *test case* yang lengkap dan dapat direproduksi.

9. *Classification Trees*.

Classification Trees adalah metode visual untuk merancang *test case*. *Input* dibagi ke dalam pohon klasifikasi, di mana setiap cabang mewakili kategori atau atribut *input*, dan kombinasi cabang membentuk *test case*. Keuntungan dari metode ini adalah mempermudah pemahaman hubungan antar *input*, mendukung *coverage* yang lebih baik, dan memudahkan pengelolaan kasus uji secara sistematis.

Metode *Blackbox Testing* yang digunakan adalah *Equivalence Partitioning* dikarenakan pendekatan ini membagi himpunan *input* yang mungkin ke dalam beberapa kelas ekuivalen, baik kelas valid maupun tidak valid, sehingga pengujian dapat dilakukan secara lebih efisien tanpa harus menguji seluruh kemungkinan *input* satu per satu. Metode ini sesuai untuk memastikan setiap fitur sistem berjalan dengan benar terhadap data yang valid (seperti pengisian *form*), maupun menangani data yang tidak valid (seperti *field* kosong) sebagai representasi dari kelas ekuivalen masing-masing skenario, sehingga cakupan pengujian tetap terjaga.

2.14.2 Silhouette Score

Silhouette Score merupakan salah satu metode yang digunakan untuk mengevaluasi kualitas hasil *clustering*. Metode ini diperkenalkan oleh Peter J. Rousseeuw (1987) untuk mengukur tingkat kemiripan suatu objek dengan *cluster* tempatnya berada dibandingkan dengan cluster lain. Dengan kata lain, *Silhouette Score* dapat menunjukkan seberapa baik suatu data ditempatkan dalam *cluster* yang sesuai [40].

Setiap data memiliki nilai *Silhouette* yang berkisar antara -1 hingga 1. Interpretasi nilai *Silhouette* adalah sebagai berikut [40]:

1. Nilai mendekati 1 menunjukkan bahwa data sangat sesuai dengan *cluster*-nya (*well-clustered*).
2. Nilai mendekati 0 menunjukkan bahwa data berada di batas antara dua *cluster* (*ambiguous*).
3. Nilai mendekati -1 menunjukkan bahwa data kemungkinan salah ditempatkan (*misclassified*) ke dalam *cluster* yang tidak tepat.

Rumus *Silhouette Score* untuk sebuah data i dapat dituliskan sebagai berikut [40]:

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

(9)

Dengan:

$a(i)$ = rata-rata jarak antara data i dengan semua data lain dalam *cluster* yang sama.

$b(i)$ = rata-rata jarak minimum antara data dengan data dari *cluster* lain yang paling dekat.

Nilai keseluruhan *Silhouette Score* diperoleh dari rata-rata nilai $s(i)$ seluruh data pada *dataset*.

2.14.3 System Usability Scale (SUS)

System Usability Scale (SUS) Merupakan metode skala sederhana, memiliki sepuluh item yang memberikan pandangan global tentang penilaian subjektif kegunaan. Metode *System Usability Scale* diperkenalkan oleh John Brooke di tahun 1986. Metode ini dapat digunakan untuk mengevaluasi berbagai jenis produk maupun pelayanan, yang mana di dalamnya termasuk Perangkat keras (*hardware*), perangkat lunak (*software*), perangkat mobile, *website* dan aplikasi. SUS dapat digunakan dalam melakukan pengujian sebuah teknologi yang independen baik pada perangkat keras, perangkat lunak, *website* dan bahkan perangkat seluler. SUS juga dikenal sebagai pengukuran kepuasan pengguna yang “*quick and dirty*” artinya penggunaan kuesioner SUS sangat cepat dan data yang dihasilkan dapat dipercaya [41]. Ada beberapa karakteristik dari SUS yang membuat menarik dan berbeda dari kuesioner lain. Pertama, SUS terdiri dari sepuluh pertanyaan, sehingga relatif cepat dan mudah bagi responden untuk menyelesaikan. Kedua, SUS menggunakan teknologi agnostik, yang berarti dapat digunakan secara luas dan mengevaluasi hampir semua jenis *interface*. Ketiga, hasil kuesioner adalah nilai tunggal, mulai dari skor 0 sampai 100, dan relatif mudah dipahami oleh berbagai disiplin, baik individu maupun kelompok [42].

Perhitungan hasil pengujian dengan metode SUS dilakukan dengan mengikuti beberapa aturan dalam proses menghitung skor SUS adalah sebagai berikut, pertama-tama jumlahkan kontribusi skor dari setiap item. Kontribusi skor tiap item berkisar dari 0 sampai dengan 4. Untuk item pernyataan angka 1,3,5,7, dan 9 kontribusi skor adalah skala posisi dikurangi 1. Untuk item pernyataan angka 2,4,6,8 dan 10 kontribusinya 5 dikurangi posisi skala. Selanjutnya Kalikan jumlah skor dengan 2,5 untuk mendapatkan nilai SUS secara

keseluruhan. Aturan dalam perhitungan skor untuk berlaku pada 1 responden. Untuk perhitungan selanjutnya, skor SUS dari masing-masing responden dicari skor rata-ratanya dengan menjumlahkan semua skor dan dibagi dengan jumlah keseluruhan dari responden. Untuk menghitung rata-rata dari keseluruhan skor SUS dapat dilihat pada persamaan 10.

Mencari rata-rata keseluruhan skor SUS [41]:

$$\bar{x} = \frac{\sum x}{n} \dots\dots\dots (10)$$

Dimana :

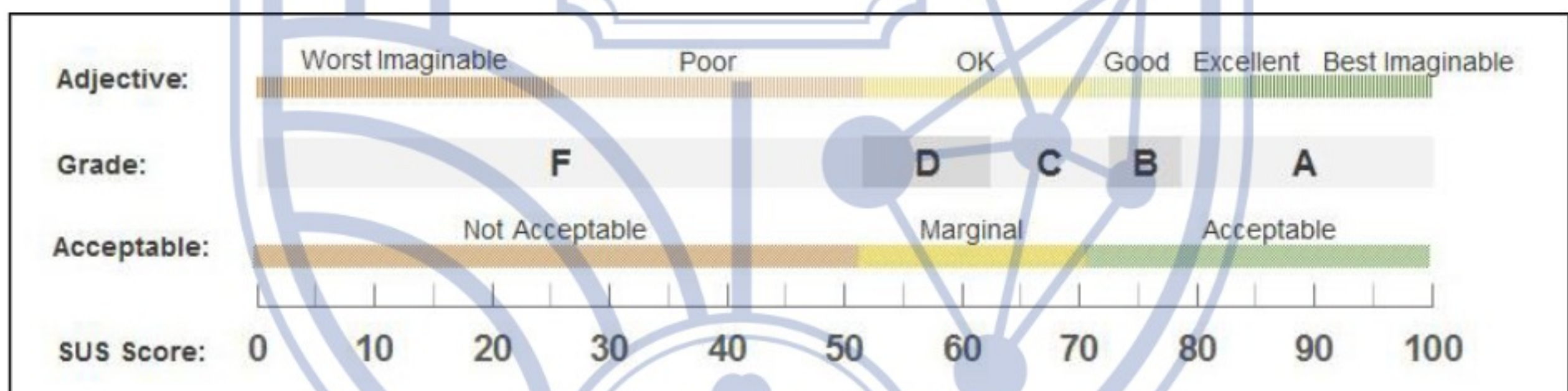
$\bar{x}$ = Skor rata-rata

$\sum x$ = Jumlah skor SUS

N = Jumlah responden

Dari hasil rekapitulasi skor SUS, kita akan mendapatkan skor rata-rata yang kemudian akan kita tarik kesimpulannya. Penentuan selanjutnya adalah menentukan *grade* hasil penilaian. Untuk menentukan *grade* hasil penilaian dapat disimpulkan dengan 3 cara penilaian yaitu [41]:

1. *Adjective Rating* dengan rentang penilaian terburuk hingga terbaik.
2. *Scale Grade* dengan rentang F hingga A.
3. *Acceptability Range* dengan rentang *Not Acceptable* hingga *Acceptable*.



Gambar 2.4 Skala Penilaian Akhir *System Usability Scale*

Pada Gambar 2.4, Skor yang dihasilkan kemudian diinterpretasikan sebagai berikut [41]:

1. *Acceptability Ranges*

Interpretasi skor *System Usability Scale* (SUS) berdasarkan penerimaan pengguna. Skala peringkat yang digunakan adalah *Not Acceptable*, *Marginal*, dan *Acceptable*.

2. *Grade Scale*

Skor *System Usability Scale* (SUS) dikelompokkan menjadi 5 *grade* yaitu: A (skor antara 90-100), B (skor antara 80-90), C (skor antara 70-80), D (skor antara 60-70), dan F (skor di bawah 60).

3. *Adjectives Rating*

Adalah kata sifat yang menerjemahkan skor numerik *System Usability Scale* (SUS) ke dalam penilaian absolut terhadap *usability*. Skala peringkat *adjectives* ini meliputi: *Worst Imaginable*, *Awful*, *Poor*, *OK*, *Good*, *Excellent*, dan *Best Imaginable*.

2.15 Penelitian Terdahulu

Berikut ini diuraikan beberapa penelitian terdahulu yang berkaitan dengan penelitian yang dilakukan pada Tabel 2.1.

Tabel 2.1 Ringkasan Penelitian Terdahulu

No.	Nama Peneliti	Tahun	Judul Penelitian	Hasil Penelitian
1	Fadhila Tangguh Admojoa dan Ahsanawatia	2020	Klasifikasi Aroma Alkohol Menggunakan Metode KNN	Hasil pengujian sistem ini terdiri dari pengujian pengaruh nilai K dan pengaruh nilai <i>cross validation</i> . Hasil dari pengujian pengaruh nilai K menghasilkan akurasi optimum senilai 100% pada nilai K=3 dan 100% pada nilai K=4 [43].
2	Awaljan, Tukino, Elfina Novalia, dan Sandi Ahmad	2022	Klasifikasi Hasil Penjualan Minuman Ringan Pada Koperasi Berdasarkan Jenis Barang Menggunakan Algoritma <i>K-Means Clustering</i>	Hasil penelitian ini menunjukkan bahwa 1 minuman ringan termasuk dalam <i>cluster</i> 0 yang tergolong volume penjualan tinggi, sedangkan 14 minuman ringan termasuk dalam cluster 1 yang tergolong volume penjualan rendah [44].
3	Rafael Arrelano	2023	<i>Clustering Data Jenis Wine Dengan Metode K-Means dan C-Means</i>	Hasil penelitian menunjukkan bahwa algoritma <i>K-Means</i> mendemonstrasikan kinerja yang lebih unggul, dengan nilai <i>Silhouette Score</i> mencapai 0.660522 atau 66.05%. Skor yang mendekati 1 atau 100% menandakan tingkat pemisahan klaster yang sangat baik. Di sisi lain, metode <i>C-Means</i> , meskipun memberikan informasi yang berharga tentang tingkat keanggotaan dalam kelompok, menunjukkan silhouette score

No.	Nama Peneliti	Tahun	Judul Penelitian	Hasil Penelitian
				yang lebih rendah, yaitu sebesar 0.3786 atau 37.86% [45].
4	Helson Matuan, Esau Dude, Atius Mallo, Herlina Yowey, Yusuf Selius Patey, dan Heru Sutejo	2026	<i>Application of the K-Means Method for Grouping Product Data Based on Sales Level</i>	Hasil penelitian menunjukkan bahwa penerapan <i>K-Means</i> berhasil mengelompokkan 99 produk Indomaret Kotaraja ke dalam empat kluster yang berbeda berdasarkan pola penjualan bulanan, dengan masing-masing kluster mencerminkan karakteristik produk fast moving, slow moving, musiman, dan fluktuatif, serta menghasilkan pemisahan yang jelas dan valid untuk mendukung keputusan operasional toko.

Seperti yang dapat dilihat pada Tabel 2.1, berbeda dengan penelitian terdahulu yang sebagian besar fokus pada pengelompokan produk berdasarkan penjualan atau karakteristik fisik dan aroma, penelitian ini mengembangkan sistem rekomendasi produk minuman beralkohol berbasis *website* dengan menggunakan TF-IDF dan *K-Means*. Penelitian ini tidak hanya mengelompokkan produk berdasarkan satu atribut, tetapi mempertimbangkan berbagai dimensi deskripsi produk secara simultan sehingga menghasilkan *cluster* yang lebih representatif terhadap preferensi pelanggan. Kontribusi utama dari penelitian ini adalah pengembangan sistem rekomendasi yang mampu memberikan saran produk yang lebih personal dan relevan, mengintegrasikan analisis teks dengan metode clustering, serta menyediakan antarmuka *website* yang dapat diakses secara fleksibel, berbeda dari penelitian terdahulu yang fokus pada analisis data internal tanpa implementasi sistem rekomendasi interaktif.