

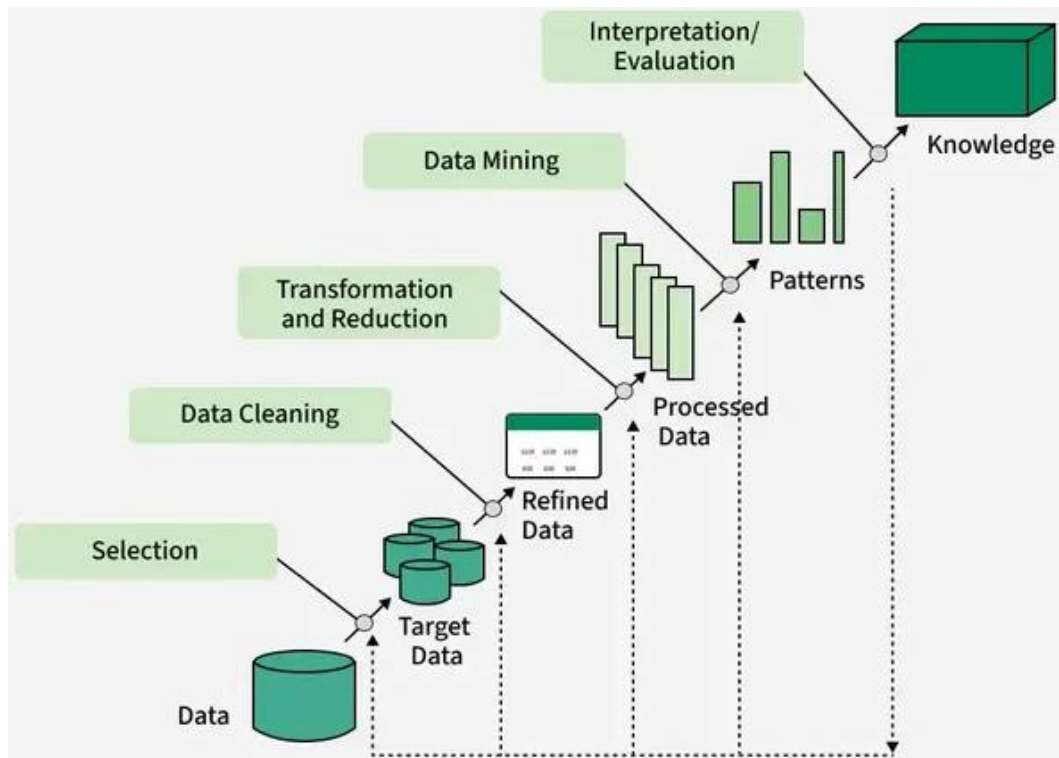
BAB II

LANDASAN TEORI

2.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Databases (KDD) merupakan suatu proses untuk menemukan informasi atau pengetahuan yang berguna dari sekumpulan data dalam jumlah besar. KDD tidak hanya berfokus pada hasil akhir berupa informasi, tetapi juga mencakup serangkaian tahapan yang sistematis mulai dari pemilihan data, pembersihan data, transformasi data, proses data mining, hingga evaluasi hasil. Tujuan utama dari KDD adalah mengubah data mentah yang awalnya tidak memiliki makna menjadi informasi yang bernilai dan dapat digunakan dalam pengambilan keputusan.

Dalam penerapannya, KDD terdiri dari beberapa tahapan penting, yaitu data selection, data cleaning, data transformation, data mining, dan interpretation/evaluation. Pada tahap data mining, berbagai metode seperti klasifikasi, clustering, dan asosiasi digunakan untuk menemukan pola dalam data. Setelah itu, hasil yang diperoleh akan dievaluasi untuk memastikan bahwa informasi yang dihasilkan relevan dan akurat. Dengan demikian, KDD menjadi kerangka kerja yang sangat penting dalam penelitian berbasis data, termasuk dalam analisis kepuasan pelanggan menggunakan metode seperti Naive Bayes dan K-Nearest Neighbor (KNN).



Gambar 2. 1. Knowledge Discovery in Databases (KDD)

Tahapan dalam Knowledge Discovery in Databases (KDD) merupakan proses yang sistematis untuk mengolah data menjadi informasi yang bermakna dan dapat digunakan dalam pengambilan keputusan. Proses KDD terdiri dari beberapa tahapan utama yang saling berhubungan, dimulai dari pemilihan data hingga interpretasi hasil. Setiap tahapan memiliki peran penting dalam memastikan bahwa pengetahuan yang dihasilkan benar-benar akurat dan relevan dengan tujuan penelitian.

Tahapan pertama adalah data selection, yaitu proses pemilihan data yang relevan dari berbagai sumber data yang tersedia. Pada tahap ini, data yang tidak berkaitan dengan tujuan penelitian akan dieliminasi, sehingga hanya data yang sesuai yang digunakan dalam proses selanjutnya. Tahap ini sangat penting untuk mengurangi kompleksitas data serta meningkatkan efisiensi proses analisis.

Tahapan kedua adalah data cleaning, yaitu proses pembersihan data dari kesalahan, nilai kosong (missing value), duplikasi data, maupun inkonsistensi data. Data yang tidak bersih dapat menyebabkan hasil analisis menjadi tidak akurat, sehingga tahap ini menjadi salah satu tahap yang paling krusial dalam KDD. Setelah data dibersihkan, dilakukan tahap data integration, yaitu menggabungkan data dari berbagai sumber menjadi satu kesatuan dataset yang utuh dan konsisten.

Tahapan berikutnya adalah data transformation, yaitu proses mengubah data ke dalam format yang sesuai untuk proses analisis. Transformasi data dapat berupa normalisasi, pengelompokan data ke dalam kategori tertentu, atau perubahan tipe data agar lebih mudah diproses oleh algoritma. Setelah itu, dilakukan tahap inti yaitu data mining, di mana berbagai metode seperti klasifikasi, clustering, atau asosiasi digunakan untuk menemukan pola atau hubungan tersembunyi dalam data.

Tahapan terakhir adalah evaluation dan interpretation, yaitu proses mengevaluasi hasil yang diperoleh dari data mining serta menafsirkan makna dari pola yang ditemukan. Pada tahap ini, peneliti menentukan apakah hasil yang diperoleh sudah sesuai dengan tujuan penelitian dan memiliki nilai guna. Jika hasil sudah valid, maka informasi tersebut dapat digunakan sebagai dasar dalam pengambilan keputusan. Dengan melalui seluruh tahapan KDD ini, data mentah dapat diubah menjadi pengetahuan yang bermanfaat secara sistematis dan terstruktur.

2.2. Tingkat Kepuasan

Tingkat kepuasan pelanggan merupakan ukuran yang menunjukkan sejauh mana harapan dan kebutuhan pelanggan terpenuhi setelah menggunakan suatu produk atau layanan [5]. Konsep ini menjadi indikator penting dalam menilai kualitas produk serta efektivitas strategi pemasaran yang diterapkan oleh Perusahaan [6]. Kepuasan pelanggan tidak hanya dipengaruhi oleh faktor fungsional seperti rasa, harga, dan kemasan, tetapi juga oleh aspek emosional yang timbul dari pengalaman saat mengonsumsi produk tersebut [7]. Semakin tinggi kesesuaian antara harapan dan kenyataan yang dirasakan pelanggan, maka semakin tinggi pula tingkat kepuasannya. Sebaliknya, apabila produk yang diterima tidak sesuai dengan ekspektasi, pelanggan cenderung merasa kecewa dan kemungkinan besar tidak akan melakukan pembelian ulang [8]. Oleh karena itu, mengukur tingkat kepuasan pelanggan menjadi salah satu strategi penting bagi perusahaan untuk memahami kebutuhan pasar dan meningkatkan loyalitas konsumen [9].

Tingkat kepuasan pelanggan terhadap minuman CS Float dapat mencerminkan bagaimana masyarakat menilai kualitas produk dari berbagai aspek seperti rasa, aroma, kesegaran, kemasan, hingga harga yang ditawarkan [10]. Produk yang mampu memberikan sensasi rasa yang menyenangkan dan harga yang dianggap sepadan akan lebih mudah mendapatkan penilaian positif dari pelanggan. Namun, perbedaan selera individu sering kali menimbulkan variasi dalam tingkat kepuasan yang dirasakan. Sebagian pelanggan mungkin menganggap rasa minuman ini sangat enak dan menyegarkan, sementara yang lain merasa terlalu manis atau kurang sesuai dengan preferensi mereka [11]. Variasi persepsi ini penting untuk

dianalisis agar perusahaan dapat memahami faktor-faktor dominan yang berpengaruh terhadap kepuasan pelanggan. Dengan pemahaman tersebut, perusahaan dapat memperbaiki kualitas produk, menyesuaikan strategi pemasaran, serta mempertahankan pelanggan yang loyal sekaligus menarik pelanggan baru secara lebih efektif.

2.3. Cs Float

CS Float merupakan salah satu produk minuman ringan berkarbonasi yang menawarkan sensasi menyegarkan dengan perpaduan rasa buah dan soda. Produk ini dirancang untuk memenuhi kebutuhan masyarakat modern yang menginginkan minuman praktis, enak, dan mudah dijangkau. Dengan kemasan yang menarik serta berbagai varian rasa seperti jeruk, anggur, dan stroberi, CS Float berhasil menarik perhatian berbagai kalangan, terutama anak muda yang gemar mencoba produk minuman baru. Harga yang relatif terjangkau juga menjadi salah satu daya tarik utama yang membuat CS Float mampu bersaing di pasar minuman ringan nasional. Di tengah banyaknya merek minuman soda dan jus instan yang beredar, CS Float menempatkan diri sebagai produk yang menyajikan cita rasa segar, warna yang menarik, dan kemudahan akses pembelian di berbagai tempat seperti minimarket, supermarket, dan toko kelontong. Keunggulan tersebut membuat CS Float menjadi salah satu produk yang cukup dikenal di kalangan konsumen minuman ringan.

Popularitas CS Float di pasaran tidak lepas dari beragam persepsi pelanggan terhadap rasa, harga, serta kualitas produk yang ditawarkan. Sebagian pelanggan merasa puas karena produk ini memiliki rasa yang enak, kemasan yang menarik, dan harga yang sesuai dengan kantong masyarakat. Namun, tidak sedikit juga yang

menilai rasa minuman terlalu manis atau harganya tidak sebanding dengan kualitas yang dirasakan. Perbedaan pandangan ini menjadikan data kepuasan pelanggan sebagai aspek penting untuk dianalisis secara lebih mendalam. Dengan mengolah data pelanggan yang mencakup variabel seperti preferensi rasa, frekuensi pembelian, dan tanggapan terhadap harga, dapat diperoleh pola yang menggambarkan kecenderungan tingkat kepuasan konsumen terhadap produk CS Float. Hasil analisis tersebut nantinya dapat memberikan pemahaman yang lebih baik tentang perilaku pelanggan serta membantu pihak produsen dalam menentukan strategi peningkatan kualitas produk dan pemasaran yang lebih tepat sasaran.

2.4. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan suatu proses sistematis untuk menemukan pengetahuan yang tersembunyi dari sekumpulan data yang besar dan kompleks [12]. Menurut Renatta, KDD tidak hanya sekadar menggali data, tetapi juga mencakup serangkaian tahapan analisis mulai dari pemilihan data, pembersihan data, transformasi data, hingga proses penemuan pola dan evaluasi hasil [13]. Proses ini bertujuan untuk mengubah data mentah menjadi informasi yang bermakna dan dapat digunakan dalam pengambilan Keputusan [14]. KDD menjadi sangat penting di era digital karena perusahaan, lembaga, maupun organisasi kini menghasilkan data dalam jumlah besar yang tidak dapat diolah secara manual. Melalui pendekatan ini, pola tersembunyi yang semula tidak terlihat dapat diidentifikasi, seperti tren perilaku pelanggan, kecenderungan pembelian, hingga preferensi individu terhadap produk atau layanan tertentu. Dengan

demikian, KDD berfungsi sebagai jembatan antara data mentah dan pengetahuan yang siap pakai untuk mendukung strategi bisnis yang berbasis data [15].

Penerapan konsep KDD sangat relevan untuk menggali informasi berharga dari data pelanggan yang berkaitan dengan tingkat kepuasan terhadap produk CS Float [16]. Melalui tahapan-tahapan seperti seleksi data pelanggan, pembersihan dari data yang tidak relevan, transformasi menjadi format yang sesuai untuk analisis, serta penerapan algoritma klasifikasi seperti *Naive Bayes* dan *K-Nearest Neighbor* (KNN), proses KDD dapat menghasilkan pengetahuan baru mengenai pola kepuasan konsumen [17]. Hasil dari proses ini dapat memperlihatkan faktor-faktor dominan yang memengaruhi kepuasan, seperti persepsi rasa, harga, kemasan, maupun frekuensi pembelian. Dengan mengubah data mentah menjadi informasi yang terstruktur, KDD membantu mengidentifikasi pelanggan yang berpotensi puas atau tidak puas terhadap produk, sehingga perusahaan dapat mengambil keputusan yang lebih tepat dalam strategi peningkatan kualitas dan layanan [18].

2.3.1. Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada kemampuan sistem komputer untuk belajar dari data tanpa harus diprogram secara eksplisit [19] [20]. Melalui proses pembelajaran, sistem akan mengenali pola atau hubungan antarvariabel dari data yang diberikan, kemudian menggunakan pola tersebut untuk melakukan prediksi atau pengambilan keputusan terhadap data baru [21]. Konsep dasar *Machine Learning* melibatkan algoritma yang memproses data secara iteratif hingga menghasilkan model dengan tingkat akurasi yang optimal. Teknologi ini telah banyak digunakan dalam berbagai

bidang seperti keuangan, kesehatan, pemasaran, dan analisis perilaku konsumen karena kemampuannya dalam mengolah data besar (*big data*) dan memberikan hasil yang cepat serta efisien [22]. Keunggulan utama *Machine Learning* terletak pada kemampuannya menyesuaikan diri terhadap perubahan data, sehingga model yang dihasilkan dapat terus berkembang dan meningkatkan kinerjanya seiring bertambahnya informasi yang diperoleh [23].

Penerapan *Machine Learning* menjadi solusi efektif dalam memahami pola kepuasan pelanggan terhadap suatu produk [24]. Melalui pendekatan ini, data pelanggan seperti preferensi rasa, frekuensi pembelian, dan persepsi terhadap harga dapat dianalisis untuk menemukan pola yang menggambarkan tingkat kepuasan mereka [25]. Dengan menggunakan algoritma klasifikasi seperti *Naive Bayes* dan *K-Nearest Neighbor (KNN)*, sistem dapat mempelajari hubungan antara karakteristik pelanggan dengan tingkat kepuasan yang mereka rasakan [26]. Hasil dari proses pembelajaran ini kemudian digunakan untuk memprediksi kategori pelanggan baru apakah termasuk dalam kelompok yang puas atau tidak puas [27]. Pendekatan semacam ini memungkinkan perusahaan untuk memperoleh wawasan yang lebih akurat dan berbasis data dalam menilai persepsi konsumen, sehingga strategi peningkatan kualitas produk dapat dilakukan secara lebih tepat sasaran dan efektif [28].

2.3.2. Database dan Data Processing

Database merupakan kumpulan data yang terorganisir dan disimpan secara sistematis agar mudah diakses, dikelola, dan diperbarui [29]. Di era digital, *Database* berperan penting dalam menyimpan berbagai informasi yang berasal dari

aktivitas pengguna, transaksi bisnis, maupun survei pelanggan. Struktur *Database* dirancang menggunakan sistem manajemen basis data (DBMS) yang memungkinkan pengguna untuk melakukan operasi seperti penyimpanan, pencarian, penghapusan, dan pembaruan data dengan efisien [30]. Data yang tersimpan dalam *Database* dapat berbentuk numerik, teks, maupun kategori, tergantung pada kebutuhan analisis yang akan dilakukan. Selain itu, proses pengolahan data atau *Data Processing* menjadi langkah penting dalam mempersiapkan data agar siap digunakan untuk analisis lebih lanjut [31]. Proses ini meliputi pembersihan data (*data cleaning*), transformasi, normalisasi, hingga pengelompokan atribut yang relevan. Tujuannya adalah memastikan bahwa data yang digunakan memiliki kualitas tinggi, bebas dari duplikasi maupun kesalahan, sehingga hasil analisis yang diperoleh dapat lebih akurat dan dapat diandalkan.

Database pelanggan yang berisi informasi seperti usia, jenis kelamin, frekuensi pembelian, dan persepsi terhadap produk digunakan sebagai sumber utama untuk melakukan analisis prediktif terhadap tingkat kepuasan pelanggan CS Float [32]. Data mentah yang dikumpulkan akan melalui proses pengolahan agar siap digunakan dalam model *Machine Learning*. Tahapan *Data Processing* dilakukan untuk membersihkan data dari nilai yang hilang, mengonversi data kategorik menjadi bentuk numerik, serta menormalkan nilai agar sesuai dengan kebutuhan algoritma yang digunakan. Setelah melalui tahapan tersebut, data siap untuk diolah menggunakan metode klasifikasi seperti *Naive Bayes* dan *K-Nearest Neighbor* (KNN). Dengan adanya pengelolaan *Database* dan pengolahan data yang baik, hasil analisis akan mampu memberikan gambaran yang lebih jelas mengenai

pola kepuasan pelanggan dan membantu menghasilkan model prediksi yang lebih akurat [33].

2.3.3. Visualization

Visualisasi merupakan salah satu tahapan penting dalam proses analisis data karena mampu menyajikan informasi kompleks dalam bentuk yang lebih mudah dipahami [34]. Melalui visualisasi, pola, tren, dan hubungan antarvariabel dapat diidentifikasi dengan lebih cepat dibandingkan hanya menggunakan data mentah. Teknik visualisasi dapat berupa grafik, diagram batang, diagram lingkaran, heatmap, scatter plot, maupun bentuk visual lainnya yang menampilkan hubungan antaratribut secara interaktif dan informatif [35]. Keberadaan visualisasi tidak hanya membantu peneliti dalam memahami karakteristik data, tetapi juga mempermudah pengambilan keputusan berdasarkan bukti yang terlihat secara visual. Dengan demikian, visualisasi data menjadi jembatan antara data statistik dan interpretasi manusia yang intuitif, sekaligus meningkatkan keakuratan dan efisiensi dalam analisis.

Penerapan visualisasi pada penelitian ini difokuskan untuk menggambarkan distribusi data pelanggan serta hasil klasifikasi tingkat kepuasan terhadap produk minuman CS Float. Melalui grafik dan diagram, dapat ditampilkan perbandingan jumlah pelanggan yang puas dan tidak puas, hubungan antara variabel seperti usia, harga, dan frekuensi pembelian dengan tingkat kepuasan, serta evaluasi hasil prediksi menggunakan metode *Naive Bayes* dan KNN. Tampilan visual ini membantu dalam menilai efektivitas model yang digunakan serta mengidentifikasi faktor dominan yang memengaruhi kepuasan pelanggan. Dengan adanya

visualisasi, hasil analisis yang diperoleh menjadi lebih mudah dipahami dan dapat dijadikan dasar yang kuat bagi perusahaan dalam menentukan strategi peningkatan kualitas produk maupun pelayanan.

2.3.4. Statistik

Statistik merupakan cabang ilmu yang mempelajari cara mengumpulkan, mengolah, menganalisis, menafsirkan, serta menyajikan data agar menghasilkan informasi yang berguna dalam pengambilan Keputusan [36]. Ilmu ini berperan penting dalam berbagai bidang seperti ekonomi, pendidikan, sosial, dan teknologi karena mampu menggambarkan fenomena berdasarkan data yang terukur. Melalui statistik, pola, kecenderungan, serta hubungan antar variabel dapat diketahui secara objektif. Analisis statistik tidak hanya berfungsi untuk menggambarkan data dalam bentuk angka, tetapi juga sebagai dasar untuk membuat kesimpulan atau prediksi terhadap fenomena yang sedang diamati. Oleh sebab itu, statistik menjadi landasan utama dalam penelitian ilmiah, terutama yang melibatkan data kuantitatif dan analisis berbasis numerik.

Statistik berperan penting dalam proses analisis data pelanggan CS Float, karena setiap atribut seperti rasa, harga, kemasan, dan frekuensi pembelian perlu diolah secara sistematis agar menghasilkan informasi yang bermakna. Melalui penerapan teknik statistik, data mentah yang diperoleh dari responden dapat diubah menjadi informasi yang mudah dipahami dan dapat digunakan untuk menentukan tingkat kepuasan pelanggan. Statistik juga menjadi dasar dalam proses penerapan metode *Machine Learning* seperti *Naive Bayes* dan *K-Nearest Neighbor (KNN)*, karena kedua algoritma tersebut bekerja dengan prinsip pengolahan data numerik

dan probabilistik. Dengan dukungan analisis statistik yang tepat, model prediksi yang dibangun dapat memberikan hasil yang lebih akurat dan dapat dipercaya dalam mengidentifikasi pola kepuasan pelanggan terhadap produk CS Float.

2.3.5. *Pattern Recognition*

Pattern Recognition atau pengenalan pola merupakan cabang ilmu dalam bidang kecerdasan buatan dan *Machine Learning* yang berfokus pada proses identifikasi, klasifikasi, serta interpretasi pola atau keteraturan yang terdapat dalam suatu data. Konsep dasarnya adalah bagaimana sistem komputer dapat belajar mengenali pola dari data yang bersifat kompleks, baik dalam bentuk angka, teks, gambar, maupun sinyal. Proses ini melibatkan tahapan utama seperti ekstraksi fitur, pemodelan pola, serta klasifikasi berdasarkan kesamaan karakteristik antar data. Dengan kata lain, *Pattern Recognition* bertujuan untuk menemukan keteraturan yang tersembunyi di dalam data dan menggunakannya untuk melakukan prediksi atau pengambilan keputusan secara otomatis. Bidang ini banyak diterapkan dalam berbagai aspek kehidupan, seperti pengenalan wajah, analisis suara, deteksi anomali, pengenalan tulisan tangan, hingga analisis perilaku konsumen.

Penerapan konsep *Pattern Recognition* memungkinkan sistem komputer untuk mengidentifikasi pola perilaku pelanggan berdasarkan data yang telah dikumpulkan, seperti preferensi rasa, frekuensi pembelian, serta persepsi terhadap harga dan kualitas produk. Melalui pendekatan ini, model klasifikasi dapat dibangun untuk mengenali kelompok pelanggan yang memiliki kesamaan karakteristik, sehingga memudahkan proses pengelompokan antara pelanggan yang puas dan tidak puas terhadap suatu produk. Dengan menggunakan algoritma seperti

Naive Bayes dan *K-Nearest Neighbor (KNN)*, data pelanggan dapat diolah menjadi pola yang bermakna dan digunakan untuk membuat prediksi secara lebih akurat. Hal ini menjadikan *Pattern Recognition* sebagai landasan penting dalam menganalisis tingkat kepuasan pelanggan secara objektif dan berbasis data, serta membantu perusahaan memahami perilaku konsumennya dengan lebih mendalam.

2.5. Metode *Naïve Bayes*

Metode *Naive Bayes* merupakan salah satu algoritma yang digunakan dalam *Machine Learning* dengan pendekatan probabilistik yang berlandaskan pada Teorema Bayes [37]. Algoritma ini bekerja dengan menghitung kemungkinan suatu data masuk ke dalam kelas tertentu berdasarkan nilai atribut atau fitur yang dimiliki. Prinsip utama dari *Naive Bayes* adalah adanya asumsi bahwa setiap atribut bersifat independen satu sama lain terhadap kelas yang menjadi target [38]. Meskipun asumsi ini terdengar sederhana, metode ini dikenal memiliki performa yang cukup baik untuk berbagai kasus klasifikasi, seperti deteksi spam, analisis sentimen, hingga prediksi perilaku pelanggan. Keunggulan utama *Naive Bayes* terletak pada efisiensi dan kecepatan dalam melakukan perhitungan, bahkan ketika jumlah data yang digunakan cukup besar. Selain itu, metode ini juga mudah diimplementasikan dan dapat memberikan hasil yang stabil meskipun data yang digunakan memiliki dimensi tinggi [39].

Metode *Naive Bayes* digunakan untuk mengelompokkan pelanggan berdasarkan tingkat kepuasan mereka terhadap produk CS Float dengan menghitung probabilitas dari setiap kategori kepuasan, seperti puas dan tidak puas [40]. Setiap variabel seperti rasa, harga, dan kemasan akan dianggap sebagai fitur

yang berkontribusi terhadap hasil klasifikasi. Model ini kemudian mempelajari pola hubungan antarfitur dengan kategori kepuasan berdasarkan data pelanggan yang telah dikumpulkan sebelumnya. Setelah model dilatih, *Naive Bayes* dapat memprediksi kategori pelanggan baru berdasarkan nilai atribut yang dimilikinya. Hasil dari penerapan metode ini dapat membantu perusahaan memahami karakteristik pelanggan yang cenderung puas serta faktor-faktor dominan yang memengaruhi tingkat kepuasan mereka. Dengan demikian, perusahaan dapat mengambil keputusan yang lebih tepat dalam meningkatkan kualitas produk dan layanan secara berkelanjutan [41].

2.6. Metode KNN

Metode *K-Nearest Neighbor* (KNN) merupakan salah satu algoritma dalam *Machine Learning* yang digunakan untuk melakukan klasifikasi maupun regresi berdasarkan prinsip kedekatan data [42]. Konsep dasar dari metode ini adalah menentukan kategori suatu data baru dengan melihat sejumlah *tetangga terdekat* (neighbor) dari data tersebut dalam ruang fitur. Nilai K menunjukkan jumlah tetangga yang digunakan dalam proses klasifikasi, dan keputusan akhir ditentukan berdasarkan mayoritas kelas dari tetangga tersebut [43]. KNN bekerja dengan menghitung jarak antara data baru dengan seluruh data dalam dataset pelatihan, biasanya menggunakan ukuran jarak Euclidean. Algoritma ini dikenal sederhana namun efektif, karena tidak memerlukan proses pelatihan model yang kompleks. Kelebihan utama KNN adalah kemampuannya beradaptasi terhadap data yang memiliki pola tidak linear, sementara kekurangannya terletak pada kebutuhan

komputasi tinggi ketika jumlah data besar serta sensitivitas terhadap data yang tidak terstandarisasi [39].

Metode KNN dimanfaatkan untuk mengklasifikasikan data pelanggan berdasarkan kemiripan karakteristik antarindividu, seperti preferensi rasa, persepsi harga, dan frekuensi pembelian produk CS Float. Setiap pelanggan baru atau data baru akan dibandingkan dengan data pelanggan lain yang sudah diketahui tingkat kepuasannya, kemudian sistem akan menentukan kategori kepuasan pelanggan tersebut berdasarkan kedekatan pola data. Pendekatan ini memungkinkan terbentuknya model analisis yang bersifat adaptif dan berbasis data nyata, di mana hasil klasifikasi menggambarkan kecenderungan tingkat kepuasan pelanggan secara akurat. Dengan demikian, metode KNN memberikan manfaat signifikan dalam memahami perilaku pelanggan, terutama dalam mengidentifikasi kelompok pelanggan yang memiliki karakteristik serupa dalam hal kepuasan terhadap produk CS Float.

2.7. Model Klasifikasi

Model klasifikasi merupakan salah satu pendekatan dalam *Machine Learning* yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang dipelajari dari data sebelumnya [44]. Model ini bekerja dengan cara mempelajari hubungan antara variabel input (fitur) dan variabel output (label), kemudian menggunakan hubungan tersebut untuk memprediksi kelas dari data baru yang belum diketahui [45]. Dalam prosesnya, data dibagi menjadi dua bagian, yaitu data latih (*training data*) untuk membangun model dan data uji (*testing data*) untuk mengevaluasi kinerja model. Model klasifikasi banyak

digunakan dalam berbagai bidang seperti deteksi spam, diagnosis medis, prediksi kredit, dan analisis perilaku pelanggan. Keunggulan utama dari model klasifikasi adalah kemampuannya untuk mengenali pola kompleks dalam data yang besar dan tidak terstruktur, serta menghasilkan keputusan berbasis probabilitas atau jarak antar data tergantung pada metode yang digunakan. Beberapa algoritma yang sering digunakan dalam klasifikasi antara lain *Naive Bayes*, *K-Nearest Neighbor (KNN)*, *Decision Tree*, dan *Support Vector Machine (SVM)*.

Penerapan model klasifikasi digunakan untuk mengidentifikasi kategori pelanggan berdasarkan tingkat kepuasan terhadap produk minuman CS Float. Data yang diperoleh dari pelanggan, seperti usia, preferensi rasa, persepsi harga, dan frekuensi pembelian, diolah menggunakan algoritma klasifikasi untuk menentukan apakah pelanggan termasuk dalam kategori puas atau tidak puas. Model ini membantu dalam menganalisis hubungan antara variabel-variabel tersebut dengan tingkat kepuasan, sehingga pola kepuasan pelanggan dapat diketahui secara lebih objektif. Dengan hasil klasifikasi tersebut, perusahaan dapat memahami karakteristik pelanggan yang memiliki tingkat kepuasan tinggi maupun rendah, serta mengambil langkah strategis dalam meningkatkan kualitas produk dan pelayanan. Pendekatan berbasis model klasifikasi juga memberikan nilai tambah karena mampu mengolah data pelanggan dalam jumlah besar dengan tingkat akurasi yang tinggi, sehingga hasil analisis yang diperoleh dapat dijadikan dasar dalam pengambilan keputusan bisnis yang lebih tepat sasaran.

2.8. Alat Bantu Program Aplikasi Orange

Orange merupakan salah satu aplikasi berbasis *open-source* yang digunakan untuk proses analisis data dan penerapan metode *Machine Learning* secara interaktif. Aplikasi ini menyediakan antarmuka visual berbentuk *workflow* yang memungkinkan pengguna melakukan tahapan pengolahan data tanpa harus menulis kode pemrograman secara manual. Melalui tampilan berbasis *drag and drop*, pengguna dapat dengan mudah menambahkan komponen seperti *data input*, *data preprocessing*, *Visualization*, *model training*, dan *evaluation* untuk menjalankan berbagai algoritma pembelajaran mesin. Orange mendukung banyak metode analisis seperti klasifikasi, regresi, klusterisasi, dan asosiasi, sehingga sangat cocok digunakan baik oleh pemula maupun peneliti profesional. Keunggulan lainnya terletak pada kemampuannya dalam menampilkan hasil analisis dalam bentuk grafik, diagram, dan laporan visual yang mudah dipahami, membuat proses eksplorasi data menjadi lebih efisien dan intuitif.

Penggunaan aplikasi Orange memberikan kemudahan dalam mengolah data pelanggan CS Float untuk menganalisis tingkat kepuasan konsumen menggunakan metode *Naive Bayes* dan *K-Nearest Neighbor (KNN)*. Dengan bantuan fitur *Data Table* dan *File Widget*, data mentah pelanggan dapat dimasukkan, dibersihkan, serta diproses untuk menghasilkan variabel-variabel penting yang berpengaruh terhadap kepuasan. Selanjutnya, algoritma *Naive Bayes* dan *KNN* dapat diterapkan menggunakan *Classification Widget*, sementara hasil akurasi model dapat dilihat melalui *Test & Score Widget*. Melalui visualisasi hasil seperti *Confusion Matrix* dan *ROC Curve*, tingkat keakuratan prediksi dapat dibandingkan dengan mudah untuk menentukan metode terbaik. Pemanfaatan Orange tidak hanya mempercepat proses

analisis data, tetapi juga meningkatkan keakuratan hasil penelitian karena setiap tahapan analisis dapat dikontrol secara sistematis dan transparan.

2.9. Kelebihan dan Kekurangan Metode *Naive Bayes* dan KNN

Metode *Naive Bayes* dan *K-Nearest Neighbor* (KNN) merupakan dua algoritma populer dalam bidang *Machine Learning* yang digunakan untuk tugas klasifikasi. *Naive Bayes* memiliki kelebihan dalam hal kecepatan komputasi, kesederhanaan, dan efektivitasnya dalam menangani data dengan jumlah besar. Algoritma ini bekerja berdasarkan prinsip probabilitas Bayes, dengan asumsi bahwa setiap atribut bersifat independen satu sama lain. Meskipun asumsi ini jarang sepenuhnya benar di dunia nyata, *Naive Bayes* tetap mampu memberikan hasil prediksi yang cukup baik, terutama ketika data memiliki pola distribusi yang jelas. Namun, kelemahannya terletak pada sensitivitas terhadap atribut yang saling bergantung dan kemungkinan penurunan akurasi jika data tidak memenuhi asumsi independensi. Sementara itu, *K-Nearest Neighbor* (KNN) menawarkan pendekatan yang berbeda dengan mengklasifikasikan data berdasarkan kedekatan jarak terhadap data lain yang sudah diketahui kategorinya. Kelebihan utama KNN adalah kemampuannya menghasilkan hasil klasifikasi yang intuitif dan akurat pada dataset dengan struktur pola yang kompleks.

Kedua metode tersebut memiliki peran penting dalam upaya menganalisis dan memprediksi tingkat kepuasan pelanggan terhadap suatu produk. *Naive Bayes* dapat digunakan untuk mengolah data pelanggan dengan lebih cepat dan efisien, terutama jika jumlah data yang dianalisis cukup besar. Algoritma ini mampu memberikan hasil prediksi dengan tingkat akurasi yang memadai meskipun struktur

data sederhana. Di sisi lain, KNN memberikan pendekatan yang lebih fleksibel dan adaptif terhadap data yang tidak memiliki pola distribusi linier. Dengan mempertimbangkan nilai-nilai atribut seperti persepsi terhadap rasa, harga, dan kemasan produk, KNN dapat mengelompokkan pelanggan berdasarkan kemiripan karakteristik mereka. Perbandingan antara kedua metode ini memungkinkan diperolehnya pemahaman yang lebih komprehensif mengenai performa algoritma klasifikasi dalam menilai tingkat kepuasan pelanggan.

2.10. Evaluasi Model *Naïve Bayes* dan KNN

Metode *Naive Bayes* dan *K-Nearest Neighbor* (KNN) merupakan dua algoritma populer yang sering digunakan dalam pemodelan klasifikasi pada bidang *Machine Learning*. *Naive Bayes* bekerja berdasarkan prinsip teorema Bayes yang menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan hubungan antarfitur yang diasumsikan saling independen. Kelebihan utama metode ini terletak pada kecepatan proses perhitungannya serta kemampuannya dalam menangani data berukuran besar dengan tingkat akurasi yang cukup baik. Sementara itu, KNN merupakan metode non-parametrik yang mengklasifikasikan data baru berdasarkan kedekatan jarak terhadap sejumlah data tetangga terdekat yang telah diketahui kelasnya. Akurasi KNN sangat bergantung pada jumlah tetangga (k) yang digunakan dan ukuran data pelatihan. Perbandingan antara kedua metode ini sering dilakukan karena keduanya memiliki pendekatan yang berbeda: *Naive Bayes* lebih mengandalkan model probabilistik, sedangkan KNN bergantung pada kedekatan spasial antar data. Evaluasi terhadap kedua model ini biasanya

dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score untuk menilai kinerja model dalam memprediksi kelas dengan benar.

Evaluasi kedua metode tersebut diterapkan untuk mengetahui model mana yang mampu memberikan hasil klasifikasi terbaik dalam mengidentifikasi tingkat kepuasan pelanggan berdasarkan data yang tersedia. *Naive Bayes* memberikan hasil prediksi cepat dan efisien ketika data memiliki distribusi yang cukup jelas antar kelas, sedangkan KNN cenderung lebih akurat ketika pola data memiliki keterkaitan jarak yang kuat antara satu pelanggan dengan pelanggan lainnya. Hasil evaluasi terhadap model akan memperlihatkan sejauh mana algoritma mampu mengenali pola dari atribut pelanggan, seperti persepsi terhadap rasa, harga, kemasan, dan frekuensi pembelian. Nilai akurasi yang dihasilkan menjadi dasar dalam menentukan model yang paling sesuai digunakan untuk memprediksi kategori pelanggan yang puas dan tidak puas.

2.11. Kelebihan Penelitian

Penerapan metode *Machine Learning* dalam analisis tingkat kepuasan pelanggan memiliki sejumlah kelebihan dibandingkan dengan pendekatan tradisional. Salah satu keunggulannya adalah kemampuan untuk mengolah data dalam jumlah besar secara efisien dan menghasilkan prediksi yang objektif berdasarkan pola yang terbentuk dari data historis. Metode *Naive Bayes* dan *K-Nearest Neighbor (KNN)* sebagai model klasifikasi memiliki karakteristik yang berbeda namun saling melengkapi. *Naive Bayes* unggul dalam kesederhanaan, kecepatan komputasi, serta efektivitasnya dalam menangani data dengan banyak

variabel. Sementara itu, *KNN* memiliki kelebihan dalam hal fleksibilitas dan ketepatan prediksi karena menentukan kategori berdasarkan kemiripan antar data.

Pemanfaatan kedua metode tersebut memberikan kelebihan yang signifikan dalam mengidentifikasi pola kepuasan pelanggan terhadap produk minuman CS Float. Analisis yang dilakukan tidak hanya menghasilkan klasifikasi antara pelanggan puas dan tidak puas, tetapi juga mampu menyoroti faktor-faktor dominan yang memengaruhi tingkat kepuasan, seperti rasa, harga, kemasan, dan frekuensi pembelian.

2.12. Penelitian Terdahulu

Pada penelitian ini terdapat penelitian terdahulu guna untuk memperkuat penelitian yang dilakukan. Untuk penelitian terdahulu penulis paparkan dalam bentuk tabel dibawah ini yaitu sebagai berikut.

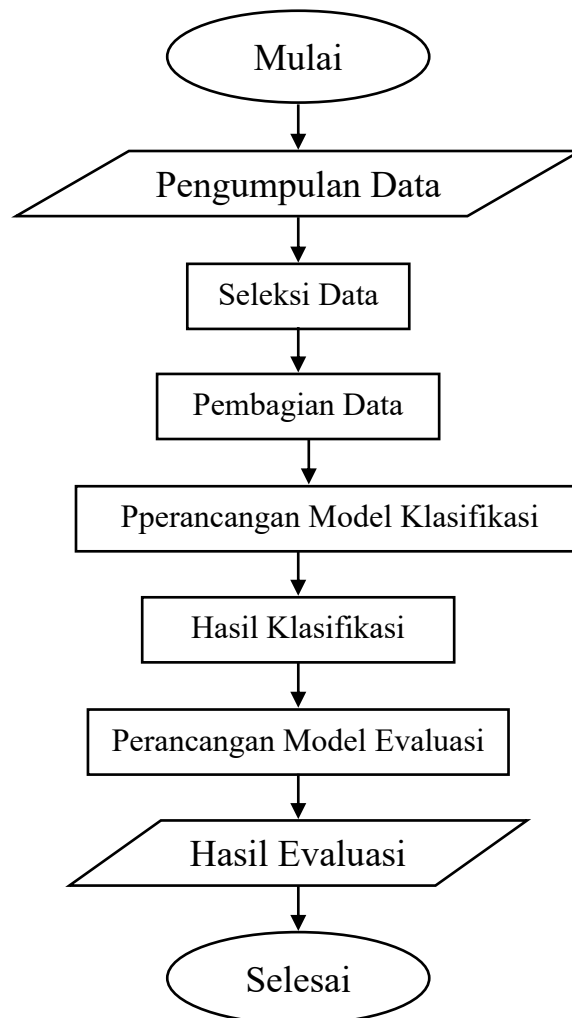
Tabel 2. 1. Penelitian Terdahulu

No	Peneliti	Judul Penelitian	Tahun	Metode dan Data	Hasil
1	Fathia Alisha Fauzia 1), Kannisa Adjani 2), Christina Juliane 3)	<i>Comparison Of K-N Earest Neighbor and Naïve Bayes Algorithms for Prediction of Aptikom Membership Activity Extension In 2023</i>	2023	K-Nearest Neighbor (K-NN), Naïve Bayes dan Data keanggotaan APTIKOM periode 2012-2022	Penelitian menunjukkan bahwa K-NN dan Naïve Bayes mampu memprediksi pembaruan keanggotaan APTIKOM dengan akurasi tinggi, masing-masing 94,00% dan 91,35%. K-NN terbukti sedikit lebih unggul dalam memberikan prediksi yang lebih tepat dibandingkan Naïve Bayes [46].
2	Istiadi1), Aviv Yuniar Rahman2), Alif Dio Raka Wisnu3)	<i>Identifcation of Tempe Fermentation Maturity Using Principal Component Analysis and</i>	2023	K-Nearest Neighbor (KNN) dengan dan tanpa PCA	Hasil penelitian menunjukkan bahwa KNN dengan PCA menghasilkan akurasi sedikit lebih rendah (80,63%) dibanding KNN tanpa PCA (81,67%), tetapi memberikan waktu

		<i>KNearest Neighbor</i>			pemrosesan yang lebih cepat. Perbandingan ini menegaskan bahwa PCA tetap berguna ketika efisiensi waktu menjadi prioritas [47].
3	Deci Irmayani 1)*, Fransisco Alexander Sinaga 2), Masrizal3)	<i>Analysis of the Level of Public Satisfaction on the TikTok Application as an E-Commerce</i>	2023	K-Nearest Neighbor (KNN), Decision Tree dan data respon masyarakat	Sebanyak 96,74% responden menyatakan puas berbelanja melalui TikTok, sedangkan 3,25% lainnya tidak puas. Temuan ini menunjukkan bahwa TikTok dinilai efektif sebagai platform E-commerce berkat pengalaman belanja yang memuaskan [48].
4	Jepi Supriyanto1*, Debby Alita2, Auliya Rahman Isnain3	Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring	2023	Metode K-Nearest Neighbor dan Data yang digunakan 1825 tweet berbahasa Indonesia.	Penelitian ini menggunakan algoritma KNN untuk analisis sentimen 1825 tweet terkait kebijakan Pembelajaran Daring, yang dikumpulkan dengan Tweepy pada periode 1 Februari–30 September 2020. Dengan pembobotan TF-IDF dan dua kelas sentimen (positif dan negatif), hasil terbaik diperoleh pada K = 10 dengan akurasi 84,65%, presisi 87%, recall 86%, f-measure 87%, dan error rate 0,12%. Secara keseluruhan, opini publik terhadap Pembelajaran Daring cenderung positif [49].
5	Syafrizal1*, M. Afdal2, Rice Novita3	Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor	2024	Metode Naïve Bayes Classifier dan KNN, untuk Datanya 3000 barus ulasan Pengguna PLN Mobile	Penelitian ini menganalisis 3000 ulasan pengguna PLN Mobile dan menunjukkan mayoritas sentimen positif. Model NBC memberikan hasil terbaik dengan akurasi 77,69% dibanding KNN. Ulasan menunjukkan aplikasi cukup membantu pengguna, namun masalah pembayaran token masih sering dikeluhkan.

2.13. Kerangka Kerja Penelitian

Pada penelitian ini terdapat kerangka kerja penelitian, hal ini dilakukan agar dapat mempermudah proses analisis data dan proses dapat dilakukan lebih terstruktur. Untuk kerangka kerja penelitiannya yaitu sebagai berikut.



Gambar 2. 2. Kerangka Kerja Penelitian

Pada kerangka kerja penelitian diatas terdapat penjelasan untuk masing masing poin yaitu sebagai berikut.

1. Pengumpulan Data

Tahap ini merupakan proses mengumpulkan data yang dibutuhkan untuk penelitian. Data dapat diperoleh dari berbagai sumber, seperti survei,

kuesioner, *Database* publik, atau data internal perusahaan. Tujuannya adalah memperoleh informasi yang relevan dan representatif agar proses analisis berikutnya dapat menghasilkan hasil yang valid dan akurat.

2. Seleksi Data

Pada tahap ini dilakukan proses pemilihan data yang dianggap penting dan berpengaruh terhadap penelitian. Data yang tidak relevan, tidak lengkap, atau mengandung kesalahan akan dibersihkan agar tidak memengaruhi hasil analisis. Seleksi data membantu meningkatkan kualitas dataset serta mengurangi kompleksitas model yang akan dibangun.

3. Pembagian Data

Data yang telah diseleksi kemudian dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun dan melatih model klasifikasi, sedangkan data uji digunakan untuk mengukur seberapa baik model mampu memprediksi data baru. Pembagian ini penting agar model dapat dievaluasi secara objektif.

4. Perancangan Model Klasifikasi

Tahap ini melibatkan pembuatan model menggunakan algoritma *Machine Learning*, seperti *Naive Bayes*, *K-Nearest Neighbor* (KNN), atau metode lainnya. Model ini dibangun berdasarkan data latih untuk mengenali pola dan hubungan antar variabel yang memengaruhi hasil prediksi.

5. Hasil Klasifikasi

Setelah model dibangun, hasil klasifikasi akan menunjukkan kategori atau label dari data yang dianalisis, misalnya “puas” atau “tidak puas”. Hasil ini

memberikan gambaran awal mengenai kemampuan model dalam mengelompokkan data sesuai dengan karakteristik yang telah dipelajari.

6. Perancangan Model Evaluasi

Tahap ini dilakukan untuk menilai performa model klasifikasi yang telah dibuat. Evaluasi dilakukan dengan menggunakan metrik seperti akurasi, presisi, recall, atau f1-score, untuk melihat sejauh mana model dapat memprediksi dengan benar.

7. Hasil Evaluasi

Tahap terakhir adalah penyajian hasil evaluasi model dalam bentuk angka atau grafik yang menunjukkan tingkat keberhasilan metode yang digunakan. Hasil ini membantu peneliti menentukan apakah model sudah cukup baik atau perlu dilakukan penyempurnaan agar hasil prediksi menjadi lebih optimal.