BAB II
LANDASAN TEORI

2.1. Tingkat Kelarisan Produk

Tingkat kelarisan produk merupakan ukuran yang menunjukkan seberapa besar suatu produk laris dan dibeli oleh konsumen dalam periode tertentu. Ini sangat penting dalam bisnis, terutama untuk memahami pola permintaan pasar dan mengelola stok secara efisien. Faktor-faktor yang mempengaruhi kelarisan produk antara lain harga, merek, kualitas, kelengkapan produk, dan promosi.

Beberapa penelitian terdahulu mengungkapkan bahwa produk dengan harga yang kompetitif, kualitas yang baik, serta kelengkapan dan fitur yang sesuai dengan kebutuhan konsumen cenderung memiliki tingkat kelarisan yang lebih tinggi. Penetapan strategi pemasaran yang tepat, seperti diskon, bundling produk, atau promosi melalui media sosial, juga berperan dalam meningkatkan kelarisan produk (Widia & Nofirda, 2022; Sudirga, 2017)[27].

2.2. Manfaat Analisis Tingkat Kelarisan Produk

Analisis tingkat kelarisan produk memberikan berbagai manfaat strategis bagi bisnis yang signifikan dalam mendukung pengelolaan operasional dan pengambilan keputusan yang lebih matang, yaitu sebagai berikut.

1. Pengelolaan Stok yang Efisien: Analisis tingkat kelarisan memungkinkan bisnis mengidentifikasi produk yang Laris, dan Tidak Laris. Dengan informasi ini, tingkat persediaan dapat disesuaikan secara optimal sehingga

risiko overstock maupun out-of-stock dapat dikurangi, biaya operasional dapat ditekan, dan tingkat perputaran stok dapat dijaga dengan lebih baik.

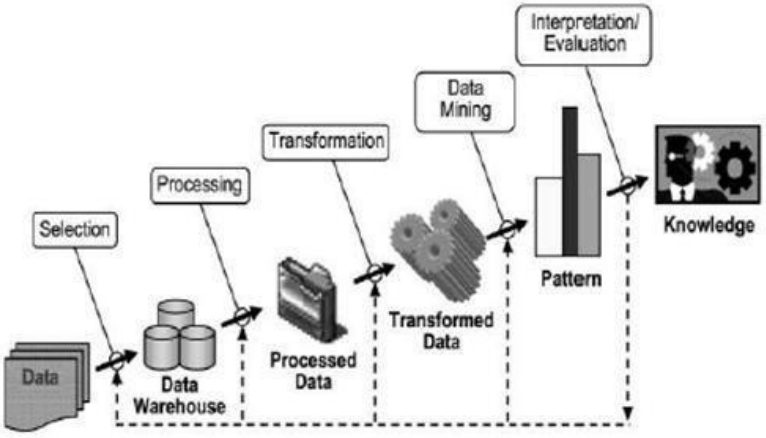
2. Strategi Pemasaran yang Lebih Tepat: Dengan data tingkat kelarisan produk, bisnis dapat merumuskan strategi promosi dan pemasaran yang lebih fokus. Produk dengan penjualan tinggi dapat dijadikan prioritas promosi untuk memaksimalkan omzet, sedangkan produk dengan tingkat kelarisan rendah dapat dijadikan fokus promosi khusus atau bundling untuk mempercepat pergerakan stok.
3. Peningkatan Pengalaman Pelanggan: Analisis ini juga membantu bisnis memahami kebutuhan dan pola preferensi pelanggan dengan lebih mendalam. Dengan begitu, bisnis dapat menawarkan produk yang sesuai dengan kebutuhan dan selera konsumen, meningkatkan pengalaman berbelanja, dan memperkokoh loyalitas pelanggan.
4. Optimalisasi Keuntungan: Dengan mengidentifikasi produk yang Tidak Laris, bisnis dapat menyusun langkah strategis guna mengurangi tingkat risiko dan kerugian, seperti memberikan diskon khusus, membuat paket bundling, atau menghentikan penjualan produk yang tidak memberikan nilai signifikan. Hal ini memungkinkan bisnis mengalokasikan sumber daya dengan lebih efisien dan fokus pada produk dengan potensi laba yang lebih tinggi.

2.3. Knowledge Discovery in Database

Knowledge Discovery in Databases (KDD) adalah suatu proses yang bertujuan untuk menemukan pola, pengetahuan, atau informasi yang berguna dari

kumpulan data yang besar. Dalam era digital saat ini, data dihasilkan dalam jumlah yang sangat besar dan sering kali sulit untuk dipahami secara langsung[1]. Oleh karena itu, KDD sangat penting untuk membantu mengubah data mentah menjadi informasi yang bermakna, yang dapat digunakan untuk pengambilan keputusan yang lebih baik[2].

KDD bukan hanya sekadar satu langkah kerja, tetapi merupakan serangkaian tahapan yang terintegrasi. Proses ini sering digunakan di berbagai bidang seperti bisnis, kesehatan, pendidikan, dan teknologi, karena dapat memberikan wawasan yang membantu meningkatkan kinerja dan efisiensi.



Gambar 2. 1. Knowledge Discovery in Database

Gambar 2.1 Tahap-tahap dalam Knowledge Discovery in Databases (KDD) [4] yaitu sebagai berikut.

1. Seleksi Data (Data Selection), Langkah pertama dalam KDD adalah pemilihan data yang relevan dari kumpulan data besar. Data yang dipilih harus memiliki nilai atau potensi untuk memberikan informasi yang berguna. Misalnya, jika analisis berfokus pada penjualan produk, maka data

yang relevan mencakup informasi tentang transaksi penjualan, pelanggan, dan produk yang terlibat[5].

2. Pemprosesan (Data Processed), Pada tahap ini, data yang telah dipilih akan diperiksa untuk memastikan kualitasnya. Proses pembersihan data mencakup identifikasi dan penghapusan kesalahan, seperti data yang hilang, duplikat, atau tidak konsisten. Pembersihan data sangat penting untuk memastikan bahwa analisis selanjutnya menghasilkan informasi yang akurat dan andal[6].
3. Transformasi Data (Data Transformation), Setelah data bersih, tahap berikutnya adalah transformasi data. Proses ini melibatkan pengubahan data ke dalam format yang lebih sesuai untuk analisis selanjutnya. Transformasi data dapat meliputi penggabungan kolom, normalisasi data, atau penyederhanaan atribut yang kompleks. Dalam penelitian Nurohmah (2023), misalnya, Principal Component Analysis (PCA) digunakan untuk mengurangi dimensi data, sehingga meningkatkan efisiensi algoritma k-means clustering dalam menganalisis tingkat kemiskinan di Jawa Barat[7].
4. Data Mining, Data mining adalah langkah utama dalam proses KDD. Pada tahap ini, berbagai teknik seperti algoritma, statistik, dan pembelajaran mesin digunakan untuk menemukan pola atau hubungan yang tersembunyi dalam data. Teknik yang umum digunakan termasuk clustering (pengelompokan), classification (klasifikasi), dan association rules (aturan asosiasi). Sebagai contoh, dalam penelitian Maulana (2024), algoritma k-means clustering diterapkan untuk menganalisis transaksi penjualan hijab,

menghasilkan wawasan yang berguna bagi pengambilan keputusan dalam bisnis ritel[8].

5. Evaluasi Pola (Pattern Evaluation), Pola yang ditemukan selama tahap data mining kemudian dievaluasi untuk memastikan relevansi dan kegunaannya. Evaluasi ini melibatkan pengukuran efektivitas pola yang ditemukan, menggunakan metrik seperti precision, recall, dan F1-score. Dalam penelitian Widaningsih (2022), yang menggunakan algoritma K- Nearest Neighbor (KNN) untuk memprediksi siswa berprestasi, penting untuk mengevaluasi akurasi model guna memastikan bahwa prediksi yang dihasilkan dapat dipercaya[9].

2.3.1. Data Mining

Data mining merupakan tahap paling penting dan inti dalam proses Knowledge Discovery in Databases (KDD), di mana berbagai metode dan algoritma digunakan untuk menemukan pola atau hubungan yang tersembunyi di dalam data. Pada tahap ini, data yang sudah melewati proses pembersihan dan transformasi diolah dengan memanfaatkan berbagai teknik, mulai dari algoritma machine learning hingga metode statistik, guna mendapatkan pola dan aturan yang dapat dijadikan landasan pengambilan keputusan. Beberapa contoh teknik yang paling umum digunakan dalam data mining adalah clustering (pengelompokan), classification (klasifikasi), dan association rules (aturan asosiasi). Clustering digunakan untuk mengelompokkan data yang memiliki pola atau karakteristik serupa, classification digunakan untuk memprediksi atau mengelompokkan data ke dalam kelas tertentu, sedangkan association rules digunakan untuk menemukan

pola hubungan atau keterkaitan antaratribut dalam sebuah transaksi atau kumpulan data. Melalui penerapan teknik-teknik tersebut, berbagai pola yang sebelumnya sulit terlihat dapat diungkapkan dan dianalisis lebih dalam guna menghasilkan nilai dan pengetahuan dari data yang dimiliki.

Contoh penerapan data mining dapat dilihat dalam penelitian Maulana (2024), yang mengimplementasikan algoritma k-means clustering untuk menganalisis pola transaksi penjualan produk hijab. Dengan memanfaatkan algoritma ini, data transaksi dapat dikelompokkan ke dalam beberapa segmen berdasarkan pola dan karakteristik pembelian dari para pelanggan. Hasil dari analisis tersebut memungkinkan pihak bisnis untuk memahami kebutuhan dan preferensi pelanggan dengan lebih baik, termasuk pola pembelian produk tertentu dan waktu-waktu transaksi yang paling aktif. Dengan begitu, pihak bisnis dapat merumuskan strategi yang lebih tepat, mulai dari pengelolaan stok, penentuan promosi, hingga pengembangan produk sesuai kebutuhan pasar. Hal ini menjadikan data mining bukan hanya sebuah proses teknis semata, tetapi juga alat strategis yang dapat memberikan nilai tambah signifikan bagi pengambilan keputusan dan pengembangan bisnis di era digital.

1. Penerapan Data Mining dalam Berbagai Bidang

Penerapan data mining telah merambah berbagai bidang dan memberikan manfaat yang signifikan bagi pengambilan keputusan maupun pengembangan strategi kerja. Dalam bidang bisnis dan pemasaran, data mining digunakan untuk menganalisis pola pembelian pelanggan guna memahami kebutuhan dan perilaku konsumen dengan lebih mendalam. Misalnya, dengan memanfaatkan algoritma

clustering dan association rules, sebuah perusahaan dapat mengidentifikasi pola pembelian produk yang sering muncul bersama (market basket analysis), atau mengelompokkan pelanggan berdasarkan pola transaksi guna mengoptimalkan promosi dan strategi penjualan. Sementara itu, dalam bidang keuangan, data mining digunakan untuk mendeteksi pola penipuan, menganalisis risiko kredit, hingga memprediksi tren nilai saham berdasarkan pola data historis, yang dapat membantu pihak perbankan dan investor dalam membuat keputusan yang lebih akurat dan efisien.

Selain itu, penerapan data mining juga terlihat nyata dalam bidang kesehatan, pendidikan, hingga pemerintahan. Dalam bidang kesehatan, data mining digunakan untuk menganalisis pola dan pola terkait riwayat penyakit pasien guna mendukung pengambilan keputusan diagnosis dan pengembangan metode perawatan yang lebih efektif. Di bidang pendidikan, data mining dapat digunakan untuk mengidentifikasi pola belajar siswa, memprediksi tingkat keberhasilan akademik, dan merumuskan metode pengajaran yang lebih adaptif sesuai kebutuhan masing-masing peserta didik. Bahkan dalam pemerintahan, teknologi ini dapat digunakan untuk memetakan pola kriminalitas, memantau tren pelayanan publik, hingga menganalisis pola mobilitas dan kebutuhan masyarakat guna mengoptimalkan pembangunan daerah. Dengan berbagai penerapan ini, data mining telah tumbuh menjadi teknologi kunci yang dapat memberikan nilai signifikan bagi berbagai disiplin ilmu dan kebutuhan nyata di lapangan.

2. Algoritma dalam Data Mining

Algoritma dalam data mining merupakan serangkaian langkah atau aturan sistematis yang digunakan untuk memproses data dan mengekstrak pola atau informasi berharga dari kumpulan data yang besar. Berbagai jenis algoritma digunakan sesuai dengan kebutuhan dan tujuan analisis, mulai dari klasifikasi, clustering, regresi, hingga asosiasi. Contoh algoritma yang sering digunakan dalam klasifikasi adalah Naive Bayes, Decision Tree, dan Support Vector Machine, yang dapat digunakan untuk memprediksi suatu nilai atau kelas dari data tertentu berdasarkan pola yang ditemukan dari data pelatihan. Ada juga algoritma clustering, seperti K-means dan Hierarchical Clustering, yang digunakan untuk mengelompokkan data berdasarkan kesamaan atau pola tertentu, tanpa memerlukan label atau target yang spesifik. Masing-masing algoritma ini memiliki kelebihan dan kelemahan tersendiri, sehingga pemilihan algoritma yang digunakan harus disesuaikan dengan jenis data dan kebutuhan analisis.

Selain itu, algoritma dalam data mining juga dapat digunakan untuk menemukan pola hubungan atau aturan antaratribut dalam data, seperti dengan algoritma Apriori dan FP-Growth, yang digunakan untuk analisis pola asosiasi. Dengan penerapan algoritma ini, berbagai pola transaksi atau pola perilaku dapat ditemukan, yang dapat digunakan untuk kebutuhan bisnis, penelitian, maupun pengembangan sistem. Berkat algoritma-algoritma tersebut, data mining memungkinkan penggalian nilai dari data yang awalnya tidak terlihat, sehingga dapat digunakan sebagai landasan dalam pengambilan keputusan yang lebih efektif dan efisien. Dengan terus berkembangnya teknologi dan metode analisis, algoritma

dalam data mining juga terus berevolusi dan memungkinkan penerapan yang semakin kompleks dan adaptif sesuai kebutuhan berbagai bidang.

2.3.2. Database dan Data Processing

Database merupakan kumpulan data yang disimpan secara sistematis dan terstruktur sehingga dapat dengan mudah diakses, dikelola, dan diubah sesuai kebutuhan. Dalam sebuah sistem, database berfungsi sebagai tempat penyimpanan berbagai informasi yang digunakan oleh aplikasi atau perangkat lunak tertentu. Dengan adanya database, data dapat diorganisasi berdasarkan pola atau atribut tertentu sehingga mempermudah proses pengumpulan, pencarian, dan pembaruan data. Berbagai teknologi basis data seperti MySQL, PostgreSQL, maupun MongoDB memungkinkan pengelolaan data dalam jumlah besar dengan tingkat keamanan dan integritas yang tinggi, menjadikan database sebagai pondasi dari berbagai sistem bisnis dan teknologi informasi.

Data processing atau pemrosesan data sendiri merupakan tahap lanjutan dari pengumpulan data yang bertujuan untuk mengubah data mentah menjadi informasi yang dapat digunakan dalam pengambilan keputusan. Proses ini meliputi berbagai aktivitas, mulai dari validasi data, penggabungan, pembersihan, hingga transformasi data agar dapat dianalisis lebih mendalam. Dengan data processing yang baik, data dapat digunakan untuk menghasilkan laporan, pola, maupun gambaran tren yang dapat dijadikan bahan pertimbangan bagi manajemen atau pihak terkait. Keterkaitan yang erat antara database dan data processing menjadikan keduanya sebagai elemen vital dalam sistem informasi, memungkinkan sebuah bisnis atau organisasi memaksimalkan nilai dari data yang dimilikinya.

2.3.3. Visualization

Visualisasi atau Visualization merupakan salah satu langkah penting dalam analisis data yang bertujuan untuk mengubah angka dan informasi kompleks menjadi bentuk visual yang mudah dipahami, seperti grafik, diagram, maupun peta. Dengan visualisasi, pola, tren, dan hubungan antarvariabel dapat terlihat lebih jelas, memungkinkan para pengambil keputusan untuk memahami data dengan lebih cepat dan menyeluruh. Berbagai metode visualisasi digunakan, mulai dari diagram batang, garis, area, hingga peta panas (heatmap), yang masing-masing dapat digunakan sesuai dengan kebutuhan dan jenis data yang dianalisis.

Selain itu, visualisasi juga berperan sebagai jembatan komunikasi antaranggota tim maupun dengan pihak eksternal yang membutuhkan laporan atau gambaran dari data yang dianalisis. Dalam penerapannya, visualisasi dapat digunakan untuk mengevaluasi kinerja bisnis, memantau perubahan pola perilaku pelanggan, atau mengidentifikasi anomali dari data yang dikumpulkan. Dengan kemampuan memberikan gambaran intuitif dari data yang kompleks, visualisasi membantu membuat analisis lebih efisien dan memungkinkan pengambil keputusan merumuskan langkah strategis yang lebih tepat berdasarkan fakta dan pola yang terlihat dari data tersebut.

2.3.4. Statistik

Statistik merupakan ilmu yang mempelajari metode pengumpulan, pengolahan, analisis, dan interpretasi data sehingga dapat dijadikan sebagai dasar pengambilan keputusan yang lebih terukur dan relevan. Dalam berbagai bidang, mulai dari bisnis, kesehatan, teknologi, hingga pemerintahan, statistik digunakan

untuk memahami pola, tren, dan pola hubungan yang terdapat dalam data. Dengan memanfaatkan metode statistika, data mentah dapat diubah menjadi informasi yang bermakna, memungkinkan para pengambil keputusan untuk membuat langkah yang lebih tepat berdasarkan bukti dan fakta yang dapat diukur. Hal ini menjadikan statistik sebagai alat yang sangat diperlukan dalam perencanaan dan pengawasan berbagai kegiatan.

Selain itu, statistik juga berperan dalam proses penelitian ilmiah dengan menyediakan metode untuk menguji hipotesis, mengukur tingkat kesalahan, dan menarik kesimpulan dari sampel data yang dianalisis. Berbagai teknik statistik, mulai dari deskriptif hingga inferensial, memungkinkan peneliti dan praktisi untuk memahami pola dalam data dengan tingkat keyakinan tertentu. Dengan memanfaatkan statistik, berbagai pihak dapat membuat prediksi yang lebih akurat, mengevaluasi kinerja, serta mengidentifikasi peluang dan risiko. Dengan kata lain, statistik memberikan landasan kuat bagi pengambilan keputusan yang berdasarkan data dan dapat dipertanggungjawabkan.

2.3.5. Pattern Recognition

Pattern recognition atau pengenalan pola merupakan salah satu cabang dari kecerdasan buatan yang berfokus pada kemampuan sistem untuk mengidentifikasi pola atau struktur tertentu dari suatu data. Pola ini dapat berupa bentuk, angka, teks, maupun pola lain yang dapat diekstrak dari berbagai sumber data. Tujuan utama dari pattern recognition ialah memungkinkan komputer untuk memahami, mengelompokkan, dan memaknai pola dari data yang belum pernah dijumpai sebelumnya. Berbagai metode digunakan dalam penerapan pattern recognition,

mulai dari metode statistika hingga algoritma pembelajaran mesin, guna membuat sistem dapat melakukan klasifikasi, pengelompokan, maupun deteksi pola dengan tingkat akurasi tinggi.

Pattern recognition memiliki peranan yang sangat signifikan dalam berbagai bidang, seperti pengenalan wajah, pengenalan suara, dan analisis citra. Dengan memanfaatkan teknologi ini, berbagai permasalahan dapat diselesaikan secara efisien dan terotomasi, mulai dari pengawasan keamanan hingga analisis pola perilaku konsumen. Keberhasilan pattern recognition juga bergantung pada kualitas dan jumlah data yang digunakan sebagai bahan pelatihan model, serta kemampuan algoritma dalam mempelajari pola dari data tersebut. Seiring dengan perkembangan teknologi, penerapan pattern recognition terus meluas dan menjadi salah satu komponen kunci dalam sistem kecerdasan buatan masa depan.

2.4. Metode *Naïve Bayes*

Metode Naive Bayes merupakan salah satu algoritma klasifikasi dalam data mining yang bekerja berdasarkan teorema Bayes dengan asumsi bahwa setiap atribut yang digunakan dalam proses klasifikasi saling independen satu sama lain. Metode ini menghitung nilai probabilitas dari masing-masing kelas target berdasarkan nilai atribut yang tersedia, kemudian memilih kelas dengan nilai probabilitas tertinggi sebagai hasil prediksi. Naive Bayes banyak digunakan karena algoritmanya sederhana, cepat, dan efisien, terutama untuk data dengan jumlah atribut yang cukup banyak. Meskipun asumsi independensi atribut terkesan menyederhanakan keadaan yang kompleks, metode ini tetap dapat memberikan tingkat akurasi yang baik dalam berbagai jenis klasifikasi.

Penggunaan metode Naive Bayes dapat diterapkan pada berbagai studi yang memerlukan prediksi atau klasifikasi pola data. Salah satu contoh penerapannya ialah untuk menganalisis pola penjualan produk berdasarkan atribut-atribut tertentu, seperti merek, jumlah penjualan, harga, maupun kualitas produk. Dengan memanfaatkan nilai probabilitas dari masing-masing atribut, metode ini dapat mengklasifikasi suatu produk sebagai laris atau Tidak Laris berdasarkan pola data yang terdapat dalam transaksi penjualan terdahulu.

Selain itu, metode Naive Bayes juga dapat digunakan untuk mengevaluasi pola dari berbagai atribut yang memengaruhi suatu keputusan bisnis. Dengan menghitung nilai probabilitas dari atribut-atribut terkait, metode ini dapat membantu pelaku usaha memahami pola-pola atau tren dari data yang dikumpulkan, seperti pola pembelian dari pelanggan dengan nilai jumlah penjualan tertentu atau pola atribut produk yang lebih diminati. Hal ini membuat metode Naive Bayes dapat dijadikan sebagai salah satu alat pendukung pengambilan keputusan yang efisien dan dapat diandalkan.

Pada akhirnya, nilai dari penerapan metode Naive Bayes terlihat dari kemampuannya memberikan tingkat akurasi yang dapat diukur dan dievaluasi dengan metode-metode standar, seperti confusion matrix. Hasil dari evaluasi ini dapat digunakan sebagai landasan bagi pengembang bisnis maupun akademisi dalam memahami pola data dan membuat prediksi yang lebih tepat. Dengan implementasi yang tepat dan pengolahan data yang baik, metode ini dapat digunakan sebagai alat yang signifikan untuk memetakan pola dan pola hubungan yang tersembunyi dari berbagai atribut yang tersedia.

2.5. Keunggulan dan Kekurangan Metode Naive Bayes

Metode Naive Bayes merupakan salah satu algoritma klasifikasi yang paling populer digunakan dalam berbagai disiplin data mining dan machine learning. Keunggulan utamanya terletak pada kesederhanaannya dalam implementasi dan efisiensinya dalam mengolah data, bahkan ketika jumlah data dan atribut yang digunakan relatif besar. Dengan asumsi bahwa setiap atribut dalam data saling independen, algoritma ini dapat menghitung nilai probabilitas dengan cepat dan memetakan pola dari data dengan akurat. Keunggulan ini membuat Naive Bayes menjadi metode yang cocok bagi para peneliti maupun praktisi yang membutuhkan model klasifikasi yang dapat dikembangkan dan digunakan dengan waktu yang relatif singkat.

Selain itu, Naive Bayes juga sangat efisien dari sisi kebutuhan daya komputasi dan memori. Berbeda dengan metode lain yang membutuhkan pelatihan panjang atau memerlukan struktur data kompleks, algoritma ini dapat bekerja dengan baik bahkan pada perangkat dengan spesifikasi standar. Model yang dihasilkan juga tidak memerlukan ukuran penyimpanan yang terlalu besar, sehingga dapat digunakan untuk kebutuhan sistem dengan keterbatasan daya maupun perangkat mobile. Efisiensi ini membuat metode ini menjadi pilihan yang ideal untuk berbagai penerapan skala kecil hingga besar.

Kelebihan lain dari metode Naive Bayes ialah fleksibilitasnya dalam menangani berbagai jenis data, mulai dari data numerik hingga data kategorikal. Dengan penyesuaian tertentu, metode ini dapat digunakan untuk klasifikasi teks, prediksi pola pembelian, bahkan analisis pola perilaku konsumen. Model ini juga

dapat digunakan untuk berbagai kebutuhan lain, seperti sistem rekomendasi, deteksi spam, maupun analisis risiko. Fleksibilitas tersebut membuat metode ini dapat diaplikasikan dalam berbagai konteks dan kebutuhan dengan tingkat keberhasilan yang relatif tinggi.

Naive Bayes juga dapat digunakan dengan jumlah data pelatihan yang kecil, tetapi tetap menghasilkan tingkat akurasi yang memadai. Berbeda dengan metode lain yang membutuhkan jumlah data pelatihan yang sangat besar agar dapat menghasilkan model dengan performa yang optimal, Naive Bayes dapat digunakan dengan data yang terbatas. Hal ini memungkinkan penerapannya dalam proyek-proyek dengan keterbatasan data atau belum tersedianya jumlah data yang signifikan. Oleh karena itu, metode ini sangat cocok digunakan dalam tahap awal penelitian atau pengembangan produk sebelum data yang lebih lengkap tersedia.

Namun, salah satu kekurangan dari metode Naive Bayes ialah asumsi kuat yang digunakan terkait independensi atribut. Dalam kenyataannya, atribut yang digunakan untuk klasifikasi sering kali saling terkait, dan asumsi bahwa semua atribut dapat berdiri sendiri tidak sepenuhnya sesuai dengan data yang digunakan. Hal ini dapat memengaruhi tingkat akurasi dari metode Naive Bayes, khususnya dalam konteks data dengan hubungan atribut yang kompleks dan erat kaitannya. Maka dari itu, metode ini kurang cocok digunakan untuk data dengan pola saling terkait yang signifikan.

Selain itu, Naive Bayes juga dapat memberikan hasil yang kurang memadai ketika digunakan untuk data dengan jumlah atribut yang sangat banyak dan nilai-

nilai atribut yang spars (jarang muncul). Dalam konteks data dengan nilai atribut yang belum pernah terlihat sebelumnya, metode ini dapat menghasilkan nilai probabilitas nol, yang dapat berdampak buruk pada kinerja klasifikasi. Untuk mengatasi hal ini, diperlukan penerapan teknik smoothing, seperti Laplace smoothing, guna memperbaiki nilai probabilitas dan membuat metode ini lebih adaptif terhadap nilai-nilai atribut yang belum pernah ditemukan.

Kelemahan lain dari Naive Bayes ialah kesulitan dalam menangani data yang tidak lengkap atau data dengan nilai yang hilang (missing value). Berbeda dengan metode lain yang dapat memanfaatkan pola data dari atribut lain untuk mengisi nilai yang hilang, metode ini tidak dapat melakukan hal yang sama dengan baik. Akibatnya, metode Naive Bayes dapat memberikan hasil klasifikasi yang kurang akurat bila digunakan pada data dengan jumlah nilai hilang yang signifikan, kecuali dilakukan pre-processing atau imputasi nilai yang memadai sebelum digunakan.

Naive Bayes juga cenderung kurang efektif ketika diterapkan pada data dengan struktur pola yang sangat kompleks dan tidak dapat dijelaskan dengan asumsi independensi atribut. Pola semacam ini lebih cocok dianalisis dengan metode lain, seperti Random Forest, Support Vector Machine, atau metode deep learning yang dapat menangkap pola dari atribut yang saling terkait dengan lebih baik. Oleh sebab itu, pemilihan metode ini perlu disesuaikan dengan kebutuhan dan karakteristik data yang digunakan.

Meskipun memiliki berbagai keterbatasan, metode Naive Bayes tetap menjadi pilihan yang relevan bagi banyak kebutuhan klasifikasi, terutama ketika

data yang digunakan tidak terlalu kompleks atau belum tersedia dalam jumlah yang besar. Dengan memahami asumsi dan pola kerja metode ini, para praktisi dapat memanfaatkan kecepatan dan kesederhanaannya untuk kebutuhan awal pengembangan model atau bahkan untuk kebutuhan klasifikasi yang membutuhkan interpretasi cepat dan efisien.

Pada akhirnya, metode Naive Bayes dapat disebut sebagai metode klasifikasi yang sangat efektif ketika digunakan dengan konteks data yang sesuai dengan asumsi dan pola kerja algoritma ini. Berbagai kelebihan dan kekurangan yang dimilikinya membuat metode ini tetap relevan hingga saat ini, bahkan ketika teknologi dan kebutuhan analisis data terus berkembang. Dengan penerapan pre-processing yang baik, penyesuaian asumsi, dan pemahaman mendalam terkait pola data, metode ini dapat digunakan sebagai pondasi yang kuat bagi pengembangan model klasifikasi yang lebih kompleks maupun untuk kebutuhan bisnis dan akademik yang membutuhkan model yang efisien dan dapat diandalkan.

2.6. Penelitian Terdahulu

Penelitian ini sebelumnya menunjukkan Bahwa Naive Bayes memiliki performa yang baik di dalam berbagai aplikasi, sebagai contoh:

Tabel 2. 1. Penelitian Terdahulu

Referensi Penelitian	1
Judul	Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics
Nama Penulis	Dr. V. Suma,

Tahun	2020
Hasil	Penerapan data mining dalam memprediksi permintaan barang elektronik rekondisi telah terbukti memberikan tingkat akurasi yang tinggi dan dapat diandalkan. Dengan memanfaatkan berbagai metode analisis data yang canggih, data mining memungkinkan sebuah perusahaan untuk menemukan pola-pola kompleks dan hubungan tersembunyi dari data penjualan dan kebutuhan pasar. Proses ini memungkinkan model analitik memahami berbagai faktor yang memengaruhi pola permintaan, mulai dari tren musiman, perubahan kebutuhan konsumen, hingga pola pembelian dari waktu ke waktu. Melalui hasil prediksi yang diperoleh dari penerapan data mining, sebuah bisnis dapat mengoptimalkan tingkat persediaan dan mengelola rantai pasok dengan lebih efisien, sehingga dapat memastikan ketersediaan produk sesuai dengan kebutuhan yang muncul di pasar. Hal ini juga memberikan peluang bagi perusahaan untuk meningkatkan tingkat kepuasan pelanggan dengan menawarkan produk yang sesuai dengan kebutuhan dan preferensi mereka. Dengan begitu, penerapan data mining tidak hanya membantu dalam memahami pola dan tren permintaan, tetapi juga dapat digunakan sebagai

	landasan yang kuat dalam pengambilan keputusan strategis guna mengelola penawaran dan permintaan produk elektronik rekondisi secara lebih tepat dan efektif. [10].
Referensi Penelitian	2
Judul	Intelligent Recommendation System Based on Mathematical Modeling in Personalized Data Mining
Nama Penulis	Yimin Cui
Tahun	2021
Hasil	Dalam era data besar, teknologi penambangan data telah menjadi salah satu kunci dalam penelitian dan bisnis, khususnya dalam pengembangan sistem rekomendasi cerdas. Salah satu penerapan yang signifikan terlihat pada sistem rekomendasi produk, yang memanfaatkan pemodelan matematis dengan aturan asosiasi untuk memberikan saran yang relevan bagi pengguna. Dengan teknologi Java 2, Edisi Perusahaan, sistem ini dirancang dengan arsitektur tiga lapisan, yaitu lapisan presentasi, logika bisnis, dan lapisan data, guna memisahkan dan mengoptimalkan masing-masing proses kerja. Dalam implementasinya, modul rekomendasi terdiri dari tiga bagian utama, yakni representasi data, pembelajaran model, dan mesin rekomendasi, yang dikombinasikan

	dengan algoritma fuzzy clustering guna menghasilkan pola rekomendasi yang lebih tepat. Pengujian sistem ini dilakukan pada sebuah platform e-commerce dan berhasil menunjukkan peningkatan signifikan dalam rasio klik dan tingkat pembelian produk. Berdasarkan data eksperimen, nilai akurasi sistem dapat mencapai angka 0,73 dengan cakupan sebesar 0,75, yang berarti sebagian besar produk yang direkomendasikan memang sesuai dengan kebutuhan pengguna. Rasio klik-tayang juga tercatat mengalami peningkatan rata-rata sebesar 11,04%, sedangkan tingkat pembelian naik hingga 9,35%. Hasil ini membuktikan bahwa penerapan teknologi penambangan data dengan pendekatan sistem rekomendasi dapat memberikan nilai lebih bagi bisnis e-commerce, meskipun aspek pengalaman dan kepuasan pengguna masih perlu dikembangkan lebih jauh agar dapat terus mengoptimalkan interaksi dan loyalitas pelanggan. [11].
Referensi Penelitian	3
Judul	Implementasi Metode Naive Bayes Pada Perancangan Aplikasi Sistem Pakar Diagnosa Bronkiektasis
Nama Penulis	Hamzah Muhar Siregar

Tahun	2020
Hasil	Penerapan metode Naive Bayes dalam sistem pakar untuk diagnosa penyakit Bronkiektasis telah terbukti sangat efektif dalam membantu proses konsultasi kesehatan yang cepat dan efisien. Dalam konteks penyakit ini, di mana kebutuhan untuk mendapatkan diagnosa dari tenaga medis spesialis sering kali terkendala oleh waktu dan ketersediaan dokter, metode Naive Bayes menawarkan sebuah alternatif yang dapat menjembatani kebutuhan tersebut. Metode ini bekerja dengan memanfaatkan asumsi bahwa setiap fitur atau gejala yang digunakan dalam proses klasifikasi berdiri sendiri dan tidak saling terkait, sehingga dapat menghitung nilai probabilitas dari masing-masing kelas dengan cepat dan akurat. Hasil dari klasifikasi yang dilakukan sistem ini ditentukan berdasarkan nilai posterior tertinggi dari berbagai kelas yang tersedia, memungkinkan sistem memberikan gambaran mengenai tingkat kemungkinan seseorang menderita Bronkiektasis dengan memanfaatkan pola data yang sudah dimasukkan. Dengan mekanisme ini, pengguna dapat memperoleh informasi awal terkait kondisi kesehatan mereka secara

	instan tanpa harus menunggu konsultasi dengan dokter spesialis. Hal ini tidak hanya membuat pemeriksaan lebih efisien, tetapi juga dapat membantu meningkatkan aksesibilitas pelayanan kesehatan bagi siapa pun yang membutuhkannya. [12].
Referensi Penelitian	4
Judul	Prediksi Perguruan Tinggi Negeri dengan Menggunakan Metode Naive Bayes
Nama Penulis	Rahmania Aulia Ikimah Putri[1], Tacbir Hendro Pudjiantoro[2]
Tahun	2020
Hasil	Penerapan metode Naive Bayes dalam menganalisis tingkat prestasi siswa di Sekolah XYZ telah memberikan kontribusi signifikan sebagai alat prediksi yang andal. Model ini memungkinkan Bagian Bimbingan Konseling (BK) untuk memberikan arahan yang lebih spesifik bagi siswa terkait pemilihan jalur masuk Perguruan Tinggi Negeri (PTN) melalui Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Dengan memanfaatkan data nilai dan prestasi akademik yang dicapai oleh siswa, metode Naive Bayes dapat menghitung probabilitas penerimaan dengan tingkat akurasi yang memadai,

	sehingga memungkinkan BK memberikan saran dan rekomendasi yang sesuai dengan kemampuan dan potensi masing-masing peserta didik. Selain itu, penerapan metode ini juga membantu siswa dalam merencanakan langkah dan strategi yang lebih matang untuk menghadapi proses seleksi masuk PTN. Dengan memahami pola nilai dan pola penerimaan dari data sebelumnya, siswa dapat fokus meningkatkan area yang perlu diperbaiki guna memaksimalkan peluang diterima di program studi dan universitas yang diinginkan. Hasil dari penerapan metode Naive Bayes ini bukan hanya memperluas pemahaman mengenai kemampuan akademik siswa, tetapi juga memberikan arahan strategis yang dapat membawa siswa lebih dekat dengan tujuan akademik dan cita-cita masa depannya [13].

2.7. Alat Bantu Program Aplikasi Orange

Aplikasi Orange merupakan salah satu alat bantu pemrograman yang populer digunakan dalam bidang data mining dan machine learning, terutama bagi para peneliti maupun praktisi yang membutuhkan antarmuka visual dan intuitif. Orange memungkinkan pengguna untuk membangun alur kerja analisis data dengan metode drag-and-drop, tanpa perlu menuliskan kode secara manual. Berbagai widget yang

tersedia di dalamnya mempermudah proses pengolahan data, mulai dari import data, pembersihan data, hingga penerapan berbagai metode analisis seperti klasifikasi, clustering, dan visualisasi pola. Dengan antarmuka yang ramah bagi pemula, Orange memungkinkan siapa saja dapat dengan cepat memulai eksperimen dengan data nyata.

Selain itu, Orange juga memberikan fleksibilitas bagi pengguna tingkat lanjut dengan menyediakan kemampuan integrasi bersama bahasa pemrograman Python, memungkinkan pengembang membuat atau memodifikasi algoritma sesuai kebutuhan spesifik. Berbagai plugin yang tersedia juga memperluas kapabilitas Orange, menjadikannya platform yang dapat digunakan untuk berbagai kebutuhan analisis data, mulai dari penelitian akademik hingga implementasi bisnis. Berkat kemudahan dan kelengkapan fiturnya, Orange dapat digunakan sebagai alat bantu pemrograman yang efisien untuk memahami pola dan membuat model prediksi dari data, menjadikannya pilihan yang ideal bagi siapa pun yang bekerja dengan analisis data dan machine learning.

2.8. Kelebihan Penelitian

2.8.1. Kelebihan Penelitian:

1. Metode Naïve Bayes dikenal dengan kemampuannya memberikan hasil klasifikasi yang cepat dan akurat, terutama pada dataset yang memiliki jumlah variabel besar namun ukuran data relatif kecil. Hal ini sejalan dengan penelitian Abdillah (2024), yang menunjukkan efektivitas Naïve
2. Bayes dalam klasifikasi data sentimen aplikasi dengan kompleksitas rendah. Rozi (2023) juga mencatat bahwa algoritma ini sangat cocok digunakan

untuk analisis data penjualan karena proses perhitungannya yang sederhana namun efisien[28].

3. Metode Naïve Bayes sangat baik untuk memprediksi kategori seperti "Laris", "Tidak Laris", atau "sedang Laris" berdasarkan data historis. Dalam konteks penelitian ini, metode ini digunakan untuk memprediksi kelarisan produk handphone di JW Celluler, dengan mempertimbangkan berbagai atribut seperti merek, harga, spesifikasi, tren, dan waktu penjualan.
4. Naïve Bayes telah diterapkan dalam berbagai domain penelitian seperti prediksi penjualan (Rozi, 2023), analisis sentimen (Nurwahidah, 2024), serta klasifikasi kelayakan pinjaman (Lestari et al., 2020). Fleksibilitas ini menunjukkan bahwa algoritma dapat dengan mudah disesuaikan untuk berbagai kebutuhan penelitian, termasuk dalam memprediksi kelarisan produk berdasarkan data penjualan handphone[35].
5. Naïve Bayes dapat menangani data kategori (diskrit) maupun data numerik (kontinu) dengan baik. Menurut Widodo (2023), algoritma ini tidak memerlukan preprocessing data yang kompleks, sehingga mempercepat proses analisis [36].
6. Kesederhanaan metode Naïve Bayes tidak mengurangi efektivitasnya dalam menghasilkan prediksi yang akurat. Ahmad et al. (2023) menyatakan bahwa algoritma ini sangat efektif bila diterapkan pada dataset dengan distribusi probabilitas yang sesuai, menjadikannya pilihan yang andal dalam berbagai penelitian klasifikasi[37].

2.8.2. Kekurangan Penelitian:

1. Salah satu kelemahan utama Naïve Bayes adalah asumsi independensi antar variabel prediktor, yang sering kali tidak realistis dalam data dunia nyata. Dalam penelitian ini, variabel seperti harga, merek, dan spesifikasi produk kemungkinan besar saling memengaruhi. Nanni dan Sudrahsyah (2020) menunjukkan bahwa asumsi independensi ini dapat mengurangi akurasi prediksi dalam dataset yang kompleks.
2. Naïve Bayes memiliki keterbatasan dalam menangani dataset yang tidak seimbang. Luthfiyyah (2024) menjelaskan bahwa data dengan distribusi kelas yang tidak merata dapat memengaruhi probabilitas prior, sehingga model lebih cenderung memprediksi kelas mayoritas. Dalam konteks penelitian ini, jika data penjualan JW Celluler tidak merata, hasil prediksi dapat menjadi bias[38].
3. Keterbatasan dalam Prediksi Produk Baru Naïve Bayes kurang efektif dalam memprediksi kelarisan produk yang baru dirilis karena tidak adanya data historis untuk produk tersebut. Hal ini menjadi tantangan dalam konteks penelitian penjualan handphone, di mana inovasi produk sering terjadi dan data historis mungkin tidak tersedia untuk model yang baru dirilis.
4. Menurut Saha et al. (2021), hasil analisis Naïve Bayes kurang transparan jika dibandingkan dengan algoritma berbasis pohon keputusan, sehingga pengguna awam mungkin kesulitan memahami hasil yang dihasilkan[33].

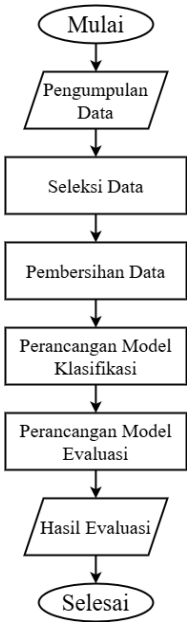
2.9. Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan panduan atau struktur yang menjelaskan langkah-langkah yang akan ditempuh dalam suatu penelitian. Kerangka ini berfungsi untuk memberikan arah yang jelas mengenai proses penelitian serta hubungan antara berbagai komponen yang terlibat. Dalam penelitian ini, fokus utamanya adalah menganalisis data penjualan produk menggunakan metode data mining, dengan tujuan untuk memperoleh wawasan yang lebih mendalam terkait tren penjualan dan strategi pemasaran yang optimal. Kerangka kerja ini mencakup tahapan penelitian, metodologssi yang digunakan, sumber data, serta teknik analisis yang diterapkan untuk mencapai tujuan penelitian yang komprehensif.

Secara keseluruhan, penelitian ini bertujuan untuk mengidentifikasi pola-pola tersembunyi dalam data penjualan dan mengoptimalkan strategi pemasaran melalui penerapan algoritma data mining yang relevan. Dengan menggunakan teknik-teknik seperti regresi linear, k-means clustering, dan FP-Growth, penelitian ini berusaha untuk memberikan solusi berbasis data yang konkret bagi perusahaan untuk meningkatkan performa penjualannya.

Kerangka kerja penelitian ini dapat dijelaskan melalui diagram alur (flowchart) berikut. Diagram ini menggambarkan langkah-langkah dalam proses data mining yang dilakukan dengan menggunakan beberapa algoritma untuk menganalisis data penjualan, termasuk regresi linear, K-means clustering, FP-Growth, dan K-NN. Setiap algoritma ini diterapkan pada data penjualan untuk memprediksi dan mengklasifikasikan pola penjualan guna memperoleh informasi

yang dapat digunakan dalam pengambilan keputusan untuk strategi pemasaran dan manajemen stok barang[29].



Gambar 2. 2. Kerangka Kerja Penelitian

1. Pengumpulan Data: Tahap ini bertujuan untuk mengumpulkan data dari berbagai sumber yang relevan dengan kebutuhan penelitian atau analisis. Data yang dikumpulkan harus lengkap dan sesuai dengan kebutuhan agar dapat digunakan pada tahap selanjutnya.
2. Seleksi Data: Pada tahap ini, data yang telah dikumpulkan disaring dan dipilih sesuai dengan kriteria atau atribut yang diperlukan. Seleksi data memastikan hanya data yang relevan dan berkualitas yang digunakan dalam proses analisis.
3. Pembersihan Data: Tahap ini dilakukan untuk memperbaiki data dengan menghapus nilai-nilai yang tidak lengkap, tidak konsisten, atau duplikat. Dengan pembersihan data, kualitas dan akurasi data dapat dijamin sebelum digunakan untuk pemodelan.

4. Perancangan Model Klasifikasi: Pada tahap ini, model klasifikasi dirancang berdasarkan metode atau algoritma tertentu untuk dapat mengelompokkan data dengan pola yang relevan. Model yang dirancang akan digunakan untuk mempelajari pola dari data yang sudah disiapkan.
5. Perancangan Model Evaluasi: Tahap ini dilakukan untuk menentukan metode dan metrik yang digunakan dalam mengukur tingkat akurasi dan kinerja model klasifikasi. Evaluasi model bertujuan untuk memastikan bahwa model yang digunakan dapat memberikan prediksi yang andal.
6. Hasil Evaluasi: Pada tahap ini, hasil dari proses evaluasi model dikumpulkan dan dianalisis untuk mengetahui tingkat keberhasilan model klasifikasi. Hasil evaluasi juga digunakan sebagai bahan pertimbangan untuk pengembangan atau perbaikan model di masa mendatang.