

BAB II

LANDASAN TEORI

2.1 Teori dan Konsep Dasar Data

Data penjualan merupakan data yang tersimpan secara historis memiliki potensi besar untuk menghasilkan pengetahuan baru apabila di analisis secara sistematis. Fakta tersebut mencakup jenis produk, jumlah terjual dan periode transaksi yang belum memiliki makna strategis apabila tidak di olah lebih lanjut. Informasi muncul setelah data di proses melalui tahapan tertentu, sehingga mampu menjelaskan kondisi penjualan secara lebih terstruktur. Pengetahuan terbentuk ketika informasi tersebut di analisis dan digunakan sebagai dasar pengambilan keputusan. Hubungan antara data, informasi dan pengetahuan menjadi landasan penting dalam peneliti ini , karena klasifikasi produk terlaris bertumpu pada kemampuan mengubah data penjualan menjadi pengetahuan yang bernilai gizi pengelolaan Toko Mamak Rey.

Konsep ini relevan dengan *data mining* yang di pahami sebagai proses sistematis untuk mengekstraksi pola dan pengetahuan tersembunyi dari kumpulan data dalam jumlah besar. Pendekatan *data mining* memungkinkan data penjualan yang bersifat historis di manfaatkan untuk memahami kecenderungan produk yang di minati konsumen. Pemanfaatan konsep ini sesuai digunakan untuk menganalisis tingkat penjualan pada usaha .

2.1.1 Konsep *Data Mining*

Data mining sebagai proses sistematis untuk menemukan pola, hubungan dan pengetahuan yang bermakna dari kumpulan data dalam jumlah besar. (Wardani et

al., 2023) menyatakan bahwa data mining berperan mengubah data penjualan yang bersifat historis menjadi informasi aktif yang dapat di manfaatkan dalam pengambilan keputusan. Pada data penjualan, *data mining* digunakan untuk menganalisis transaksi yang tersimpan dan mengidentifikasi pola pembelian konsumen, seperti produk yang paling sering terjual. Proses ini memungkinkan pemilik usaha memperoleh wawasan yang tidak dapat di peroleh hanya dengan melihat data transaksi manual. Proses ini tidak hanya berfokus pada pengolahan data mentah, tetapi juga pada transformasi data menjadi informasi yang bermakna dan dapat dimanfaatkan dalam pengambilan keputusan. Strategi tersebut menempatkan *data mining* sebagai bagian penting dari pengolahan data modern, khususnya pada bidang bisnis dan penjualan yang menghasilkan data transaksi secara berkelanjutan.

Pengolahan data penjualan menjadi penting karena pengelolaan stok yang efektif sangat bergantung pada pemahaman terhadap pola permintaan produk. (Prapti, Hartanti dan Permatasari et al., 2022) memberikan bahwa tanpa pengolahan data yang terstruktur , pelaku usaha beresiko mengalami kelebihan stok pada produk yang kurang di minati dan kekurangan stok pada produk yang laris . Melalui penerapan *data mining* dan teknik klasifikasi seperti *Naive Bayes*, data penjualan dapat digunakan untuk mengidentifikasi produk terlaris secara objektif. Hasil pengelolaan dapat membantu usaha menyusun strategi pengadaan barang yang lebih tepat sasaran, efisien dan berbasis data, sehingga mendukung peningkatan kinerja usaha.

Penerapan *data mining* pada data penjualan UMKM memiliki nilai strategis, karena karakteristik data yang relative sederhana, namun banyak informasi. (Rahayu & Apridonal, 2024) menyatakan bahwa pemanfaatan teknik klasifikasi dalam data mining mampu membantu usaha dalam mengelompokkan data penjualan ke dalam kategori tertentu secara objektif dan konsisten. Pendekatan ini memungkinkan pemilik usaha memperoleh gambaran yang lebih jelas mengenai produk terlaris tanpa bergantung.

Hubungan konsep tersebut dengan penelitian ini terletak pada penggunaan algoritma *Naive Bayes* sebagai metode klasifikasi prpduk terlaris. Data historis penjualan Toko Mamak Rey diproses melalui tahapan *data mining* untuk menghasilkan model klasifikasi yang mampu mengelompokkan produk berdasarkan tingkat kellarisannya. Dengan demikian, data mining tidak hanya berfungsi sebagai alat analisis data, tetapi juga sebagai dasar ilmiah dalam mendukung pengambilan keputusan pengelolaan stok yang lebih tepat dan berbasis data (Rahayu & Apridonal, 2024)

2.1.2 Kelebihan dan Kekurangan *Data Mining*

(Rahayu & Apridonal,2024) menjelaskan bahwa *data mining* memiliki keunggulan utama dalam kemampuannya mengidentifikasi pola tersembunyi dari data dalam jumlah besar yang sulit dianalisis secara manual. Proses ini memungkinkan indentifikasi pola penjualan, kecenderungan permintaan produk, serta hubungan antar atribut data secara objektif. Keunggulan lain terletak pada dukungan data, dimana keputusan bisnis tidak lagi bergantung pada naluri, tetapi pada hasil analisis data yang terukur. Pada penjualan, *data mining* membantu

meningkatkan efektivitas pengelolaan stok dan perencanaan strategi penjualan karena informasi yang dihasilkan bersifat perkiraan dan bermanfaat. Selain itu, teknik klasifikasi dalam *data mining* dinilai mampu memberikan hasil yang konsisten dan dapat digunakan secara berulang pada data historis yang terus bertambah.

Keterbatasan data mining menurut (Rahayu & Apridonal, 2024) ketergantungannya terhadap kualitas data. Data yang tidak lengkap, tidak konsisten atau mengandung banyak kesalahan dapat memengaruhi akurasi hasil analisis. Proses prapemrosesan data juga membutuhkan waktu dan ketelitian yang tinggi sebelum teknik *data mining* dapat diterapkan secara optimal. Selain itu, pemilihan algoritma yang tidak sesuai dengan karakteristik data dapat menghasilkan model yang kurang akurat dan sulit di jelaskan.

2.1.3 Klasifikasi dalam *Data Mining*

Klasifikasi menjadi salah satu fungsi utama *data mining* yang digunakan untuk mengelompokkan data ke dalam kelas tertentu berdasarkan karakteristik yang dimilikinya. Produk penjualan dapat di klasifikasikan ke dalam kategori terlaris dan tidak terlaris melalui proses pembelajaran dari data sebelumnya. Pendekatan ini memberikan gambaran objektif tentang tingkat penjualan, sehingga pemilik toko dapat menyusun strategi stok secara lebih terukur. Di sisin lain penerapan *data mining* pada data penjualan mampu meningkatkan efektivitas pengolaan persediaan dan mengurangi resiko kerugian akibat stok yang tidak optimal (Wardani et al., 2024)

Menurut (Wardani et al., 2024) mengindikasikan bahwa penerapan *data mining* dalam klasifikasi memberikan kontribusi penting terhadap pengelolaan persediaan barang. Hasil klasifikasi produk terlaris dapat digunakan sebagai dasar dalam menentukan prioritas pengadaan barang, menguramgi resiko kelebihan stok pada produk yang kurang diminati, serta mencegah kekurangan stok pada produk dengan tingkat permintaan tinggi. Namun, *data mining* tidak hanya berfungsi sebagai alat analisis teknis, tetapi juga sebagai sarana pendukung keputusan bisnis. Klasifikasi yang dihasilkan membantu perusahaan menyusun strategi pengelolaan stok yang lebih efektif dan efisien.

2.2 Algoritma Naive Bayes

Algoritma *Naive Bayes* merupakan metode klasifikasi berbasis probabilitas yang berlandaskan Teorema Bayes (Ericha Apriliyani & Salim, 2022). Algoritma ini menghitung peluang suatu data termasuk ke dalam kelas tertentu berdasarkan nilai probabilitas atribut yang dimiliki. Asumsi utama yang digunakan adalah independensi antar atribut, sehingga setiap atribut dianggap memberikan kontribusi tersendiri terhadap hasil klasifikasi. Kesederhanaan struktur perhitungan membuat *Naive Bayes* mudah diterapkan pada dataset dengan jumlah atribut terbatas, seperti data penjualan toko berskala kecil. Prinsip probabilitas ini menjadikan proses klasifikasi bersifat transparan dan dapat dijelaskan secara matematis.

Berbagai penelitian menunjukkan bahwa meskipun asumsi independensi bersifat sederhana, *Naive Bayes* tetap mampu menghasilkan tingkat yang kompetitif pada berbagai kasus klasifikasi, termasuk data penjualan (Ericha Apriliyani & Salim,

2022) Karakteristik ini menjadikan *Naive Bayes* sesuai untuk diterapkan pada Toko Mamak Rey yang memiliki data penjualan dengan atribut terbatas.

Rumus Teorema Bayes

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan:

X : Data dengan Class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas dari X

Rumus ini digunakan untuk menghitung probabilitas setiap kelas berdasarkan atribut data penjualan. Kelas nilai dengan probabilitas terbesar dipilih sebagai hasil klasifikasi, misalnya untuk menentukan apakah suatu produk tersebut termasuk kategori terlaris atau tidak terlaris.

2.2.1 Kelebihan *Naive Bayes*

Karakteristik data penjualan yang relative sederhana dan tidak terlalu kompleks menjadikan *Naive Bayes* sebagai algoritma yang sesuai. Algoritma ini memiliki keunggulan dari sisi efisiensi komputasi dan stabilitas hasil pada dataset berukuran kecil hingga menengah. Kondisi tersebut sejalan dengan kebutuhan

penelitian yang berfokus pada klasifikasi produk terlaris tanpa melibatkan perhitungan yang rumit. Pengguna *Naive Bayes* juga memungkinkan pemilik usaha memahami hasil klasifikasi secara logis karena didasarkan pada nilai peluang yang dapat ditelusuri.

Algoritma *Naive Bayes* memiliki kelebihan utama seperti probabilitas yang sederhana, namun efektif dalam melakukan klasifikasi data. Algoritma ini bekerja dengan menghitung probabilitas suatu kelas berdasarkan atribut atribut yang dimiliki data, sehingga mampu menghasilkan keputusan klasifikasi secara objektif dan terukur. (Wardhani et al., 2023) menjelaskan bahwa kesederhanaan struktur perhitungan *Naive Bayes* menjadikannya mudah di implementasikan serta efisien dalam pengolahan data penjualan yang bersifat historis. Karakteristik ini sangat sesuai untuk lingkungan yang membutuhkan metode analisis cepat dan tidak kompleks.

Kelebihan lain dari *Naive Bayes* terletak pada kemampuan menangani data dengan distribusi kelas yang tidak seimbang. Di sisi lain (Apriliyani et al., 2025) menegaskan bahwa *Naive Bayes* tetap menunjukkan tingkat klasifikasi yang tinggi pada unbalanced dataset, kondisi yang umum di jumpai pada data penjualan. Situasi ini terjadi karena hanya sebagian kecil produk yang tergolong sangat laris, sementara sebagian besar produk lainnya memiliki tingkat penjualan yang lebih rendah. Probabilitas *Naive Bayes* memungkinkan model tetap mengenali pola penjualan secara konsisten meskipun jumlah data antarkelas berbeda.

2.2.2 Penerapan *Naive Bayes*

Penerapan *Naive Bayes* pada data penjualan bertujuan untuk menemukan probabilitas suatu produk tergolong terlaris berdasarkan pola transaksi sebelumnya. Data historis digunakan sebagai dasar pembelajaran untuk membangun model klasifikasi. Model tersebut kemudian di aplikasikan pada data uji untuk menghasilkan label kelas produk. Pendekatan ini memberikan informasi yang jelas mengenai tingkat kelarisan produk, sehingga dapat di manfaatkan dalam pengambilan keputusan stok. Studi terdahulu menegaskan bahwa *Naive Bayes* efektif digunakan dalam klasifikasi produk, karena memudahkan implementasi dan stabilitas hasil (Prapti et al., 2022)

Penerapan algoritma *Naive Bayes* dalam ritel di fokuskan pada pemanfaatan data transaksi penjualan sebagai sumber data historis. Data tersebut mencakup informasi produk, jumlah penjualan serta waktu tertentu yang kemudian digunakan untuk membangun model klasifikasi. (Purnama & Putra, 2024) menunjukkan bahwa *Naive Bayes* mampu mengelompokkan produk ke dalam kategori tertentu, seperti produk dengan tingkat penjualan tinggi dan rendah, sehingga membantu usaha memahami pola permintaan konsumen. Hasil klasifikasi ini dapat di manfaatkan untuk mendukung pengambilan keputusan tentang pengelolaan stok, perencanaan dan pengendalian persediaan.

Algoritma *Naive Bayes* digunakan karena mampu dalam memodelkan hubungan probabilitas antara atribut penjualan dan kelas produk. Proses penerapan di mulai dengan tahap transformasi data yang meliputi pembersihan data, pengelompokkan atribut serta penentuan kelas produk. Data yang telah di proses

kemudian digunakan sebagai data uji coba untuk membangun model *Naive Bayes*. Model tersebut mengatur pola penjualan yang terdapat dalam data historis, sehingga mampu memberikan prediksi jumlah produk terlaris secara konsisten. Pendekatan ini memberikan kesederhanaan algoritma *Naive Bayes* tidak mengurangi kemampuan dalam menghasilkan hasil yang bermakna.

2.3 Tahapan Data Mining

Proses *data mining* dilakukan melalui beberapa tahapan yang saling berurutan dan terstruktur. Setiap tahapan saling berkaitan dan mempengaruhi kualitas hasil analisis. Transformasi data menjadi tahap penting penting karena kesalahan pada tahap ini dapat menurunkan akurasi model klasifikasi (Rozi, 2023). Tahapan ini diawali dengan pengumpulan data yang sesuai dengan tujuan analisis kemudian dilanjutkan dengan tahap transformasi untuk meningkatkan kualitas data. Setelah data siap digunakan, proses dilanjutkan dengan tahap analisis atau pemodelan menggunakan algoritma seperti *Naive Bayes* untuk klasifikasi. Tahap akhir mencakup evaluasi hasil untuk menilai kinerja model serta penjelasan hasil sebagai dasar pengambilan keputusan. Rangkaian tahap ini memastikan pengetahuan yang dihasilkan sebagai data proses yang sistematis dan dapat di pertanggung jawabkan.

2.3.1 Tahapan Knowledge Discovery in Database (KDD)

Pendekatan yang digunakan untuk mengekstraksi pengetahuan penting dari data dalam jumlah besar melalui tahapan terstruktur yaitu *Knowledge Discovery in Database* (KDD). KDD tidak hanya berfokus pada penerapan algoritma *machine learning*, tetapi meliputi keseluruhan proses pengolahan data

sejak tahap awal hingga dihasilkan ilmu yang dapat dimanfaatkan secara praktis (Wardani et al., 2023). Pada klasifikasi produk terlaris, KDD berfungsi sebagai landasan agar analisis data penjualan dilakukan secara terstruktur, objektif dan dapat dipertanggung jawabkan.

Tahap inti dalam KDD adalah *data mining*, pada tahap ini algoritma *Naive Bayes* diterapkan untuk melakukan proses klasifikasi berdasarkan data yang telah diproses sebelumnya. Algoritma ini bekerja dengan probabilitas untuk memprediksi kelas suatu data berdasarkan atribut yang dimilikinya. (Wardani et al 2023) memberikan penerapan *Naive Bayes* pada data penjualan mampu memberikan hasil klasifikasi produk terlaris secara sistematis dan efisien.

2.3.2 Pengumpulan Data

Pengumpulan data dilakukan dengan memanfaatkan data penjualan historis yang dimiliki oleh Toko Mamak Rey. Data tersebut meliputi informasi transaksi penjualan produk dalam periode tertentu, seperti nama produk, jumlah penjualan, dan frekuensi transaksi. Data dikumpulkan dari catatan penjualan toko yang telah terdokumentasi, baik dalam bentuk arsip digital maupun catatan administratif. Pengumpulan data ini bertujuan memastikan bahwa dataset yang digunakan mempresentasikan kondisi penjualan yang sebenarnya sehingga hasil analisis dapat mencerminkan pola penjualan yang akurat. Disisi lain (Prapti et al., 2025) menjelaskan bahwa pemilihan data pada tahap ini sangat menentukan kualitas informasi yang dihasilkan pada tahap selanjutnya, khususnya dalam klasifikasi produk terlaris.

2.3.3 Preprocessing

Data Preprocessing untuk meningkatkan kualitas data melalui pembersihan data dari nilai kosong, duplikasi dan inkonsistensi. Data penjualan sering kali mengandung kesalahan pencatatan yang dapat mempengaruhi kinerja algoritma klasifikasi (Wardani et al., 2023). Oleh karena itu, preprocessing menjadi tahap untuk memastikan data yang digunakan benar benar menampilkan kondisi penjualan yang sebenarnya.

2.3.4 Transformasi Data

Tahap transformasi data dilakukan untuk meningkatkan kualitas data sebelum di analisis. Proses ini meliputi pembersihan data dari nilai kosong, duplikasi atau kesalahan pencatatan yang dapat mempengaruhi hasil klasifikasi. Transformasi data juga dilakukan untuk menyesuaikan format data agar dapat di proses oleh algoritma *Naive Bayes*. Selain itu, pemilihan atribut dilakukan untuk memastikan bahwa hanya atribut yang relevan dengan tujuan klasifikasi yang digunakan dalam pemodelan. (Apriliyani dan Salim 2025) menjelaskan transformasi data diperlukan agar atribut yang digunakan mampu mendukung kinerja algoritma *Naive Bayes*, terutama ketika data memiliki kelas yang tidak seimbang. Transformasi yang tepat membantu model dalam mengetahui pola probabilitas secara lebih akurat.

2.4 Teknik Evaluasi Algoritma

Evaluasi algoritma merupakan tahap penting dalam penelitian klasifikasi karena berfungsi untuk menilai sejauh mana melakukan prediksi secara akurat dan dapat digunakan (Azhari et al., 2025) Evaluasi dilakukan setelah model *Naive Bayes* digunakan. Proses ini bertujuan untuk memastikan bahwa hasil klasifikasi

tidak hanya benar tetapi juga relevan dalam mendukung pengambilan keputusan. Evaluasi kinerja digunakan untuk menilai kemampuan model dalam membedakan produk dengan tingkat penjualan tinggi dan rendah secara konsisten.

Evaluasi algoritma klasifikasi bertujuan mengukur tingkat keberhasilan model dalam mengelompokkan data. Firdaus dan Setiadi (2023) menyatakan bahwa *matrix* evaluasi seperti akurasi dan *confusion matrix* memberikan gambaran yang jelas mengenai performa algoritma *Naive Bayes* dibandingkan algoritma lain. Evaluasi ini diperlukan agar hasil klasifikasi produk terlaris dapat di pertanggung jawabkan secara ilmiah.

2.4.1 Confusion Matrix

Konsep evaluasi kinerja klasifikasi dalam (Azhari et al.,2025) menjelaskan *confusion matrix* sebagai alat evaluasi yang digunakan untuk memberitahukan kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap kondisi data yang sebenarnya. *Confusion matrix* memberikan informasi dalam bentuk tabel yang memberikan jumlah data yang di klasifikasikan secara benar dan salah satu untuk setiap kelas. Melalui matrix ini, dapat mengetahui kesalahan klasifikasi yang dilakukan model serta memahami pola kesalahan. *Confusion matrix* berfungsi sebagai dasar utama dalam menilai ke andalan model klasifikasi secara menyeluruh.

2.4.2 Accuracy, Precision dan Recall

Ukuran *accuracy* menunjukkan tingkat yang benar terhadap keseluruhan model dalam mengklasifikasikan data. *Accuary* memberikan prediksi yang benar terhadap keseluruhan data yang di uji. Meskipun demikian, *accuary* tidak cukup untuk menilai kinerja model secara menyeluruh, terutama pada data dengan

distribusi kelas yang tidak seimbang. Oleh karena itu, ukuran lain seperti *precision* dan *recall* diperlukan untuk memberikan penilaian yang lebih mendalam.

Ukuran *precision* memberikan kesesuaian model dalam memprediksi kelas tertentu. *Precision* memberikan sejauh mana produk yang di prediksi sebagai produk terlaris benar benar merupakan produk dengan tingkat penjualan tinggi. Nilai *precision* yang tinggi menandakan bahwa hasil klasifikasi dapat dipercaya sebagai dasar pengambilan keputusan tentang pengelolaan stok.

Sedangkan *recall* memberikan kemampuan model dalam menemukan seluruh data yang benar pada kelas tersebut. *Recall* yang tinggi memberikan bahwa model *Naive Bayes* mampu menangkap sebagian besar produk terlaris yang ada dalam data penjualan.

Hasil perbandingan algoritma yang dilakukan oleh (Azhari et al., 2025) menunjukkan bahwa *Naive Bayes* memiliki kinerja yang kompetitif dibandingkan algoritma lain seperti C45, Random forest dan SVM. Meskipun bersifat sederhana, *Naive Bayes* mampu menghasilkan nilai *accuracy*, *precision* dan *recall* yang stabil. Temuan ini mengindikasikan memperkuat alasan pemilihan *Naive Bayes* sebagai metode klasifikasi dalam penelitian pada Toko Mamak Rey, karena algoritma ini tidak hanya efisien secara komputasi, tetapi juga memiliki performa evaluasi yang memadai untuk data penjualan.

Evaluasi kinerja klasifikasi menggunakan *accuracy*, *precision* dan *recall* menjadi landasan penting dalam memastikan bahwa implementasi *Naive Bayes* pada klasifikasi produk terlaris menghasilkan model yang valid. Evaluasi yang

baik memungkinkan hasil klasifikasi digunakan secara langsung untuk mendukung strategi pengelolaan persediaan dan pengambilan keputusan bisnis berbasis data pada Toko Mamak Rey.

2.5 Alat Bantu dan Tools Pendukung

Alat bantu pemrograman berguna sebagai sarana untuk menerapkan konsep *data mining* dan algoritma *Naive Bayes*. Penggunaan alat bantu pemrograman dan tools pendukung mempermudah proses implementasi algoritma klasifikasi. Firdaus dan Setiadi (2023) menunjukkan bahwa pemanfaatan tools analisis data membantudalam melakukan pemodelan dan evaluasi algoritma secara lebih efisien. Dalam konteks penelitian ini, tools digunakan sebagai sarana pendukung untuk mengolah data penjualan dan menerapkan algoritma *Naive Bayes*. Membantu dalam proses klasifikasi data sosial menggunakan algoritma *Naive Bayes*. Penelitian ini menggunakan beberapa perangkat lunak dan tools pendukung untuk memastikan kebenaran perhitungan, efesiensi pemrosesan serta memudahkan dalam visualisasi dan analisis data.

2.5.1 Peran *Tools* dan Implementasi Algoritma

(Firdaus, A., & Setiadi, D. 2025) menjelaskan bahwa alat bantu analisis data merupakan perangkat lunak yang digunakan untuk mendukung proses pengolahan, analisis dan pemodelan data secara sistematis. Pada penelitian berbasis *data mining*, alat bantu analisis data berfungsi sebagai media untuk menerapkan tahapan *Knowledge Discovery in Database (KDD)* secara terstruktur, mulai dari pengolahan data awal hingga evaluasi hasil analisis. Keberadaan alat bantu membantu dalam mengolah data berukuran besar serta mengurangi kesalahan yang mungkin terjadi apabila analisis dilakukan secara manual.

Konsep alat bantu analisis data dalam klasifikasi memberikan kemampuan perangkat lunak dalam mendukung proses *preprocessing*, pemodelan dan evaluasi model. (Firdaus, A., & Setiadi, D. 2025) menyatakan bahwa alat bantu analisis data memungkinkan proses pembersihan data, transformasi atribut serta penerapan algoritma *Naive Bayes* dilakukan secara efisien dan konsisten. Fitur-fitur yang tersedia pada alat bantu membantu dalam menyiapkan data agar sesuai dengan kebutuhan algoritma klasifikasi, sehingga hasil analisis yang diperoleh lebih akurat dan dapat dipercaya.

Penggunaan alat bantu analisis data sangat membantu mengubah data transaksi penjualan menjadi informasi yang bernilai. Data yang hanya berupa catatan penjualan dapat diolah dan dianalisis untuk menghasilkan klasifikasi produk terlaris secara objektif. (Firdaus, A., & Setiadi, D. 2025) menegaskan bahwa pemanfaatan alat bantu analisis data dalam *data mining* meningkatkan efisiensi proses serta mendukung pengambilan keputusan berbasis data. Oleh karena itu, alat bantu analisis data diposisikan sebagai komponen penting dalam implementasi algoritma *Naive Bayes* untuk klasifikasi produk terlaris.

2.5.2 Microsoft Excel

Perangkat lunak ini digunakan untuk tahap awal dalam proses input data, pengolahan awal serta perhitungan manual seperti probabilitas dasar, distribusi frekuensi dan pembuatan confusion matrix. Excel dalam mengelola data tabular menjadikan alat bantu yang umum digunakan pada penelitian berbasis data penjualan. Pemanfaatan excel sangat mendukung tahap prapemrosesan data agar data siap untuk digunakan dalam proses klasifikasi.



Gambar 2.1 Aplikasi Microsoft Excel

2.5.3 Rapid Miner

Rapid Miner merupakan salah satu perangkat lunak analisis *data mining* yang menyediakan platform data yang digunakan untuk membangun, melatih dan mengevaluasi model klasifikasi *Naive Bayes*. Fitur yang tersedia seperti transformasi data, pemodelan serta evaluasi hasil klasifikasi. Karakteristik ini menjadikan RapidMiner sesuai digunakan untuk penelitian ini, karena mampu mengimplementasikan algoritma *Naive Bayes* secara sederhana dan terstruktur.



Gambar 2.2 Aplikasi RapidMiner

2.6 Penelitian Terdahulu dan Kelebihan Penelitian

Penelitian ini memiliki keterkaitan dengan penelitian terdahulu, namun memiliki kelebihan pada fokus objek penelitian yang spesifik serta penerapan

langsung pada lingkungan operasional toko. Berbagai penelitian terdahulu menunjukkan bahwa algoritma *Naive Bayes* efektif digunakan dalam klasifikasi data penjualan dan analisis tingkat produk. Rianty,Budi & Efendi (2021) menemukan bahwa *Naive Bayes* memiliki tingkat akurasi yang baik dalam klasifikasi data penjualan. Namun sebagian besar penelitian tersebut masih berfokus pada evaluasi algoritma atau perbandingan metode . Penelitian ini memberikan keunggulan pada penerapan langsung hasil klasifikasi untuk mendukung pengelolaan stok pada usaha ritel skala kecil, sehingga memberikan nilai praktis yang lebih kuat.

Tabel 2.1 Penelitian Dahulu

No	Peneliti	Judul	Tahun	Data yang digunakan	Hasil Penelitian
1	Wardhani, N. W., Nugraha, P.G.S.C., & Mahendra, G.S	Implementasi <i>Naive Bayes</i> pada data mining untuk klasifikasikan penjualan barang terlaris pada perusahaan ritel	2023	Data transaksi penjualan ritel ; algoritma <i>Naive Bayes</i>	Model mampu mengklasifikasi kan produk terlaris dan membantu pengambilan keputusan stok berbasis data
2	Firdaus, L., & Setiadi, T	Perbandingan algoritma <i>Naive Bayes</i> , Decision Tree dan KNN untuk klasifikasi produk populer	2023	Data penjualan produk ritel	<i>Naive Bayes</i> menunjukkan kinerja yang stabil pada klasifikasi produk populer, terutama pada dataset dengan atribut terbatas.
3	Purnama, P. A. W.,&Putra , T. A	Klasifikasi Penjualan Produk Menggunakan Algoritma <i>Naive Bayes</i> pada Konter HP Bayu Cell	2024	Data penjualan ritel harian; algoritma <i>Naive Bayes</i>	Konsep probabilitas mampu meningkatkan efesiensi pengelolaan

4	Prapti, R., Hartati, D., & Permatasari H.	Penerapan <i>data mining</i> untuk memprediksi jumlah produk terlaris menggunakan <i>Naive Bayes</i>	2025	Data penjualan toko ritel.	Model mampu membantu dalam memprediksi produk terlaris dan mendukung perencanaan stok
5	Apriliyani, A., & Salim, A.	Analisis performa metode <i>Naive Bayes</i>	2025	Dataset tidak seimbang	<i>Naive Bayes</i> tetap menunjukkan performa stabil berdasarkan <i>accuracy</i> , <i>precision</i> dan <i>recall</i>
6	Azhari, M., Situmorang, Z., & Rosnelly, R.	Perbandingan akurasi, <i>recall</i> dan presisi klasifikasi pada algoritma C4.5, Random forest, SVM dan <i>Naive Bayes</i>	2025	Dataset klasifikasi; C4.5, Random Forest, SVM dan <i>Naive Bayes</i>	<i>Naive Bayes</i> memiliki performa kompetitif dan layak digunakan untuk klasifikasi.
7	Peneliti ini	Implementasi Algoritma <i>Naive Bayes</i> untuk Klasifikasi Produk Terlaris pada Toko Mamak Rey	2025	Data transaksi Toko Mamak Rey; <i>Naive Bayes</i>	Menghasilkan klasifikasi produk terlaris sebagai dasar pengolaan stok dan strategi pengadaan

2.6.1 Kelebihan dan Kebaruan Penelitian

(Wardani et al., 2023) melakukan penelitian tentang implementasi algoritma *Naive Bayes* untuk mengklasifikasikan penjualan barang terlaris pada perusahaan ritel. Penelitian tersebut memanfaatkan data transaksi penjualan dalam jumlah besar dan bertujuan mengidentifikasi produk dengan tingkat penjualan tertinggi berdasarkan pola data historis. Hasil penelitian memberikan algoritma *Naive Bayes* mampu memberikan tingkat klasifikasi yang bagus dengan tingkat

akurasi yang stabil, serta dapat digunakan sebagai dasar pengambilan keputusan tentang pengelolaan persediaan.

Biarpun memiliki kesamaan dari sisi algoritma yang digunakan (Wardani et al., 2023) memiliki perbedaan dengan penelitian yang dilakukan saat ini. Penelitian tersebut berfokus pada perusahaan ritel dengan tingkat data yang besar dan tujuan analisis yang bersifat umum, sedangkan penelitian ini difokuskan pada satu objek usaha berskala kecil yaitu Toko Mamak Rey dengan tujuan yang lebih khusus yaitu mengklasifikasikan produk terlaris untuk mendukung pengelolaan stok secara mudah.

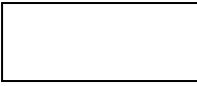
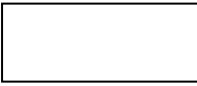
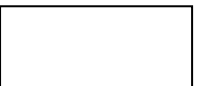
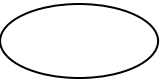
Kelebihan penelitian ini pada penggunaan data penjualan historis asli dari satu objek usaha, sehingga hasil klasifikasi yang didapatkan lebih sesuai dengan kebutuhan dan dapat diterapkan. Strategi tersebut memungkinkan pemilik usaha memanfaatkan hasil penelitian secara langsung dalam pengambilan keputusan sehari-hari. Penelitian ini berfokus pada kesederhanaan implementasi algoritma *Naive Bayes* agar mudah dipahami dan diterapkan pada lingkungan usaha kecil tanpa perhitungan yang sulit.

2.7 Kerangka Penelitian

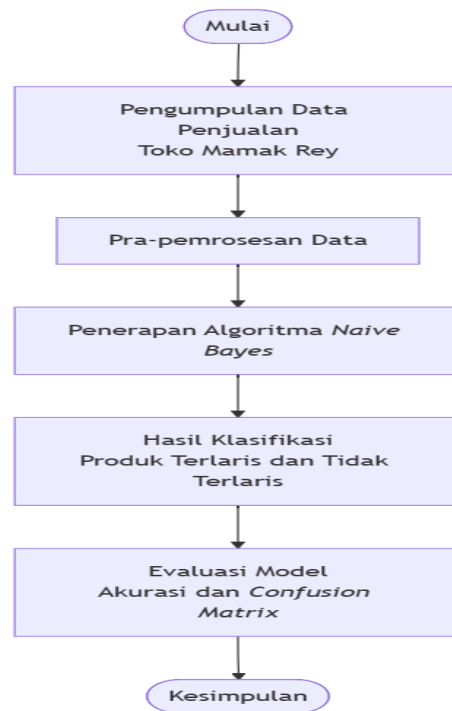
Kerangka penelitian ini memberikan gambaran alur kerja yang menjelaskan tahapan penelitian dari awal hingga diperoleh hasil klasifikasi produk terlaris. Kerangka ini memberikan hubungan antara permasalahan, pengumpulan data, proses analisis menggunakan algoritma *Naive Bayes*, hingga menghasilkan data untuk mengambil keputusan.

2.7.1 Flowchart

Tabel 2.2 Flowchart Penelitian

No	Simbol	Keterangan
1		(Terminator, mulai) Menunjukkan awal mulainya proses penelitian sistem klasifikasi terlaris pada Toko Mamak Rey.
2		(Proses, Pengumpulan Data Penjualan) Tahap pengambilan data penjualan produk dari toko sebagai dataset penelitian untuk di proses lebih lanjut.
3		(Proses, Pra Pemrosesan Data), Tahap membersihkan dan menyiapkan data sebelum di analisis. Data mentah di ubah menjadi data yang siap digunakan oleh algoritma <i>Naive Bayes</i> .
4		(Proses, Penerapan algoritma <i>Naive Bayes</i>), sistem menghitung probabilitas produk berdasarkan data training untuk menentukan kategori produk terlaris dan tidak terlaris.
5		(Proses, Hasil Klasifikasi Produk Terlaris dan Tidak Terlaris), Menampilkan hasil klasifikasi dari proses algoritma. Hasil ini digunakan untuk membantu pengelolaan stok produk toko
6		(Proses, Evaluasi Model Akurasi dan Confusion Matrix), mengukur tingkat penjualan algoritma <i>Naive Bayes</i> . Sistem mengevaluasi apakah hasil klasifikasi sudah berjalan dengan baik dan akurat.
7		(Terminator, Kesimpulan) Menunjukkan hasil dari proses penelitian. Kesimpulan berdasarkan hasil klasifikasi dan evaluasi model yang telah dilakukan.

2.7.2 Diagram Alur Penelitian



Gambar 2.3 Diagram Alur Penelitian

Kerangka kerja penelitian ini terdiri dari beberapa tahap,yaitu;

1. Menjelaskan proses penelitian, yaitu penetapan tujuan analisis untuk mengklasifikasikan produk terlaris dan tidak terlaris secara objektif berbasis data penjualan. Tahap ini menjadi dasar arah penelitian dan menentukan ruang lingkup pengolahan data yang dilakukan.
2. Pengumpulan Data Penjualan Toko Mamak Rey berfokus pada pengambilan data historis penjualan yang dimiliki toko. Data ini memperlihatkan aktivitas transaksi nyata dan menjadi sumber utama dalam proses klasifikasi. Kualitas dan kelengkapan data pada tahap ini sangat menentukan hasil analisis selanjutnya.
3. Transformasi Data bertujuan menyiapkan data agar layak digunakan oleh algoritma. Proses ini meliputi pembersihan data dari kesalahan

pencatatan, duplikasi atau nilai yang tidak akurat serta penyesuaian format data agar konsisten. Tahap ini memastikan bahwa data yang digunakan memiliki kualitas yang baik dan tidak menimbulkan bias pada model.

4. Penerapan Algoritma *Naive Bayes* merupakan inti dari proses klasifikasi. Algoritma ini digunakan untuk mempelajari pola data penjualan dan menghitung probabilitas setiap produk untuk masuk ke dalam kategori terlaris atau tidak terlaris. Pendekatan probabilitas *Naive Bayes* memungkinkan proses klasifikasi dilakukan secara sederhana namun efektif.
5. Hasil Klasifikasi Produk Terlaris menampilkan keluaran model berupa mengelompokkan produk berdasarkan tingkat kelarisan. Hasil ini memberikan informasi yang jelas tentang produk dengan tingkat penjualan tinggi dan rendah, sehingga dapat dimanfaatkan sebagai dasar pengambilan keputusan.
6. Evaluasi Model dilakukan untuk menilai kinerja hasil klasifikasi. Evaluasi menggunakan ukuran akurat untuk melihat tingkat kesesuaian prediksi secara umum, serta *confusion matrix* untuk menggambarkan perbandingan antara hasil prediksi dan kondisi secara lebih jelas. Tahap ini sangat membantu untuk memastikan model yang digunakan sangatlah benar.
7. Tahap Kesimpulan merupakan simpulan akhir berdasarkan hasil klasifikasi dan evaluasi model. Tahap ini merangkum penelitian serta menegaskan peran algoritma *Naive Bayes* dalam membantu klasifikasi produk terlaris pada Toko Mamak Rey.