

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sejumlah kemajuan di bidang teknologi informasi telah muncul sebagai hasil dari perkembangannya yang semakin pesat. Seperti dalam bidang *data mining*, khususnya dalam pengolahan dan analisis data penjualan. Di era digital saat ini, volume data transaksi meningkat secara signifikan seiring dengan kemajuan sistem informasi dan platform digital yang digunakan oleh pelaku usaha untuk mengelola kegiatan bisnis mereka. Menurut Zainul Aras dan Sarjono (2016), *data mining* merupakan sekumpulan teknik eksplorasi otomatis yang digunakan untuk menemukan pola serta relasi kompleks dalam kumpulan data berukuran besar guna mendukung pengambilan keputusan strategis. Dalam konteks bisnis ritel dan penjualan produk, penerapan *data mining* berperan penting dalam memahami perilaku konsumen serta mengidentifikasi produk dengan tingkat permintaan tertinggi (Sulistyo et al., 2025).

Pada tingkat nasional, sektor usaha kecil dan menengah (UMKM) turut menjadi bagian penting dalam perekonomian Indonesia, termasuk dalam bidang penjualan produk hewan peliharaan yang semakin berkembang. Digitalisasi pada sektor UMKM di Indonesia terus mengalami perkembangan pesat seiring dengan meningkatnya adopsi teknologi dalam kegiatan operasional bisnis. Dalam konteks bisnis ritel dan penjualan produk, penerapan *data mining* juga selaras dengan percepatan transformasi digital UMKM, di mana faktor-faktor ekonomi, sosial, teknologi, dan kebijakan memiliki peran penting dalam mendorong optimalisasi strategi penjualan (Wardani et al., 2023). Fenomena ini turut dirasakan oleh pelaku usaha seperti Petshop Bay Bay Sidoarjo, yang memiliki volume penjualan tinggi dan beragam jenis produk. Permasalahan yang muncul adalah sulitnya menentukan produk mana yang paling laris secara konsisten dari waktu ke waktu tanpa bantuan analisis berbasis data. Menurut Wardani et al. (2023), analisis penjualan berbasis *data mining* membantu pelaku usaha dalam pengelolaan stok dan perencanaan strategi pemasaran yang lebih akurat.

Beberapa penelitian sebelumnya telah mengkaji penggunaan algoritma *Decision Tree* dan *Naïve Bayes* dalam konteks klasifikasi data penjualan. Hidayat

et al. (2024) melakukan studi perbandingan antara kedua algoritma tersebut untuk analisis sentimen produk dan menemukan bahwa *Decision Tree C4.5* memiliki akurasi yang lebih stabil dibandingkan *Naïve Bayes*. Sementara itu, Lubis et al. (2023) dalam penelitiannya mengenai pengukuran minat penjualan sepatu. Hal ini menunjukkan bahwa akurasi kedua algoritma tersebut relatif seimbang, namun *Decision Tree* lebih unggul dalam menampilkan aturan klasifikasi yang mudah dipahami oleh pengguna. Temuan tersebut sejalan dengan penelitian oleh Maranto et al. (2024) menegaskan bahwa setiap algoritma memiliki keunggulan tersendiri. Dari penelitian-penelitian tersebut, terlihat bahwa pemilihan metode klasifikasi yang tepat perlu disesuaikan dengan struktur dan karakteristik data penjualan yang dianalisis. Berdasarkan hasil perbandingan tersebut, penulis bermaksud mengimplementasikan kedua algoritma pada konteks data penjualan UMKM untuk mengetahui algoritma yang paling optimal dalam mengklasifikasikan produk terlaris pada Petshop Bay Bay Sidoarjo.

Penelitian ini menggunakan *RapidMiner* sebagai alat bantu utama untuk implementasi algoritma *Decision Tree (C4.5)* dan *Naïve Bayes*. Menurut Ainurrohma (2021), *RapidMiner* merupakan salah satu perangkat lunak *open-source* untuk *data mining* yang tidak sulit digunakan dan tidak memerlukan pengetahuan pemrograman yang canggih. Faisal et al. (2021) menambahkan bahwa penggunaan *RapidMiner* dalam pengujian algoritma *Decision Tree* memungkinkan peneliti untuk memperoleh hasil visualisasi yang lebih jelas terhadap pola data penjualan. Penggunaan *RapidMiner* dalam penelitian ini diharapkan dapat meningkatkan akurasi, efisiensi waktu, dan kemudahan interpretasi hasil klasifikasi produk terlaris.

Penelitian ini berfokus pada perbandingan hasil implementasi algoritma *Decision Tree (C4.5)* dan *Naïve Bayes* dalam mengklasifikasikan produk terlaris menggunakan data penjualan dari Petshop Bay Bay Sidoarjo. Penelitian oleh Bamba et al. (2025) menunjukkan bahwa metode *Naïve Bayes* sering kali menghasilkan akurasi tinggi pada *dataset* dengan ukuran besar dan distribusi sederhana, sedangkan *Decision Tree* lebih cocok untuk data dengan banyak atribut kategorikal. Menurut penelitian oleh Eryana et al. (2025) menerangkan dan menguraikan bahwa analisis komparatif antara algoritma *Naïve Bayes* dan *Decision*

*Tree* perlu dilakukan melalui evaluasi *accuracy*, *precision*, dan *recall* untuk menilai kesesuaian metode terhadap karakteristik data yang berbeda. Juwita et al. (2022) menegaskan bahwa penerapan algoritma *Naïve Bayes* dalam analisis data penjualan mampu membantu pelaku usaha menentukan produk terlaris berdasarkan data transaksi yang tersedia. Penelitian ini tidak hanya membandingkan tingkat *accuracy* antar algoritma, tetapi juga menilai relevansi penerapannya terhadap konteks bisnis dan jenis data yang digunakan.

Klasifikasi produk terlaris berbasis *data mining* juga memiliki potensi besar dalam mendukung perencanaan stok yang lebih efisien. Menurut Hapsari dan Fauzani (2025), penerapan *data mining* pada data transaksi penjualan dapat mempercepat proses pencatatan dan analisis penjualan, sekaligus membantu pelaku usaha dalam mengidentifikasi produk dengan kinerja penjualan terbaik untuk mendukung pengambilan keputusan bisnis. Penerapan algoritma yang tepat dapat membantu mengidentifikasi tren pembelian musiman sehingga strategi penjualan dapat disesuaikan dengan kebutuhan pasar. Bamba et al. (2025) juga menjelaskan bahwa hasil klasifikasi berbasis algoritma cerdas mampu meningkatkan efektivitas perencanaan distribusi produk pada sektor ritel. Dengan menggunakan hasil perbandingan performa algoritma, Petshop Bay Bay Sidoarjo dapat menentukan metode analisis yang paling sesuai untuk karakteristik datanya, baik dari sisi ketepatan hasil maupun kecepatan pemrosesan. Pendekatan ini juga dapat membantu pemilik usaha memahami perilaku konsumen secara lebih mendalam dan terarah. Penelitian ini memberikan gambaran empiris tentang efektivitas metode *Decision Tree* dan *Naïve Bayes* dalam konteks bisnis lokal yang berbasis data transaksi aktual. Penelitian ini tidak terbatas pada penilaian algoritma saja, namun juga menekankan pentingnya penerapan hasil analisis dalam strategi bisnis berbasis data yang berkelanjutan.

Penelitian dilakukan dan berfokus untuk mengimplementasikan, menghitung, menganalisis, dan membandingkan hasil klasifikasi produk terlaris menggunakan algoritma *Decision Tree (C4.5)* dan *Naïve Bayes* dengan bantuan *RapidMiner*. Diharapkan bahwa temuan hasil penelitian ini mampu membantu penulis baik secara teoritis maupun praktis. Membantu peneliti selanjutnya, maupun mitra studi kasus Petshop Bay Bay Sidoarjo. Penelitian ini dapat memberikan kontribusi pada

pengetahuan tentang penggunaan algoritma klasifikasi dalam konteks UMKM secara teoritis. Dalam praktiknya, analisis hasil yang diharapkan mampu menjadi dasar pengambilan keputusan bagi pelaku bisnis dalam menentukan strategi stok, dan pengelolaan penjualan berbasis data.

## **1.2. Rumusan Masalah**

Berdasarkan informasi latar belakang yang telah dijelaskan sebelumnya, penelitian ini dibuat untuk menhadasi sejumlah masalah yang muncul terkait dengan perbandingan algoritma *Decision Tree (C4.5)* dan *Naive Bayes* untuk mengategorikan barang terlaris pada Bay Bay Sidoarjo. Berikut rumusan masalah pada penelitian ini adalah:

1. Bagaimana cara menghitung dan mengimplementasikan algoritma *Decision Tree (C4.5)* dan *Naive Bayes* menggunakan bantuan perangkat lunak *RapidMiner* dalam proses klasifikasi produk terlaris di Petshop Bay Bay Sidoarjo?
2. Metode manakah di antara *Decision Tree (C4.5)* dan *Naive Bayes* yang memberikan hasil klasifikasi paling baik dan akurat berdasarkan nilai *accuracy*, *precision*, dan *recall* yang disesuaikan dengan klasifikasi multikelas pada Petshop Bay Bay Sidoarjo?

## **1.3. Batasan Penelitian**

Dalam penelitian ini, beberapa batasan agar pembahasan tidak menjadi terlalu umum dan agar isu yang dibahas tetap menjadi fokus utama. Penulis menetapkan batasan penelitian yang dimaksud adalah sebagai berikut:

1. Penelitian ini hanya membahas perbandingan kinerja antara algoritma *Decision Tree (C4.5)* dan *Naive Bayes* dalam mengklasifikasikan produk terlaris pada Petshop Bay Bay Sidoarjo.
2. Proses analisis dan pengujian dilakukan menggunakan perangkat lunak *RapidMiner* sebagai alat bantu utama.
3. Data yang digunakan merupakan data histori penjualan produk hewan peliharaan yang diperoleh langsung dari Petshop Bay Bay Sidoarjo, mencakup periode Januari 2022 sampai Desember 2024 sebagai dasar dalam analisis dan pemodelan.

4. Pengukuran performa algoritma difokuskan pada metrik *accuracy*, *weighted mean precision*, *weighted mean recall*, dan waktu komputasi.
5. Penelitian ini tidak meninjau faktor eksternal seperti promosi, perilaku dan *feedback* pelanggan, atau faktor ekonomi makro yang dapat memengaruhi tingkat penjualan produk.

#### 1.4. Tujuan Penelitian

Mengacu pada rumusan masalah dan batasan penelitian yang telah diuraikan, penelitian ini memiliki tujuan sebagai berikut:

1. Untuk menghitung dan mengimplementasikan algoritma *Decision Tree (C4.5)* dan *Naive Bayes* menggunakan *RapidMiner* dalam proses klasifikasi produk terlaris di Petshop Bay Bay Sidoarjo.
2. Untuk mengetahui, menganalisis, dan membandingkan performa kedua algoritma berdasarkan metrik *accuracy*, *precision*, dan *recall* yang disesuaikan dengan klasifikasi multikeals sehingga dapat ditentukan metode yang lebih unggul dalam proses klasifikasi produk terlaris di Petshop Bay Bay Sidoarjo.

#### 1.5. Manfaat Penelitian

Penelitian ini tidak hanya dilakukan sebagai bentuk kegiatan ilmiah, tetapi juga diharapkan mampu memberikan manfaat yang bersifat praktis maupun konseptual bagi pihak-pihak tertentu. Manfaat tersebut diuraikan sebagai berikut:

##### 1. Bagi Penulis

Penelitian ini menjadi sarana pembelajaran bagi penulis dalam memahami dan menerapkan algoritma *Decision Tree (C4.5)* dan *Naive Bayes* menggunakan *RapidMiner* untuk analisis klasifikasi data penjualan. Melalui penelitian ini, penulis memperoleh pengalaman dalam proses pengolahan data, penerapan metode klasifikasi, analisis hasil pengujian model, serta interpretasi akurasi algoritma yang bermanfaat untuk pengambilan keputusan berbasis data di masa mendatang.

##### 2. Bagi Peneliti Selanjutnya

Penelitian ini diharapkan dapat menjadi rujukan bagi mahasiswa maupun peneliti selanjutnya yang berminat mengembangkan kajian mengenai perbandingan algoritma klasifikasi, khususnya dalam bidang *data mining*

dan *machine learning*. Selain itu, hasil penelitian ini dapat dimanfaatkan sebagai landasan untuk penelitian lanjutan, seperti penambahan algoritma pembandingan, penerapan pada dataset dengan skala yang lebih besar, maupun integrasi dengan teknik optimasi parameter guna meningkatkan kinerja model.

### 3. Bagi Mitra Studi Kasus (Petshop Bay\_Bay)

Hasil penelitian ini diharapkan dapat membantu Petshop Bay Bay dalam memahami pola keterjualan produk berdasarkan data historis. Informasi dari hasil implementasi *Decision Tree (C4.5)* dan *Naïve Bayes* dapat digunakan sebagai dasar untuk pengelolaan stok, pengaturan ketersediaan barang, serta pengambilan keputusan operasional yang lebih tepat dan berbasis data di lingkungan bisnis.