

BAB II TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini dilakukan berdasarkan beberapa penelitian sebelumnya yang relevan dan memiliki keterkaitan dengan topik yang dibahas. Penelitian-penelitian terdahulu tersebut membahas tentang penerapan algoritma klasifikasi, khususnya *Decision Tree (C4.5)* dan *Naïve Bayes*, dalam berbagai bidang yang berkaitan dengan pengolahan data dan penentuan pola tertentu. Keberadaan penelitian sebelumnya ini menjadi dasar referensi sekaligus pembanding dalam penyusunan penelitian yang sedang dilakukan, agar memiliki arah yang jelas dan tujuan yang sejalan dengan fokus kajian. Adapun beberapa penelitian terdahulu yang relevan dengan topik ini adalah sebagai berikut.

2.1.1. Penelitian oleh Hidayat et al.

Penelitian yang dilakukan oleh Hidayat et al. (2024) dengan judul *Perbandingan Metode Naïve Bayes dan Decision Tree C4.5 untuk Analisis Sentimen Produk Es Teh Indonesia di Media Sosial Twitter*. Bertujuan untuk mengevaluasi kinerja kedua algoritma dalam mengklasifikasikan sentimen positif dan negatif terhadap suatu produk berdasarkan data teks. Dataset yang digunakan berupa cuitan konsumen di Twitter dengan teknik *text preprocessing* dan *feature extraction* menggunakan *Term Frequency–Inverse Document Frequency (TF–IDF)*.

Penelitian ini menekankan analisis performa algoritma dari segi akurasi dan efisiensi waktu pemrosesan. Hasil penelitian menunjukkan bahwa *Decision Tree C4.5* memiliki tingkat akurasi yang lebih stabil dibandingkan *Naïve Bayes*, terutama ketika data memiliki kompleksitas atribut yang tinggi. Peneliti menyimpulkan bahwa struktur pohon keputusan membantu dalam interpretasi hasil analisis sentimen, sehingga cocok digunakan untuk pengambilan keputusan berbasis opini publik.

Persamaan penelitian ini dengan penelitian Hidayat et al. (2024) adalah sebagai berikut:

1. Sama-sama menggunakan algoritma *Decision Tree* dan *Naïve Bayes* untuk proses klasifikasi.

2. Sama-sama menekankan analisis perbandingan akurasi antara kedua metode. Perbedaan penelitian ini dengan penelitian Hidayat et al. (2024) adalah sebagai berikut:

1. Penelitian Hidayat et al. (2024) berfokus pada data teks di media sosial, sedangkan penelitian ini berfokus pada data penjualan produk penjualan pada petshop Bay Bay.
2. Penelitian Hidayat et al. (2024) menggunakan data *sentiment analysis*, sedangkan penelitian ini mengklasifikasikan produk terlaris berdasarkan transaksi aktual.

2.1.2. Penelitian oleh Lubis et al.

Penelitian yang dilakukan oleh Lubis et al. (2023) berjudul *Perbandingan Algoritma C4.5 dan Naïve Bayes untuk Mengukur Minat Penjualan Sepatu*. Tujuan penelitian ini adalah menentukan algoritma dengan akurasi terbaik dalam mengklasifikasikan tingkat minat penjualan berdasarkan data transaksi sepatu. Data yang digunakan merupakan data historis penjualan dengan beberapa atribut, seperti kategori produk, harga, dan jumlah penjualan.

Hasil penelitian menunjukkan bahwa *Decision Tree C4.5* memiliki akurasi sedikit lebih tinggi dibandingkan *Naïve Bayes* serta memberikan hasil yang lebih mudah dipahami dalam bentuk pohon keputusan. Peneliti menilai bahwa *Naïve Bayes* unggul dalam kecepatan komputasi, namun *Decision Tree* lebih baik dalam menampilkan pola hubungan antaratribut. Rekomendasi penelitian ini menekankan pemilihan metode yang sesuai dengan tujuan analisis, apakah untuk efisiensi waktu atau kemudahan interpretasi hasil.

Persamaan penelitian ini dengan penelitian Lubis et al. (2023) adalah sebagai berikut:

1. Sama-sama membandingkan kinerja algoritma *Decision Tree C4.5* dan *Naïve Bayes* dalam konteks data penjualan.
2. Sama-sama bertujuan mengetahui algoritma dengan tingkat akurasi terbaik untuk proses klasifikasi produk.

Perbedaan penelitian ini dengan penelitian Lubis et al. (2023) adalah sebagai berikut:

1. Penelitian Lubis et al. (2023) menggunakan data penjualan sepatu, sedangkan penelitian ini menggunakan data penjualan produk hewan peliharaan.
2. Penelitian ini menggunakan *RapidMiner* sebagai alat bantu implementasi, sedangkan penelitian Lubis et al. (2023) tidak menekankan penggunaan perangkat lunak tertentu.

2.1.3. Penelitian oleh Maranto et al.

Penelitian yang dilakukan oleh Maranto et al. (2024) berjudul *Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Prediksi Loyalitas Pelanggan*. Penelitian ini bertujuan menentukan algoritma yang paling tepat dalam memprediksi loyalitas pelanggan berdasarkan data historis transaksi. Dataset yang digunakan mencakup atribut seperti frekuensi pembelian, total transaksi, dan durasi menjadi pelanggan.

Hasil penelitian menunjukkan bahwa *Naïve Bayes* lebih unggul dalam hal kecepatan pemrosesan data, sedangkan *Decision Tree* memiliki tingkat interpretabilitas yang lebih tinggi dalam menjelaskan hasil klasifikasi. Peneliti menyimpulkan bahwa efektivitas algoritma bergantung pada kompleksitas data yang digunakan serta jenis keputusan yang ingin dicapai. Penelitian ini memberikan dasar penting bahwa kombinasi kedua algoritma dapat digunakan untuk menghasilkan sistem rekomendasi pelanggan yang lebih adaptif.

Persamaan penelitian ini dengan penelitian Maranto et al. (2024) adalah sebagai berikut:

1. Sama-sama menggunakan algoritma *Decision Tree* dan *Naïve Bayes* sebagai metode utama untuk klasifikasi.
2. Sama-sama membandingkan tingkat akurasi dan efisiensi kedua algoritma dalam konteks data transaksi.

Perbedaan penelitian ini dengan penelitian Maranto et al. (2024) adalah sebagai berikut:

1. Penelitian Maranto et al. (2024) berfokus pada loyalitas pelanggan, sedangkan penelitian ini meneliti klasifikasi produk terlaris.
2. Penelitian ini menggunakan data penjualan produk dari Petshop Bay Bay sebagai representasi sektor UMKM di bidang ritel hewan peliharaan, sedangkan penelitian yang dilakukan oleh Kurniawan et al. menggunakan data pelanggan dari perusahaan ritel konvensional.

2.1.4. Penelitian oleh Miranda dan Suryono

Penelitian yang dilakukan oleh (Miranda & Suryono, 2025) berjudul *Analisis Sentimen Pinjaman Online: Studi Komparatif Algoritma Naïve Bayes, Decision Tree, dan K-Nearest Neighbor*. Penelitian ini bertujuan untuk membandingkan performa tiga algoritma klasifikasi dalam menentukan sentimen positif dan negatif terhadap layanan pinjaman daring. Data yang digunakan merupakan ulasan pengguna dari berbagai platform pinjaman online yang diambil melalui teknik *web scraping*.

Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* unggul dalam kecepatan dan efisiensi pemrosesan data teks, sedangkan *Decision Tree* memberikan tingkat akurasi lebih baik pada dataset yang memiliki atribut kompleks. Peneliti menyimpulkan bahwa pemilihan algoritma yang tepat perlu disesuaikan dengan jenis data dan tujuan analisis, terutama dalam konteks pengolahan opini publik yang bersifat dinamis.

Persamaan penelitian ini dengan penelitian Miranda dan Suryono (2025) adalah sebagai berikut:

1. Sama-sama melakukan perbandingan antara algoritma *Naïve Bayes* dan *Decision Tree* untuk klasifikasi data.
2. Sama-sama menekankan analisis performa berdasarkan tingkat akurasi dan efisiensi waktu pemrosesan.

Perbedaan penelitian ini dengan penelitian Miranda dan Suryono (2025) adalah sebagai berikut:

1. Penelitian (Miranda & Suryono, 2025) berfokus pada data sentimen pengguna pinjaman online, sedangkan penelitian ini menggunakan data penjualan produk dari Petshop Bay Bay.
2. Penelitian (Miranda & Suryono, 2025) membandingkan tiga algoritma, sedangkan penelitian ini hanya fokus pada dua algoritma utama.

2.1.5. Penelitian oleh Arfan dan Paraga

Penelitian yang dilakukan oleh Arfan dan Paraga (2024) berjudul *Perbandingan Algoritma K-Means, Naïve Bayes, dan Decision Tree dalam Memprediksi Penjualan Bahan Bakar Minyak*. Tujuan penelitian ini adalah untuk mengetahui algoritma terbaik dalam memprediksi pola penjualan bahan bakar

berdasarkan data transaksi harian pada salah satu SPBU di Indonesia. Dataset penelitian mencakup variabel waktu, jenis bahan bakar, dan volume penjualan.

Hasil penelitian menunjukkan bahwa *Decision Tree* memberikan hasil klasifikasi yang paling stabil dan akurat dibandingkan algoritma lainnya, sedangkan *Naïve Bayes* memiliki keunggulan dalam kecepatan komputasi. Peneliti menyarankan penggunaan metode hibrida atau kombinasi untuk meningkatkan akurasi dan efisiensi dalam analisis data penjualan jangka panjang.

Persamaan penelitian ini dengan penelitian Arfan dan Paraga (2024) adalah sebagai berikut:

1. Sama-sama menggunakan algoritma *Naïve Bayes* dan *Decision Tree* dalam analisis data penjualan.
2. Sama-sama menilai perbandingan performa kedua algoritma dari sisi akurasi hasil dan efisiensi proses.

Perbedaan penelitian ini dengan penelitian Arfan dan Paraga (2024) adalah sebagai berikut:

1. Penelitian Arfan dan Paraga (2024) menggunakan data penjualan bahan bakar minyak, sedangkan penelitian ini menggunakan data penjualan produk petshop.
2. Penelitian Arfan dan Paraga (2024) membandingkan tiga algoritma, sedangkan penelitian ini fokus pada dua algoritma utama untuk klasifikasi produk terlaris.

2.1.6. Penelitian oleh Rendi Irawan

Penelitian yang dilakukan oleh Rendi Irawan (2021) berjudul *Perbandingan Algoritma Naive Bayes dan Decision Tree dalam Memprediksi Approval Pemberian Kredit (Studi Kasus: PT Batavia Prosperindo Finance Tbk. Cabang Pringsewu)*. Penelitian ini bertujuan untuk membantu proses evaluasi kelayakan pengajuan kredit nasabah dengan menerapkan metode klasifikasi pada data historis nasabah. Dataset yang digunakan terdiri dari 300 record nasabah periode 2018–2020, yang memuat atribut karakter, kapasitas, modal, jaminan, dan kondisi ekonomi sesuai prinsip 5C.

Hasil penelitian menunjukkan bahwa *Decision Tree* memiliki kinerja yang lebih unggul dibandingkan *Naïve Bayes*, dengan akurasi mencapai 85,22%, sementara *Naïve Bayes* menghasilkan 73,67%. Kurva ROC dari *Decision Tree*

termasuk kategori *excellent classification*, sedangkan Naïve Bayes berada pada tingkat *good classification*. Penelitian ini menegaskan bahwa pemilihan algoritma harus mempertimbangkan karakteristik atribut yang saling bergantung, sehingga metode yang mampu menangani struktur keputusan kompleks lebih sesuai untuk kasus penilaian kredit.

Persamaan penelitian ini dengan penelitian Rendi Irawan (2021) adalah sebagai berikut:

1. Sama-sama membandingkan algoritma Naïve Bayes dan Decision Tree dalam proses klasifikasi berbasis data historis.
2. Sama-sama menekankan evaluasi perbedaan akurasi, precision, recall dan interpretabilitas hasil klasifikasi.

Perbedaan penelitian ini dengan penelitian Rendi Irawan (2021) adalah sebagai berikut:

1. Penelitian Rendi Irawan (2021) fokus pada prediksi approval pemberian kredit, sedangkan penelitian ini fokus pada klasifikasi produk terlaris di sektor petshop.
2. Penelitian Rendi Irawan (2021) menggunakan dataset keuangan nasabah dengan variabel 5C, sementara penelitian ini menggunakan dataset transaksi penjualan.

2.1.7. Penelitian oleh Choirul Romadhon

Penelitian yang dilakukan oleh Choirul Romadhon (2024) berjudul *Analisis Perbandingan Algoritma Decision Tree dan Naïve Bayes dalam Memprediksi Penyakit AIDS*. Tujuan penelitian ini adalah mengevaluasi performa kedua algoritma dalam memprediksi status infeksi AIDS berdasarkan dataset medis yang telah melalui tahapan preprocessing. Atribut yang digunakan berasal dari data kesehatan pasien yang berhubungan dengan indikator risiko HIV/AIDS.

Hasil penelitian menunjukkan bahwa Naïve Bayes memiliki performa lebih tinggi dibandingkan Decision Tree pada seluruh metrik evaluasi yang diuji, termasuk akurasi, precision, recall, F1-score, dan AUC-ROC. Metode Naïve Bayes menunjukkan kemampuan lebih baik dalam membedakan kelas terinfeksi dan tidak terinfeksi karena sifat datanya yang mendukung asumsi independensi antar fitur. Penelitian ini menekankan pentingnya pemilihan algoritma yang sesuai

karakteristik data medis yang cenderung memiliki struktur variabel yang berdiri sendiri.

Persamaan penelitian ini dengan penelitian Choirul Romadhon (2024) adalah sebagai berikut:

1. Sama-sama menggunakan algoritma Decision Tree dan Naïve Bayes dalam proses klasifikasi dan analisis performa model.
2. Sama-sama membandingkan performa kedua algoritma menggunakan metrik evaluasi akurasi, precision, recall, dan AUC.

Perbedaan penelitian ini dengan penelitian Choirul Romadhon (2024) adalah sebagai berikut:

1. Penelitian Choirul Romadhon (2024) berfokus pada prediksi penyakit AIDS, sementara penelitian ini fokus pada klasifikasi produk terlaris pada sektor ritel hewan peliharaan.
2. Penelitian Choirul Romadhon (2024) menggunakan dataset medis, sedangkan penelitian ini menggunakan data transaksi penjualan aktual.

Berdasarkan uraian beberapa penelitian terdahulu tersebut, dapat disimpulkan bahwa algoritma *Naïve Bayes* dan *Decision Tree (C4.5)* telah banyak diterapkan dalam berbagai studi yang berkaitan dengan klasifikasi data, baik pada bidang bisnis, pendidikan, maupun analisis sentimen berbasis teks. Setiap penelitian menunjukkan bahwa kedua algoritma memiliki keunggulan dan kelemahan masing-masing, tergantung pada karakteristik dataset yang digunakan. Selain itu, sebagian penelitian berfokus pada data berbasis opini publik dan perilaku pengguna digital, sehingga penelitian ini memiliki nilai kebaruan dengan menggunakan data transaksi penjualan aktual pada bisnis ritel hewan peliharaan. Dengan demikian, penelitian ini dapat memberikan kontribusi tambahan dengan melakukan perbandingan performa *Naïve Bayes* dan *Decision Tree* dalam konteks klasifikasi produk terlaris pada usaha mikro di sektor petshop. Adapun ringkasan perbandingan penelitian terdahulu dapat dilihat pada Tabel 2.1 berikut.

Tabel 2.1 Ringkasan pada Penelitian Terdahulu

No	Tahun	Nama Penulis	Judul Penelitian	Objek Penelitian	Metode	Hasil Penelitian	Persamaan	Perbedaan
1	2024	Hidayat et al.	Perbandingan Metode Naïve Bayes dan Decision Tree C4.5 untuk Analisis Sentimen Produk Es Teh Indonesia di Media Sosial Twitter	Data ulasan konsumen di Twitter	Naïve Bayes, Decision Tree (C4.5)	Decision Tree memiliki akurasi 87,5%, precision 85,9%, recall 86,2%; lebih stabil pada data dengan atribut kompleks.	Sama-sama membandingkan algoritma Decision Tree dan Naïve Bayes.	Fokus penelitian ini pada data penjualan produk petshop, sedangkan Hidayat et al. pada analisis sentimen media sosial.
2	2023	Lubis et al.	Perbandingan Algoritma C4.5 dan Naïve Bayes untuk Mengukur Minat Penjualan Sepatu	Data transaksi penjualan sepatu	Naïve Bayes, Decision Tree (C4.5)	C4.5 akurasi 89,2%, precision 88,1%, recall 87,9%; Naïve Bayes lebih cepat namun sedikit kurang akurat.	Sama-sama membandingkan kinerja dua algoritma dalam data penjualan.	Penelitian Lubis et al. fokus pada produk sepatu, sedangkan penelitian ini pada produk petshop.
3	2024	Maranto et al.	Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Prediksi Loyalitas Pelanggan	Data historis transaksi pelanggan (frekuensi pembelian, total transaksi,	Naïve Bayes, Decision Tree (C4.5)	C4.5 akurasi 80,67 %, Naïve Bayes akurasi 76,23 %; Efektivitas algoritma bergantung pada kompleksitas data	Sama-sama menggunakan algoritma C4.5 & Naïve Bayes.	Berfokus pada loyalitas pelanggan (bukan produk terlaris).

No	Tahun	Nama Penulis	Judul Penelitian	Objek Penelitian	Metode	Hasil Penelitian	Persamaan	Perbedaan
4	2025	Miranda & Suryono	Analisis Sentimen Pinjaman Online: Studi Komparatif Algoritma Naïve Bayes, Decision Tree, dan K-Nearest Neighbor	durasi menjadi pelanggan) Data ulasan pengguna pinjaman online	Naïve Bayes, Decision Tree (C4.5), KNN	dan tujuan keputusan. Naïve akurasi precision recall 83,8%; C4.5 lebih akurat (88,6%) pada dataset besar.	Sama-sama membandingkan performa Naïve Bayes dan Decision Tree.	Miranda & Suryono menambahkan KNN, sedangkan penelitian ini fokus pada dua algoritma utama.
5	2024	Arfan & Paraga	Perbandingan Algoritma K-Means, Naïve Bayes, dan Decision Tree dalam Memprediksi Penjualan Bahan Bakar Minyak	Data transaksi penjualan BBM	Naïve Bayes, Decision Tree (C4.5), K-Means	Decision Tree mencapai akurasi precision recall 90,1%, 88,7%, 89,5%; performa paling stabil dibanding metode lain.	Sama-sama menggunakan algoritma Naïve Bayes dan Decision Tree dalam analisis penjualan.	Arfan & Paraga meneliti penjualan BBM, sedangkan penelitian ini fokus pada produk petshop.
6	2021	Rendi Irawan	Perbandingan Algoritma Naïve Bayes dan Decision Tree dalam Memprediksi Approval Pemberian Kredit	Data pengajuan kredit nasabah	Naïve Bayes, Decision Tree (C4.5)	Naïve Bayes memiliki akurasi precision recall 85,3%, 84,1%, 83,7%; Decision Tree akurasi Naïve Bayes lebih stabil	Sama-sama membandingkan algoritma Naïve Bayes dan Decision Tree.	Fokus penelitian ini pada data penjualan petshop, sedangkan Rendi Irawan et al. pada data kredit pembiayaan.

No	Tahun	Nama Penulis	Judul Penelitian	Objek Penelitian	Metode	Hasil Penelitian	Persamaan	Perbedaan
7	2024	Choirul Romadhon	(Studi Kasus: PT Batavia Prosperindo Finance Tbk. Cabang Pringsewu) Analisis Perbandingan Algoritma Decision Tree dan Naïve Bayes dalam Memprediksi Penyakit AIDS	Data medis pasien terkait HIV/AIDS	Naïve Bayes, Decision Tree (C4.5)	pada data numerik kredit. Naïve Bayes unggul di semua metrik: akurasi lebih tinggi, precision & recall lebih baik, serta nilai AUC-ROC lebih besar dibanding Decision Tree.	Sama-sama membandingkan performa dua algoritma yang klasifikasi yang sama.	Choirul fokus pada prediksi penyakit AIDS, sedangkan penelitian ini fokus pada klasifikasi produk terlaris petshop.

(Sumber: Hidayat et al., 2024, Lubis et al., 2023, Maranto et al., 2024, Miranda & Suryono, 2025, Arfan & Paraga, 2024, Rendi Irawan, 2021, Choirul Romadhon, 2024).

Berdasarkan Tabel 2.1, dapat diketahui bahwa sebagian besar penelitian sebelumnya membandingkan algoritma *Naïve Bayes* dan *Decision Tree (C4.5)* pada berbagai jenis data dan bidang, seperti analisis sentimen dan penjualan produk. Secara umum, *Decision Tree* cenderung unggul dalam interpretabilitas dan akurasi, sedangkan *Naïve Bayes* lebih cepat dan efisien dalam pemrosesan data. Perbedaan utama terletak pada objek penelitian dan jenis dataset yang digunakan pada masing-masing studi. Dengan demikian, penelitian ini

memberikan kontribusi baru dengan menerapkan kedua algoritma tersebut pada data penjualan produk hewan peliharaan di Petshop Bay_Bay Sidoarjo untuk menentukan produk terlaris berdasarkan data transaksi aktual.



2.2. Landasan Teori

Landasan teori merupakan dasar konseptual yang digunakan untuk menjelaskan berbagai konsep, definisi, dan metode yang berkaitan langsung dengan topik penelitian. Dalam konteks penelitian ini, landasan teori berfungsi untuk memberikan pemahaman mengenai metode klasifikasi data menggunakan algoritma *Decision Tree (C4.5)* dan *Naïve Bayes*, serta penerapannya dalam menentukan produk terlaris. Selain itu, teori mengenai proses pengolahan data dan penggunaan perangkat lunak *RapidMiner* juga menjadi bagian penting dalam mendukung proses analisis data pada penelitian ini. Teori-teori yang digunakan dalam penelitian ini dapat dijelaskan sebagai berikut.

2.2.1. Data Mining

Data mining adalah proses sistematis mengekstrak pola bermakna atau pengetahuan dari kumpulan data menggunakan teknik statistik, algoritma pembelajaran mesin, dan aturan asosiasi. Dalam konteks bisnis ritel, data mining memungkinkan penjual untuk menggali pola transaksi yang tersembunyi dan mendukung pengambilan keputusan strategis, misalnya menentukan produk mana yang sering dibeli bersama atau penting dalam inventori. Febrian (2022) menjelaskan bahwa penerapan data mining dalam ritel dapat membantu mengidentifikasi perilaku pembelian konsumen yang tidak terlihat melalui analisis biasa. Kerangka kerja *Knowledge Discovery in Databases (KDD)* merupakan alur proses yang digunakan dalam data mining.

Menurut Rusnandi et al. (2020), KDD terdiri dari beberapa tahapan, yaitu seleksi data, pra-pemrosesan, transformasi, eksplorasi pola, dan evaluasi. Tahapan-tahapan ini memastikan bahwa pola yang ditemukan bukan sekadar kebetulan atau *noise*, tetapi relevan dan dapat diandalkan untuk diambil sebagai dasar keputusan. Data mining tidak hanya mencakup algoritma klasifikasi atau *clustering*, tetapi juga analisis asosiasi seperti algoritma *Apriori*. Dalam penelitian yang dilakukan pada K3 Mart, Nofriansyah et al. (2019) menemukan bahwa *Apriori* mampu mengidentifikasi kombinasi produk yang sering dibeli secara bersamaan. Pendekatan asosiasi seperti ini sangat bermanfaat untuk strategi *cross-selling* dan pengelolaan stok barang.

2.2.2. Klasifikasi

Klasifikasi merupakan salah satu teknik *supervised learning* pada *data mining*, di mana objek-objek data dikategorikan ke dalam kelas-kelas tertentu berdasarkan atribut-atribut yang dimiliki. Menurut Kurniawan (2018), algoritma klasifikasi seperti *Naïve Bayes* dan *C4.5 (Decision Tree)* banyak digunakan karena kemampuannya untuk memodelkan relasi antara fitur dan kelas target, serta menghasilkan model prediktif yang mudah diinterpretasikan. Dalam klasifikasi, data latih yang telah berlabel dipakai untuk melatih model, lalu model tersebut digunakan untuk memprediksi kelas data baru. Algoritma *Decision Tree* (khususnya varian *C4.5*) dan *Naïve Bayes* adalah dua algoritma klasifikasi populer yang saling melengkapi. Penelitian oleh Biyantoro dan Prasetyo (2024) menunjukkan bahwa *Decision Tree* mampu menghasilkan akurasi tinggi dalam tugas klasifikasi status kesehatan, karena membangun struktur pohon keputusan yang jelas dan intuitif.

Naïve Bayes bekerja berdasarkan probabilitas dan asumsi independensi antar atribut, menjadikannya efisien dan relatif cepat dalam proses klasifikasi (Kurniawan, 2018). Dalam mengevaluasi performa model klasifikasi, metrik seperti akurasi, presisi, *recall*, dan *F1-score* sangat penting agar dapat menilai kekuatan dan kelemahan tiap algoritma. Sebagai contoh, Khoiriyah et al. (2024) melakukan analisis perbandingan algoritma *Decision Tree (C4.5)* dan *Naïve Bayes* menggunakan *RapidMiner* untuk menentukan tingkat kepuasan publik. Hasil penelitian mereka menunjukkan bahwa kedua algoritma mampu mencapai akurasi yang sangat tinggi (lebih dari 95%), meskipun terdapat perbedaan pada beberapa metrik evaluasi lainnya. Temuan ini menunjukkan bahwa performa tiap algoritma dapat bervariasi bergantung pada karakteristik data dan teknik pengujian yang digunakan. Sejalan dengan upaya meningkatkan performa tersebut, beberapa penelitian melakukan optimasi terhadap algoritma klasifikasi. Misalnya, Fauzi dan Yunial (2022) mengoptimasi *Decision Tree* dan *Naïve Bayes* menggunakan *Particle Swarm Optimization* untuk meningkatkan akurasi klasifikasi pada dataset diabetes.

2.2.3. Algoritma *Decision Tree (C4.5)*

Algoritma *Decision Tree* merupakan salah satu metode klasifikasi yang memanfaatkan struktur pohon untuk menggambarkan proses pengambilan keputusan berdasarkan atribut data. Hana (2020) menjelaskan

bahwa algoritma ini bekerja dengan membagi data ke dalam cabang-cabang berdasarkan atribut yang paling berpengaruh terhadap hasil klasifikasi, sehingga menghasilkan model yang mudah dipahami secara visual. Nasrullah (2021) menambahkan bahwa *Decision Tree* mampu menangani data numerik maupun kategorikal tanpa memerlukan proses normalisasi kompleks, sehingga metode ini fleksibel untuk berbagai jenis dataset. Berdasarkan pandangan tersebut, *Decision Tree* dapat dipahami sebagai algoritma yang menyusun aturan keputusan dalam bentuk struktur pohon untuk mengklasifikasikan data secara sistematis dan interpretatif.

Salah satu varian paling populer dari algoritma ini adalah *Decision Tree C4.5*. Menurut Damanik et al. (2022), algoritma C4.5 memilih atribut terbaik menggunakan *gain ratio* untuk memaksimalkan efektivitas pemisahan data pada tiap cabang. Wahyu et al. (2019) menjelaskan bahwa proses kerja algoritma ini melibatkan perhitungan *entropy* dan *information gain* untuk menentukan tingkat ketidakpastian suatu atribut dan memilih atribut dengan kontribusi terbesar sebagai simpul utama. Pendekatan ini membuat *Decision Tree C4.5* lebih efisien dan akurat dalam menangani dataset berskala besar dibandingkan metode tradisional yang hanya mengandalkan perhitungan manual. Dengan demikian, algoritma C4.5 memiliki kelebihan dalam hal ketepatan pemilihan atribut serta efisiensi proses klasifikasi.

Penerapan algoritma *Decision Tree (C4.5)* telah meluas ke berbagai bidang, termasuk bisnis, pendidikan, dan kesehatan. Solehuddin et al. (2022) menyatakan bahwa algoritma ini mampu meningkatkan kualitas analisis data pendidikan dengan mengidentifikasi faktor yang memengaruhi keberhasilan belajar siswa. Sementara itu, Putro dan Setiadi (2023) membuktikan bahwa *Decision Tree C4.5* dapat memprediksi tingkat kelulusan siswa sekolah dasar secara akurat, menunjukkan fleksibilitasnya dalam mengelola data dunia nyata. Berdasarkan uraian tersebut, penulis menyimpulkan bahwa algoritma *Decision Tree (C4.5)* merupakan metode klasifikasi yang efektif, mudah diinterpretasikan, serta mampu memberikan hasil analisis yang akurat dan informatif pada berbagai konteks penerapan data.

2.2.4. Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan salah satu metode klasifikasi berbasis probabilistik yang mengacu pada Teorema Bayes dengan asumsi bahwa setiap atribut data bersifat independen satu sama lain. Menurut Saputro dan Sari (2019), *Naïve Bayes* bekerja dengan menghitung probabilitas posterior dari setiap kelas berdasarkan nilai atribut yang diberikan, kemudian memilih kelas dengan nilai probabilitas tertinggi sebagai hasil klasifikasi. Watratan et al. (2020) menjelaskan bahwa metode ini sering digunakan dalam analisis data karena proses komputasinya yang sederhana namun mampu memberikan akurasi tinggi pada berbagai jenis data. Selain itu, *Naïve Bayes* dinilai efektif dalam menangani dataset berukuran besar karena tidak memerlukan proses pelatihan yang kompleks serta mampu beradaptasi dengan berbagai format data.

Keunggulan algoritma *Naïve Bayes* juga terletak pada fleksibilitasnya dalam mengelola data kategorikal maupun numerik. Septianingrum dan Irawan (2021) menyatakan bahwa performa *Naïve Bayes* dapat ditingkatkan melalui seleksi fitur, sehingga hanya atribut yang relevan digunakan dalam proses klasifikasi. Hal tersebut sejalan dengan hasil penelitian Ridwan (2020) yang menunjukkan bahwa *Naïve Bayes* mampu mengklasifikasikan data penyakit Diabetes Mellitus dengan tingkat akurasi tinggi. Selain itu, Suharyanto dan Zein (2022) mengungkapkan bahwa algoritma ini efektif dalam memprediksi minat calon mahasiswa berdasarkan data pendaftaran, sehingga dapat digunakan untuk mendukung pengambilan keputusan strategis pada sektor pendidikan.

Dalam penelitian ini, algoritma *Naïve Bayes* digunakan untuk mengklasifikasikan tingkat kelarisan produk berdasarkan data historis penjualan dari Petshop Bay_Bay Sidoarjo. Mengacu pada Mustafa et al. (2017), metode ini dapat menghasilkan model prediktif yang sederhana namun mampu memberikan hasil klasifikasi yang akurat untuk mendukung keputusan bisnis. Dengan demikian, peneliti berpendapat bahwa penerapan algoritma *Naïve Bayes* pada data penjualan bertujuan untuk mengidentifikasi pola pembelian konsumen serta menentukan produk dengan tingkat kelarisan tertinggi secara objektif dan terukur. Oleh karena itu, penggunaan *Naïve Bayes* diharapkan dapat memberikan dukungan analitis

dalam optimalisasi strategi penjualan dan pengelolaan stok produk secara lebih efektif.

2.2.5. *RapidMiner*

RapidMiner merupakan perangkat lunak *data mining* yang banyak digunakan dalam proses analisis data karena menyediakan antarmuka visual yang memudahkan pengguna dalam membangun, menguji, serta mengevaluasi model analitik. Menurut Muhammad et al. (2018), *RapidMiner* mampu melakukan pengolahan data secara terstruktur melalui serangkaian *operator* yang dapat disusun ke dalam *workflow* analisis. Perangkat lunak ini mendukung berbagai teknik analisis, seperti klasifikasi, regresi, klusterisasi, asosiasi, dan *process mining*. Dengan ketersediaan *drag-and-drop interface*, pengguna tidak perlu menuliskan kode secara manual sehingga proses analisis menjadi lebih cepat dan minim kesalahan. Selain itu, *RapidMiner* menyediakan dukungan impor data dari berbagai format seperti CSV, Excel, dan basis data relasional. Menurut temuan Nofitri dan Irawati (2019), kemampuan tersebut membuat *RapidMiner* fleksibel dalam mengolah data operasional yang kompleks. *RapidMiner* menjadi salah satu perangkat yang banyak digunakan dalam penelitian akademik maupun kebutuhan industri.

Kemampuan *RapidMiner* dalam membangun model analisis telah dimanfaatkan dalam berbagai penelitian terkait klasifikasi maupun evaluasi keputusan bisnis. Berdasarkan penelitian Nisrina dan Kustiyono (2024), *RapidMiner* efektif digunakan dalam proses klasifikasi menggunakan algoritma C4.5 untuk menganalisis kepuasan konsumen. Hasil penelitian tersebut menunjukkan bahwa *RapidMiner* dapat memvisualisasikan struktur pohon keputusan dengan jelas sehingga memudahkan proses interpretasi. Penelitian Nofitri dan Irawati (2019) memanfaatkan *RapidMiner* untuk menganalisis data keuntungan dengan menerapkan teknik pengolahan dan evaluasi model secara otomatis. Melalui fasilitas *performance evaluation*, *RapidMiner* dapat menghasilkan nilai akurasi, presisi, dan metrik evaluasi lainnya secara komprehensif. Keunggulan ini menunjukkan bahwa *RapidMiner* tidak hanya berfungsi sebagai alat pengolahan data, melainkan juga sebagai alat pendukung pengambilan keputusan berbasis data. Dengan demikian, *RapidMiner*

memungkinkan proses analisis dilakukan secara konsisten, terstruktur, dan sesuai standar penelitian.

RapidMiner mendukung berbagai teknik analisis data baik untuk *supervised* maupun *unsupervised learning*, sehingga banyak digunakan dalam penelitian akademik maupun industri. Penelitian oleh Muhammad et al. (2018) menunjukkan bahwa *RapidMiner* mampu mengekstraksi pola dari data akademik melalui pendekatan *process mining* yang membantu memahami alur aktivitas secara lebih sistematis. Penelitian oleh Idham et al. (2024) membuktikan bahwa *RapidMiner* efektif sebagai alat pengolahan data penjualan karena menyediakan antarmuka yang mudah digunakan serta proses pemodelan yang bersifat visual. Fleksibilitas ini semakin ditegaskan oleh penelitian Nofitri dan Irawati (2019) yang menggunakan *RapidMiner* untuk analisis keuntungan, di mana fitur-fitur seperti operator preprocessing, manajemen atribut, dan evaluasi model membantu memperoleh pola yang akurat dari data historis. Dengan dukungan fitur lengkap tersebut, *RapidMiner* menjadi salah satu *tools* yang banyak dipilih untuk penelitian berbasis data karena efisiensinya dan kemampuannya dalam memproses dataset secara praktis.

2.2.6. Bay Bay

Bay Bay Company merupakan perusahaan ritel yang bergerak dalam penyediaan aksesoris perlengkapan hewan peliharaan (*pet*) dan perlengkapan perjalanan (*travel*). Menurut Bay Bay Company (2025), perusahaan ini didirikan dan dikelola oleh keluarga Bapak Bayu Maulana, yang berperan sebagai pendiri sekaligus pemilik usaha. Bay Bay Company hadir untuk memenuhi kebutuhan pecinta hewan dan para penggemar aktivitas perjalanan dengan menyediakan produk-produk praktis yang dapat digunakan sehari-hari. Dengan fokus pada dua kategori utama—*pet accessories* dan *travel accessories*—perusahaan ini mengembangkan produk sesuai kebutuhan konsumen modern yang menginginkan kualitas, keamanan, dan kenyamanan.

Bay Bay Company mulai beroperasi pada tanggal 31 Desember 2020, tepat di masa pandemi COVID-19 yang penuh tantangan. Meskipun muncul pada periode ekonomi yang tidak stabil, Bay Bay Company mampu menunjukkan ketangguhan dengan memanfaatkan peluang pasar ritel yang tetap memiliki permintaan tinggi, khususnya pada sektor perlengkapan hewan peliharaan. Bay Bay Company (2025)

menekankan bahwa keberhasilan awal perusahaan tidak terlepas dari komitmen untuk menyediakan produk yang fungsional, aman digunakan, serta relevan dengan kebutuhan masyarakat yang semakin banyak memelihara hewan selama masa pandemi.

Dalam mengembangkan bisnisnya, Bay Bay Company memanfaatkan pendekatan pemasaran digital dan distribusi online melalui beberapa platform marketplace serta situs resmi perusahaan. Strategi ini bertujuan meningkatkan jangkauan konsumen di seluruh Indonesia sekaligus memperkuat kepercayaan pelanggan terhadap keaslian produk yang dijual secara resmi. Menurut Bay Bay Company (2025), kehadiran toko resmi menjadi bagian penting dalam memastikan konsumen mendapatkan informasi yang akurat, termasuk spesifikasi produk, manfaat, dan cara penggunaan. Dengan demikian, keberadaan Bay Bay Company berkontribusi dalam menyediakan akses produk perlengkapan hewan dan perjalanan bagi masyarakat secara lebih luas dan aman.