



Perbandingan Algoritma K-Nearest Neighbor dan Naive Bayes Untuk Klasifikasi Status Penjualan Furniture dengan Data Historis

Bangkit Dwi Sucahyo, Sampurna Dadi Riskiono*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung

Jl. ZA. Pagar Alam No.9 -11, Labuhan Ratu, Kec. Kedaton, Kota Bandar Lampung, Lampung, Indonesia

Email: ¹bangkit_dwi_sucahyo@teknokrat.ac.id, ^{2,*}sampurna.go@teknokrat.ac.id

Email Penulis Korespondensi: sampurna.go@teknokrat.ac.id

Submitted: 10/11/2025; Accepted: 31/01/2026; Published: 31/01/2026

Abstrak—Penelitian ini bertujuan untuk mengevaluasi seberapa baik kinerja dua metode klasifikasi pembelajaran mesin, K-Nearest Neighbor (KNN) dan Naïve Bayes, dalam memprediksi status penjualan produk furnitur di CV. Surya Gemilang. Data yang digunakan berasal dari catatan penjualan sebelumnya dan mencakup detail seperti kategori produk, nama produk, harga, jumlah penjualan, pendapatan, dan status penjualan, yang diberi label "Laris" dan "Tidak Laris". Penelitian ini mengikuti beberapa langkah, termasuk pengumpulan data, pembersihan dan pengorganisasian data, pemberian label, pelatihan model, dan penilaian kinerjanya menggunakan akurasi, presisi, recall, dan skor F1. Proses penelitian meliputi preprocessing data, penanganan missing values, encoding fitur kategorikal, normalisasi fitur numerik, pemisahan data latih dan uji, pelatihan model, serta evaluasi performa menggunakan akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma KNN memperoleh akurasi 97%, presisi 100%, recall 95%, dan F1-score 0,97. Sementara itu, algoritma Naïve Bayes menghasilkan akurasi 85%, presisi 92%, recall 81%, dan F1-score 0,86. Temuan ini mengindikasikan bahwa KNN lebih mampu mengenali pola kompleks dalam data penjualan dibandingkan Naïve Bayes. Kontribusi penelitian ini adalah menyediakan model klasifikasi berbasis machine learning yang dapat digunakan untuk mendukung perencanaan produksi dan strategi pemasaran melalui prediksi tingkat penjualan produk furnitur. Hasilnya menunjukkan bahwa KNN mencapai akurasi 97%, sedangkan Naïve Bayes hanya mencapai 85%. Hal ini menunjukkan bahwa KNN lebih baik dalam mengidentifikasi hubungan kompleks antar fitur dalam data penjualan, sementara Naïve Bayes kurang efektif karena mengasumsikan semua variabel bersifat independen. Singkatnya, KNN lebih efektif dalam mengklasifikasikan status penjualan produk furnitur dan dapat menjadi dasar untuk membuat keputusan bisnis yang terinformasi berdasarkan data. Penelitian ini memberikan kontribusi yang berarti terhadap penerapan pembelajaran mesin di perusahaan kecil dan menengah, membantu meningkatkan perkiraan penjualan dan mengembangkan strategi pemasaran yang lebih efektif.

Kata Kunci: K-Nearest Neighbor; Naive Bayes; Prediksi Penjualan; Data Mining; Machine Learning

Abstract—This study aims to evaluate the performance of two machine learning classification methods, K-Nearest Neighbor (KNN) and Naive Bayes, in predicting the sales status of furniture products at CV. Surya Gemilang. The data used comes from previous sales records and includes details such as product category, product name, price, sales amount, revenue, and sales status, which are labeled "Best Selling" and "Not Selling". This study follows several steps, including data collection, data cleaning and organization, labeling, model training, and performance assessment using accuracy, precision, recall, and F1-score. The research process includes data preprocessing, handling missing values, encoding categorical features, normalizing numeric features, separating training and testing data, model training, and performance evaluation using accuracy, precision, recall, and F1-score. The results show that the KNN algorithm achieves 97% accuracy, 100% precision, 95% recall, and 0.97 F1-score. Meanwhile, the Naive Bayes algorithm achieved 85% accuracy, 92% precision, 81% recall, and an F1-score of 0.86. These findings indicate that KNN is better able to recognize complex patterns in sales data than Naive Bayes. The contribution of this research is to provide a machine learning-based classification model that can be used to support production planning and marketing strategies by predicting furniture product sales levels. The results show that KNN achieved 97% accuracy, while Naive Bayes only achieved 85%. This indicates that KNN is better at identifying complex relationships between features in sales data, while Naive Bayes is less effective because it assumes all variables are independent. In summary, KNN is more effective in classifying furniture product sales status and can be the basis for making informed business decisions based on data. This research makes a significant contribution to the application of machine learning in small and medium-sized enterprises, helping to improve sales forecasting and develop more effective marketing strategies.

Keywords: K-Nearest Neighbor; Naive Bayes; Sales Prediction; Data Mining; Machine Learning

1. PENDAHULUAN

Teknologi informasi telah membawa perubahan besar dalam proses pengambilan keputusan pada industri perdagangan dan penjualan. Perusahaan tidak lagi hanya mengandalkan intuisi, tetapi memerlukan analisis data historis untuk merencanakan strategi pemasaran, mengelola inventaris, dan meningkatkan profitabilitas. Untuk tujuan tersebut, data mining dan machine learning menjadi alat penting untuk memahami pola dan tren dalam data penjualan. Salah satu penerapannya adalah memprediksi tingkat penjualan produk, sehingga perusahaan dapat memperkirakan permintaan pasar secara lebih akurat. CV. Surya Gemilang, sebagai perusahaan penjualan furnitur, menghadapi tantangan dalam memahami dinamika permintaan pelanggan yang berubah-ubah sesuai musim dan tren. Prediksi penjualan merupakan bagian penting dari perencanaan bisnis karena membantu perusahaan mengetahui apa yang dibutuhkan pasar.

Melalui pendekatan klasifikasi, perusahaan dapat mengidentifikasi produk yang berpotensi laris dan tidak laris, sehingga pengelolaan stok dan distribusi menjadi jauh lebih efisien dan risiko penumpukan barang dapat diminimalkan. Hal ini membantu meningkatkan efisiensi produksi dan distribusi serta mengurangi risiko stok yang



terlalu banyak atau terlalu sedikit. Karena penjualan furnitur dapat berubah seiring musim dan tren konsumen, metode pembelajaran mesin seperti K-Nearest Neighbor (KNN) dan Naive Bayes sangat berguna[1]. KNN merupakan cara yang sederhana namun efektif untuk mengklasifikasikan data. Metode ini bekerja dengan mengelompokkan titik-titik data berdasarkan seberapa dekat mereka dalam ruang multidimensi. Manfaat utama KNN adalah kemampuannya untuk menangani data yang tidak langsung dan tidak mengasumsikan pola data tertentu. Dalam prediksi penjualan, KNN dapat menemukan produk yang memiliki pola penjualan serupa berdasarkan faktor-faktor seperti harga, jenis material, dan kategori produk[2]. Namun, salah satu kekurangannya adalah pemrosesan data dalam jumlah besar dapat memakan waktu lebih lama. Di sisi lain, Naive Bayes adalah metode yang menggunakan probabilitas dan Teorema Bayes, dan mengasumsikan bahwa fitur-fitur bersifat independen. Meskipun asumsi ini tidak selalu benar di dunia nyata, Naive Bayes masih berfungsi dengan baik dalam banyak tugas klasifikasi[3], [4]. Dalam prediksi penjualan, Naive Bayes dapat menghitung peluang suatu produk berada dalam kelompok "laris" atau "tidak laris" berdasarkan probabilitas atributnya. Selain itu, kecepatannya yang tinggi dan kemampuannya dalam menangani big data menjadikannya pilihan populer untuk analisis prediktif.

Perbandingan kedua algoritma ini penting untuk mengetahui metode mana yang paling sesuai dengan kebutuhan perusahaan di bidang visi komputer data penjualan dari CV. Surya Gemilang digunakan di sini. Setiap algoritma memiliki kekuatan dan kelemahannya sendiri, yang bergantung pada jenis data yang digunakannya. Hal-hal seperti jumlah fitur, seberapa banyak noise dalam data, bagaimana kelas-kelas disebar, dan seberapa besar dataset semuanya memainkan peran besar dalam seberapa baik kinerja algoritma. Karena itu, studi ini melihat seberapa baik KNN dan Naive Bayes dalam memprediksi kategori penjualan produk furnitur menggunakan data historis perusahaan. Penelitian sebelumnya menunjukkan bahwa KNN biasanya bekerja dengan baik dengan data yang memiliki pola kompleks, sementara Naive Bayes bekerja dengan baik secara konsisten dengan data di mana fitur-fiturnya tersebar secara merata. Menurut sebuah studi dengan judul "K-Nearest Neighbor and Naive Bayes Algorithm Approach for Online Sales Level Classification Optimization In South Tangerang", KNN mencapai akurasi 89,26% dalam memprediksi penjualan e-commerce, sementara Naive Bayes mencapai hasil 90,91% pada dataset yang sama[5]. Ini menunjukkan bahwa memilih algoritma terbaik benar-benar tergantung pada karakteristik data.

Dalam bisnis furnitur, faktor-faktor seperti jenis produk, harga jual, frekuensi pembelian produk, dan musim penjualan berkaitan erat dengan tingkat penjualan. Oleh karena itu, prediksi yang akurat dapat membantu CV. Surya Gemilang membuat keputusan yang lebih baik terkait stok, promosi, dan pembelian bahan baku. Sebuah studi dengan judul "Analisis Perbandingan Algoritma Machine Learning untuk Forecasting Persediaan Produk Barang Pokok" menemukan bahwa penggunaan machine learning dalam ritel dapat meningkatkan manajemen inventaris hingga 28%[6]. Meskipun penelitian sebelumnya menjelaskan keberhasilan KNN dan Naive Bayes dalam konteks prediksi penjualan, terdapat beberapa celah penelitian yang belum banyak disorot. Sebagian besar penelitian terdahulu berfokus pada e-commerce, barang pokok, dan retail umum, bukan pada bisnis furnitur fisik yang memiliki karakteristik penjualan musiman, harga tinggi, dan variasi produk yang besar. Selain itu, sebagian besar studi hanya melaporkan akurasi model tanpa mengevaluasi performa secara komprehensif menggunakan metrik lain seperti presisi, recall, dan F1-score. Pada banyak penelitian, dataset juga tidak dipisahkan berdasarkan kategori penjualan atau variabel yang mempengaruhi perilaku konsumen, sehingga rekomendasi algoritma terbaik belum sepenuhnya kontekstual dalam pengambilan keputusan bisnis. Selain itu, penggunaan teknik seperti validasi silang dan matriks kebingungan penting untuk memeriksa kinerja model. Ukuran seperti akurasi, presisi, recall, dan F1-score digunakan untuk membandingkan efektivitas kedua algoritma tersebut.

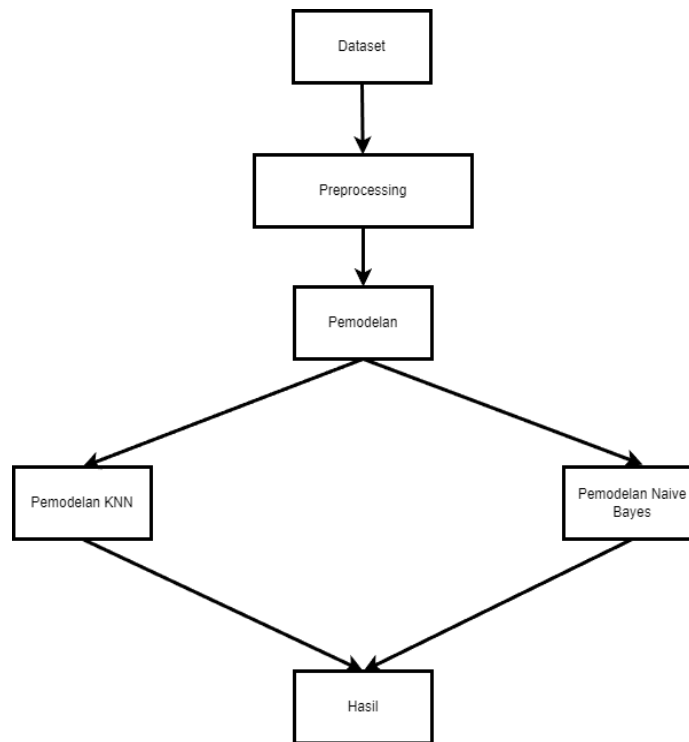
Penelitian ini membantu membangun model klasifikasi yang menggunakan pembelajaran terawasi, dengan fokus pada usaha kecil dan menengah. Penelitian ini membandingkan dua metode tradisional yang masih umum digunakan saat ini, yang dapat membantu pengembang menciptakan sistem AI untuk prediksi penjualan yang lebih baik beradaptasi dengan perubahan perilaku konsumen[7], [8]. Dengan menggunakan analitik prediktif pada usaha kecil dan menengah, penelitian ini meningkatkan akurasi prakiraan penjualan[9]. Hasilnya, penelitian ini memiliki dampak besar di bidang ekonomi mikro. CV. Surya Gemilang dapat menggunakan hasilnya untuk mengelola inventaris, merencanakan promosi, dan memprediksi penjualan di masa mendatang. Selain itu, penggunaan metode pembelajaran mesin yang efisien mendukung gagasan pengambilan keputusan bisnis berdasarkan data, yang kini menjadi bagian penting dalam transformasi manufaktur dan perdagangan menjadi sistem digital. Dengan menggunakan KNN dan Naive Bayes, perusahaan dapat memahami perilaku pelanggan dan menciptakan strategi bisnis yang lebih sesuai dengan keinginan pasar. Hal ini sejalan dengan temuan yang menyoroti pentingnya algoritma klasifikasi dalam meningkatkan manajemen rantai pasokan. Secara keseluruhan, perbandingan kedua algoritma ini akan diuji menggunakan data penjualan nyata yang mencakup angka dan kategori. Proses ini meliputi persiapan data, pemisahan dataset, pelatihan model, dan penilaian kinerja menggunakan ukuran standar. Hasilnya akan dianalisis untuk menentukan algoritma mana yang paling akurat dan paling baik dalam menangani data baru. Studi ini bertujuan untuk memberikan gambaran yang jelas tentang seberapa efektif KNN dan Naive Bayes dalam memprediksi penjualan furnitur di CV. Surya Gemilang. Temuan ini diharapkan dapat berkontribusi pada pengetahuan akademis di bidang pembelajaran mesin terapan dan menawarkan manfaat praktis bagi bisnis yang ingin meningkatkan strategi mereka melalui analisis berbasis data.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif yang didasarkan pada machine learning untuk membandingkan efektivitas dua algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Naive Bayes, dalam mengklasifikasikan data penjualan produk di CV. Surya Gemilang. Pendekatan machine learning banyak digunakan dalam analisis prediktif karena kemampuannya dalam menemukan pola tersembunyi dari data historis yang besar [10], [11]. Tujuan dari penelitian ini adalah mengetahui algoritma mana yang mampu memberikan prediksi kategori penjualan produk dengan tingkat akurasi lebih tinggi, seperti produk yang laris dan tidak laris, berdasarkan data penjualan sebelumnya.

Proses penelitian mencakup beberapa tahap sebagaimana di ilustasikan pada Gambar 1., yaitu mengumpulkan data, mempersiapkan data sebelum diproses, membagi data menjadi bagian latihan dan uji, menerapkan kedua algoritma tersebut, serta mengevaluasi hasilnya dengan menggunakan berbagai metrik seperti akurasi, presisi, recall, dan F1-score. Evaluasi menggunakan metrik tersebut penting agar hasil yang diperoleh dapat dibandingkan secara objektif. Harapan dari hasil penelitian ini adalah memberikan rekomendasi algoritma yang paling tepat untuk membantu mengambil keputusan strategis dalam meningkatkan penjualan dan efisiensi bisnis di CV. Surya Gemilang, sebagaimana pendekatan serupa telah berhasil diterapkan dalam sektor ritel dan e-commerce [12].



Gambar 1. Tahapan Penelitian

Dengan menerapkan metode yang sudah digunakan sebelumnya, penelitian ini dilakukan dalam beberapa tahap, yaitu mulai dari mengumpulkan data, mempersiapkan data sebelum diproses, melatih model, menguji model, serta mengevaluasi hasilnya menggunakan beberapa ukuran seperti accuracy, precision, recall, dan F1-score. Pendekatan ini diharapkan mampu menghasilkan model yang dapat mengklasifikasikan penjualan furnitur dengan akurasi tinggi dan performa yang stabil.

2.2 Pengumpulan Data

Langkah pertama dalam penelitian ini adalah pengumpulan data yang dilakukan secara sistematis melalui wawancara langsung dan peninjauan dokumen internal dari perusahaan. Tujuan dari proses ini adalah mendapatkan informasi primer yang tepat dan relevan dengan topik penelitian, agar hasil analisis yang diperoleh bisa didasari dengan cukup kuat. Data yang dikumpulkan mencakup berbagai hal penting seperti jenis produk furnitur yang dijual, harga jual setiap produk, jumlah penjualan dalam waktu tertentu, serta kondisi penjualan yang dibagi menjadi kategori laris dan tidak laris. Tahapan pengumpulan data ini merupakan dasar utama bagi seluruh proses penelitian, karena kualitas dan kelengkapan data mentah sangat berpengaruh pada keandalan, konsistensi, dan akurasi hasil analisis di tahap selanjutnya [13]. Pada data penjualan ini terdapat 500 data penjualan dengan 6 atribut utama serta dengan status penjualan yang menunjukkan status klasifikasinya 0 menunjukkan barang tidak laris dan 1 menunjukkan barang laris seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Dataset Atribut Penelitian

Atribut	Deskripsi
Kategori	Jenis produk yang dijual, misalnya Kursi, Lemari, Rak, dan sebagainya.
Nama Produk	Nama spesifik produk yang dijual, seperti Kursi Bar Kayu, Lemari Anak Kayu.
Harga	Harga satuan produk
Jumlah Terjual	Jumlah unit produk yang terjual selama periode tertentu.
Pendapatan	Total pendapatan dari penjualan produk
Status Penjualan	Kategori hasil klasifikasi penjualan, yaitu “Laris” atau “Tidak Laris”.

Berdasarkan Tabel 1, semua atribut dalam dataset penelitian dijelaskan Kategori menunjukkan jenis produk yang dijual, seperti kursi, lemari, rak, dan lainnya, Nama Produk menunjukkan nama spesifik dari produk yang dijual, seperti kursi kayu atau lemari kayu untuk anak, Harga menunjukkan harga per unit produk, Jumlah Terjual mencerminkan banyaknya produk yang terjual dalam periode tertentu, Pendapatan menunjukkan total uang yang diperoleh dari penjualan produk; dan Status Penjualan adalah kategori hasil dari klasifikasi penjualan, yaitu “Laris” atau “Tidak Laris”.

2.3 Preprocessing

Preprocessing merupakan langkah pertama yang harus dilakukan untuk pembersihan data atau data cleaning. Pembersihan dan transformasi data merupakan bagian dari tahap preprocessing yang bertujuan untuk meningkatkan akurasi dan relevansi keluaran dari pemodelan kasifikasi[14]. Pada tahap ini akan mengubah data menjadi bentuk yang akan dianalisis menggunakan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes

2.4 Penanganan Missing Values

Pemeriksaan missing values dilakukan untuk mengidentifikasi atribut yang memiliki nilai kosong atau tidak terisi. Jika ditemukan nilai kosong dilakukan strategi penanganan dengan mengisi nilai kosong mean, median, atau modus sesuai jenis data atau menghapus baris atau atribut tertentu jika presentase missing values terlalu besar.

2.5 Transformasi Fitur Kategorikal

Data kategorikal seperti jenis produk, kategori atau nama produk tidak dapat langsung diproses oleh algoritma. Oleh karena itu dilakukan encoding dengan menggunakan metode one-hot encoding untuk kategori produk dan label encoding untuk variabel yang memiliki kategori tetapi tidak memiliki hubungan hierarkis transformasi ini memastikan bahwa fitur dapat diolah secara matematis oleh algoritma klasifikasi

2.6 Normalisasi atau Standarisasi Fitur numerik

Fitur numerik seperti harga, jumlah terjual, dan pendapatan dinormalisasi agar memiliki skala yang seragam. Hal ini sangat penting terutama untuk algoritma KNN yang berbasis perhitungan jarak. Metode yang digunakan dapat berupa Min-Max Normalization, atau Standardization (Z-score). Dengan proses ini, tidak ada satu fitur numerik pun yang mendominasi proses klasifikasi akibat perbedaan skala.

2.7 Klasifikasi K-Nearest Neighbor (KNN)

Metode K-Nearest Neighbors (KNN) adalah salah satu algoritma klasifikasi yang mudah namun sangat berguna dalam mengenali pola dan menganalisis data. Algoritma ini bekerja dengan cara membandingkan data baru dengan data-data lain yang sudah diketahui kelasnya, berdasarkan jarak atau tingkat kesamaan. Artinya, data baru akan diberi label sesuai dengan kelas yang paling dominan di antara tetangga terdekatnya. Untuk mengukur tingkat kesamaan tersebut, biasanya digunakan rumus jarak Euclidean seperti ini.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$

$d(x, y)$ = Jarak antara data x dan y , n = Jumlah fitur, x_i = Nilai fitur ke- i pada data uji, y_i = Nilai fitur ke- i pada data latih. Dalam rumus tersebut, $d(x, y)$ menunjukkan jarak antara dua buah data, yaitu data yang sedang diuji dan data yang sudah dilatih. Nilai x_i dan y_i masing-masing mewakili nilai dari atribut ke- i pada kedua data yang dibandingkan, sedangkan n adalah jumlah total atribut atau variabel yang digunakan. Semakin kecil nilai jarak yang diperoleh, semakin mirip data tersebut dengan data latih yang sudah diketahui kelasnya. Setelah jarak dihitung untuk semua data latih, algoritma KNN kemudian memilih K data yang memiliki jarak paling dekat, lalu menentukan kelas yang paling banyak muncul dari data-data tersebut. Proses penentuan kelas akhir ditulis dalam persamaan.

$$C(x) = \text{mode}\{y_1, y_2, \dots, y_K\} \tag{2}$$

$C(x)$ = Kelas hasil prediksi, $y_1, y_2 \dots y_K$ = label dari K tetangga terdekat. Dengan pendekatan sederhana ini, KNN terbukti efektif dalam berbagai bidang, termasuk klasifikasi produk penjualan furnitur berdasarkan status penjualan seperti “laris” dan “tidak laris”. Menurut Ismail Marzuki, Norlida Hassan (2020) Keunggulan utama KNN adalah kemampuannya yang fleksibel dan mampu memberikan hasil yang presisi tanpa harus



mengasumsikan bentuk distribusi data tertentu. Namun, kekurangannya adalah membutuhkan waktu komputasi yang lama ketika jumlah data latih sangat besar[15].

2.8 Klasifikasi Naive Bayes

Metode Naive Bayes adalah algoritma pengklasifikasian yang menggunakan dasar probabilitas dan menerapkan Teorema Bayes dengan mengasumsikan bahwa setiap fitur dalam data saling independen. Algoritma ini bekerja dengan cara menghitung seberapa besar kemungkinan suatu data masuk ke dalam suatu kelas tertentu berdasarkan nilai-nilai fitur yang dimilikinya. Rumus umum yang digunakan dalam metode Naive Bayes adalah sebagai berikut:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (3)$$

$P(C|X)$ = probabilitas kelas C diberikan data X (yang dicari), $P(X|C)$ = probabilitas fitur X jika kelas C benar, $P(C)$ = Probabilitas awal (prior) kelas C, $P(X)$ = probabilitas data X. $P(C|X)$ adalah kemungkinan kelas C terjadi berdasarkan data X, $P(X|C)$ adalah kemungkinan munculnya fitur X jika kelas C diketahui, dan $P(C)$ adalah kemungkinan awal dari kelas tersebut. Karena setiap fitur dianggap tidak saling tergantung, maka peluang gabungannya bisa dihitung dengan mengalikan setiap peluang fitur secara terpisah.

$$P(C|X) = \frac{P(C) \prod_{i=1}^n P(x_i|C)}{P(X)} \quad (4)$$

Hasil akhir dari klasifikasi ditentukan oleh kelas dengan nilai probabilitas terbesar. Secara matematis, proses pengambilan keputusan ini dapat dirumuskan sebagai:

$$C^* = \operatorname{argmax}_c P(C) \prod_{i=1}^n P(x_i|C) \quad (5)$$

Metode Naive Bayes banyak digunakan karena kemampuannya dalam memberikan hasil klasifikasi yang cepat dan akurat meskipun pada dataset besar dengan banyak fitur. Menurut penelitian “Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai” Naive Bayes juga baik digunakan untuk menangani data yang berupa kategori atau angka[16]. Dalam penelitian tentang penjualan furnitur, algoritma ini sangat membantu untuk membagi produk berdasarkan tingkat penjualannya, seperti menentukan apakah suatu produk masuk ke dalam kategori best seller, kurang laris, atau memiliki potensi. Menurut penelitian “Klasifikasi Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor dan Metode Naive Bayes” KNN dan Naive Bayes sering digunakan untuk meningkatkan presisi klasifikasi karena kedua metode tersebut saling melengkapi KNN lebih baik dalam mengenali pola dari data lokal, sedangkan Naive Bayes lebih baik dalam menghitung probabilitas antar fitur secara global[17].

2.9 Evaluasi Model Klasifikasi

Dalam studi yang menggunakan algoritma klasifikasi, K-Nearest Neighbor dan Naive Bayes memeriksa seberapa baik kinerja model sangat penting untuk memahami seberapa baik sistem dalam membuat prediksi. Salah satu cara populer untuk memeriksa hasil klasifikasi adalah dengan melihat keempat metrik ini akurasi, presisi, recall, dan F1-score. Keempat metrik ini masing-masing membantu dengan cara yang berbeda untuk menunjukkan seberapa baik suatu model dapat memprediksi hal-hal yang sebenarnya benar.

a. Accuracy

Akurasi adalah cara dasar untuk memeriksa seberapa sering suatu model memberikan prediksi yang tepat untuk semua data uji. Rumus untuk akurasi ditulis seperti ini:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$TP(\text{True Positive})$ = Jumlah data positif yang diprediksi benar $TN(\text{True Negative})$ = Jumlah data negatif yang diprediksi benar $FP(\text{False Positive})$ = Jumlah data negatif yang salah diprediksi positif $FN(\text{False Negative})$ = Jumlah data positif yang salah diprediksi negatif. Nilai akurasi menunjukkan berapa banyak prediksi yang benar yang dihasilkan model dibandingkan dengan semua prediksi yang dihasilkannya. Akurasi yang tinggi berarti model tersebut mampu memprediksi kategori suatu data dengan baik. Namun, akurasi saja tidak selalu mencerminkan keseluruhan data, terutama ketika data tidak seimbang, yang berarti satu kelompok memiliki lebih banyak data daripada kelompok lainnya[18]. Dalam situasi seperti itu, model tersebut mungkin tampak akurat karena kelas mayoritas sangat umum, tetapi mungkin tidak bagus dalam menemukan kelas minoritas.

b. Precision

Presisi menunjukkan seberapa akurat suatu model ketika memprediksi hasil positif. Rumus presisi adalah:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

Presisi berkaitan dengan seberapa akurat model memprediksi sesuatu yang positif. Semakin tinggi presisinya, semakin kecil kemungkinan model salah melabeli sesuatu yang negatif sebagai positif. Presisi sangat penting

karena prediksi positif yang salah dapat menyebabkan masalah besar. Misalnya, jika email yang sebenarnya penting ditandai sebagai spam, informasi penting mungkin hilang[19].

c. Recall

Recall, yang juga disebut sensitivitas, menunjukkan seberapa baik suatu model dapat menemukan semua kasus positif yang sebenarnya. Rumus untuk recall ditulis seperti ini:

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Nilai recall yang tinggi berarti model dapat menemukan hampir semua kasus positif yang sebenarnya. Dalam situasi seperti mendiagnosis penyakit, recall sangat penting karena model perlu mengidentifikasi sebanyak mungkin pasien sakit yang sebenarnya, meskipun model mungkin juga salah melabeli beberapa orang sehat sebagai sakit[20]. Jadi, mengingat kembali sangat penting ketika melakukan penelitian yang berfokus pada penemuan kasus berisiko tinggi, karena kegagalan dalam mengidentifikasi kasus positif dapat menimbulkan masalah besar.

d. F1-Score

Untuk mendapatkan keseimbangan yang lebih baik antara presisi dan perolehan, skor F1 digunakan. Skor ini merupakan rata-rata harmonis dari presisi dan perolehan. Rumusnya adalah:

$$F1 - Score = 2x \frac{Precision \times Recall}{Precision+Recall} \quad (9)$$

F1-Score berkisar dari 0 hingga 1, dan semakin mendekati 1, semakin baik keseimbangan antara presisi dan perolehan. Skor ini sangat membantu ketika mengevaluasi data yang tidak seimbang, karena tidak terlalu dipengaruhi oleh satu kelas yang lebih umum daripada yang lain. Jadi, F1-Score adalah cara yang lebih baik untuk menunjukkan kinerja dibandingkan dengan hanya melihat akurasi saja[21], [22].

3. HASIL DAN PEMBAHASAN

3.1 Import data

Penelitian ini memanfaatkan data penjualan yang diperoleh dari CV. Surya Gemilang, sebuah perusahaan yang berlokasi di Pekon Wates Timur, Jl. Sawah, Tambahmulyo, Kecamatan Gadingrejo, Kabupaten Pringsewu, 35372. Data penjualan tersebut menjadi sumber utama dalam analisis, karena mencakup informasi penting terkait kategori produk, jumlah penjualan, serta variabel lain yang relevan dengan tujuan penelitian. Seluruh proses pengolahan data dilakukan menggunakan platform Google Colab, yang dipilih karena mampu menyediakan lingkungan komputasi berbasis cloud dengan dukungan berbagai library analisis data, sehingga memudahkan peneliti dalam melakukan eksplorasi, pembersihan, hingga pemodelan data secara efisien dan terstruktur.

```
Data Awal:
  Kategori      Produk      Harga  Jumlah_Terjual  Pendapatan \
0  Kursi        Kursi Bar Kayu  8074188      31      250299828
1  Lemari       Lemari Anak Kayu  3833224      18      68998032
2  Kursi        Kursi Lipat Aluminium  3693480      21      77563080
3  Rak          Rak Gantung Dapur  5000115      30      150003450
4  Rak          Rak Gantung Dapur  5000115      30      150003450

Status_Penjualan
0  Laris
1  Tidak Laris
2  Tidak Laris
3  Tidak Laris
4  Tidak Laris
```

Gambar 2. Import data

Gambar 2 menunjukkan bagian awal penelitian, data penjualan diimpor menggunakan Python dan pustaka Pandas. Tujuannya adalah membaca dataset penjualan furnitur dari berkas "data penjualan CV Surya Gemilang.xlsx". Berkas ini berisi detail penting seperti kategori produk, nama produk, harga, jumlah penjualan, total pendapatan, dan status penjualan, yang dapat berupa laris dan tidak laris. Kode yang digunakan adalah `pd.read_excel('data jual cv surya gemilang.xlsx')`, yang mengubah berkas Excel menjadi kerangka data untuk analisis lebih lanjut. Kemudian, perintah `data.head()` digunakan untuk menampilkan lima baris pertama data guna memastikan data terbaca dengan baik dan terstruktur dengan benar.

3.2 Pemeriksaan struktur data

Setelah impor data berhasil, langkah selanjutnya adalah memeriksa struktur dan jenis data menggunakan fungsi `data.info()`. Langkah ini membantu memastikan semua atribut dalam dataset terbaca dengan benar dan tidak ada nilai yang hilang.

```
*** <class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 6 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   Kategori            500 non-null    object
1   Nama Produk         500 non-null    object
2   Harga               500 non-null    int64
3   Jumlah Terjual     500 non-null    int64
4   Pendapatan          500 non-null    int64
5   Status Penjualan   500 non-null    object
dtypes: int64(3), object(3)
memory usage: 23.6+ KB

Cek data kosong:
Kategori            0
Nama Produk         0
Harga               0
Jumlah Terjual     0
Pendapatan          0
Status Penjualan   0
dtype: int64
```

Gambar 3. Pemeriksaan Struktur data

Melalui hasil pemeriksaan struktur data pada Gambar 3, terlihat bahwa dataset yang digunakan dalam penelitian ini terdiri dari enam atribut utama, yaitu Kategori, Nama Produk, Harga, Jumlah Terjual, Pendapatan, dan Status Penjualan. Setiap kolom pada dataset tersebut menunjukkan jumlah entri yang terisi secara lengkap, sehingga dapat dipastikan bahwa tidak terdapat data yang hilang (missing values) yang berpotensi mengganggu proses analisis. Kondisi ini memungkinkan proses pengolahan data dilakukan secara optimal tanpa memerlukan tahapan imputasi atau penanganan nilai kosong. Selanjutnya, untuk keperluan pelatihan dan pengujian model, dataset dibagi menjadi dua bagian dengan proporsi 80:20, di mana 80% data digunakan sebagai data pelatihan (training set) dan 20% sisanya dialokasikan sebagai data pengujian (testing set). Pembagian ini bertujuan untuk memastikan bahwa model dapat belajar secara efektif sekaligus diuji menggunakan data yang tidak pernah dilihat sebelumnya, sehingga menghasilkan evaluasi performa yang lebih objektif.

3.3 Pelabelan data

Tahap berikutnya dalam penelitian ini adalah konversi data kategorikal menjadi data numerik untuk memastikan seluruh variabel dapat diproses oleh algoritma machine learning.

```
Categorical columns successfully converted to numeric!
```

	Kategori	Produk	Harga	Jumlah_Terjual	Pendapatan	Status_Penjualan
0	0	1	8074188	31	250299828	0
1	1	7	3833224	18	68998032	1
2	0	4	3693480	21	77563080	1
3	3	24	5000115	30	150003450	1
4	3	24	5000115	30	150003450	1

Gambar 4. Pelabelan

Gambar 4 menampilkan tahapan transformasi data yang dilakukan sebelum proses pemodelan machine learning. Pada tahap pertama, data asli disalin untuk memastikan proses manipulasi tidak mengubah dataset utama. Selanjutnya, variabel Status_Penjualan yang merupakan variabel target dikonversi menggunakan Label Encoding. Teknik ini mengubah setiap kategori pada variabel target menjadi nilai numerik, sehingga model dapat menginterpretasikan data tersebut tanpa kehilangan informasi kategorikal. Tahap berikutnya adalah melakukan One-Hot Encoding pada dua fitur kategorikal, yaitu Kategori dan Produk. Penggunaan One-Hot Encoder dengan parameter drop="first" bertujuan untuk mengurangi redundansi dan menghindari fenomena multicollinearity yang umum terjadi akibat keberadaan dummy variables berlebih. Proses encoding ini menghasilkan sejumlah fitur baru yang direpresentasikan dalam bentuk nilai biner (0 dan 1), di mana setiap kolom menunjukkan keberadaan atau ketiadaan kategori tertentu pada suatu entri data. Setelah fitur-fitur baru tersebut terbentuk, DataFrame hasil encoding kemudian digabungkan kembali dengan bagian data numerik yang telah dipertahankan, seperti Harga, Jumlah_Terjual, dan Pendapatan. Pada bagian bawah gambar terlihat hasil akhir dari proses ini, yaitu dataset yang kini sepenuhnya berada dalam format numerik. Setiap baris data mencerminkan representasi lengkap dari variabel asli, baik numerik maupun kategorikal yang telah diencode. Tahap ini merupakan bagian penting dalam pipeline pra-pemrosesan, karena algoritma machine learning pada umumnya hanya dapat memproses data numerik. Dengan transformasi ini, dataset menjadi lebih siap untuk digunakan dalam tahap pelatihan model, evaluasi performa, serta proses prediksi selanjutnya.

3.4 Pemodelan K-Nearest Neighbor

Pada tahap ini, model klasifikasi dibuat menggunakan algoritma K-Nearest Neighbors (KNN). Metode ini dipilih karena memiliki keunggulan dalam memprediksi kategori data baru dengan melihat seberapa dekatnya kategori tersebut dengan titik-titik data dalam set pelatihan yang sudah memiliki kategori yang diketahui. Model dimulai dengan menetapkan parameter K ke 5, yang berarti setiap prediksi untuk sampel uji didasarkan pada lima contoh terdekat dari data pelatihan. Nilai K ini dapat diubah nanti, tergantung pada seberapa baik kinerja model untuk mendapatkan hasil yang paling akurat.

3.5 Pemodelan Naïve Bayes

Proses pembuatan model klasifikasi menggunakan algoritma Naive Bayes, yang merupakan salah satu jenis Gaussian Naive Bayes. Metode ini merupakan pendekatan probabilistik dalam pembelajaran mesin yang umum digunakan untuk klasifikasi karena cepat dan berkinerja baik. Langkah pertama adalah membuat objek model menggunakan perintah `model = GaussianNB()`. Kemudian, model dilatih menggunakan fungsi `fit()`, dengan `X_train` adalah kumpulan fitur dan `y_train` adalah label target atau kelas. Selama pelatihan ini, model Naive Bayes mempelajari bagaimana data didistribusikan menggunakan Teorema Bayes. Model ini mengasumsikan bahwa setiap fitur memengaruhi variabel target secara independen. Metode Gaussian dipilih karena data dalam studi ini mencakup nilai numerik yang diyakini mengikuti distribusi normal, seperti kolom Harga, Jumlah Terjual, dan Pendapatan. Proses pelatihan ini memungkinkan model untuk menghitung probabilitas setiap kelas seperti laris dan tidak laris, berdasarkan fitur input.

3.6 Hasil Pemodelan 2 Algoritma

Tahapan pemodelan dilakukan dengan menggunakan 2 algoritma klasifikasi untuk membandingkan kemampuan masing-masing model dalam memprediksi penjualan laris dan tidak laris. Setiap model dinilai memanfaatkan metrik akurasi, presisi, recall dan F1-score untuk mengevaluasi performa secara menyeluruh.

a. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) Menunjukkan Tingkat akurasi sebesar 97,00%. Ini berarti model sangat baik dalam mengurutkan data dengan benar. Laporan klasifikasi juga menunjukkan bahwa kedua kelas, Laris dan Tidak Laris, memiliki skor yang tinggi dan serupa untuk precision, recall, dan F1-score. Untuk kelas Laris, precision 100%, recall 95%, dan F1-score 97%. Untuk kelas Tidak Laris, precision 93%, recall 100%, dan F1-score 97%.

b. Naïve Bayes

Naïve Bayes menunjukkan Tingkat akurasi sebesar 85,00% Nilai ini menunjukkan bahwa sekitar 85% prediksi model pada data uji akurat dibandingkan dengan label sebenarnya. Laporan klasifikasi menunjukkan bahwa untuk kelas 0, yang merupakan produk Laris, model tersebut akurat precision 92%, recall 81%, F1-score sebesar 86%. Untuk kelas 1, yang merupakan produk "Tidak Laris", model tersebut akurat precision 78%, recall 91%, F1-score sebesar 84%. Rata-rata makro dan rata-rata tertimbang keduanya berada di sekitar 85%, yang berarti model berkinerja baik secara keseluruhan, tetapi terdapat sedikit perbedaan akurasi antara kedua kelas tersebut.

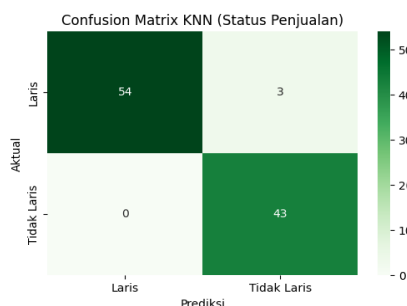
Perbandingan performa dua algoritma klasifikasi bisa terlihat pada gambar 7 dalam memprediksi produk penjualan CV. Surya Gemilang yaitu K-Nearest Neighbor (K-NN) dan Naïve bayes berdasarkan parameter evaluasi akurasi, presisi, recall, serta F1-score.

3.7 Evaluasi Confusion Matrix

Peneitian ini mengevaluasi kinerja setiap model yang dikembangkan dengan memanfaatkan metrik klasifikasi yang diperoleh melalui confusion matrix, antara lain accuracy, precision, recall, dan F1-score. Evaluasi ini dilakukan pada dua model yang berbeda untuk menilai pengaruh tuning terhadap performa algoritma.

a. Evaluasi K-Nearest Neighbor (K-NN)

Hasil dari Confusion Matrix memberikan gambaran lebih jelas tentang seberapa baik model K-Nearest Neighbor bekerja dalam mengklasifikasikan data penjualan produk furnitur, khususnya dalam hal apakah produk tersebut laris atau tidak laris.



Gambar 7. Confusion Matrix KNN

Berdasarkan hasil yang ditunjukkan pada Gambar 7, diperoleh nilai prediksi sebagai berikut:

- a) *True Positive*(TP) = 54 artinya sebanyak 54 data yang berhasil diprediksi dengan benar sebagai laris.
- b) *False Negative*(FN) = 3 artinya sebanyak 3 data yang sebenarnya laris namun diprediksi tidak laris.
- c) *False Positive*(FP) = 0 tidak ada data tidak laris yang salah diprediksi sebagai laris.
- d) *True Negative*(TN) = 43 sebanyak 43 data tidak laris berhasil diprediksi sebagai tidak laris.

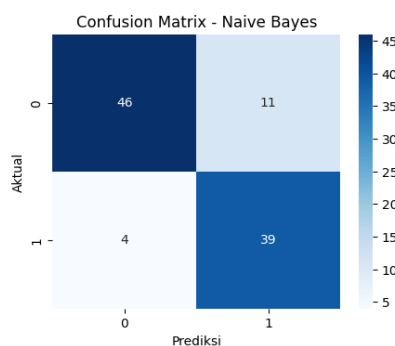
Hasil ini menunjukkan bahwa model KNN sangat akurat, dengan tingkat akurasi 97%, sesuai dengan laporan klasifikasi. Nilai presisi dan recall yang tinggi juga menunjukkan bahwa model ini sangat baik dalam mengenali kedua kelas dan tidak membuat banyak kesalahan. Nilai untuk TP (54) dan TN (43) menunjukkan bahwa model ini berhasil menemukan pola dalam data penjualan, terutama untuk status tidak laris, yang diprediksi dengan tepat tanpa kesalahan (FP = 0). Ini berarti model KNN sangat baik dalam membedakan antara produk populer dan non-populer. Hal ini mungkin karena KNN mengamati seberapa dekat titik-titik data satu sama lain berdasarkan fitur-fitur serupa seperti harga, jenis produk, dan jumlah produk yang terjual.

Tabel 2. Hasil Pemodelan KNN

	precision	recall	F1-score	support
0	1.00	0.95	0.97	57
1	0.93	1.00	0.97	43
Accuracy			0.97	100
Macro avg	0.97	0.97	0.97	100
Wighted avg	0.97	0.97	0.97	100

Tabel 2 menyajikan hasil evaluasi kinerja model K-Nearest Neighbors (KNN) berdasarkan metrik precision, recall, F1-score, dan support untuk masing-masing kelas. Pada kelas 0, model menunjukkan nilai precision sebesar 1.00, yang berarti seluruh prediksi positif untuk kelas ini benar tanpa adanya kesalahan. Nilai recall sebesar 0.95 menunjukkan bahwa 95% data aktual kelas 0 berhasil dikenali dengan tepat oleh model. F1-score sebesar 0.97 mencerminkan keseimbangan yang sangat baik antara precision dan recall pada kelas ini. Untuk kelas 1, precision yang diperoleh sebesar 0.93, yang menandakan masih terdapat sebagian kecil prediksi yang tidak tepat. Namun, nilai recall mencapai 1.00, sehingga seluruh data aktual kelas 1 berhasil diprediksi dengan benar. F1-score pada kelas ini juga mencapai 0.97, menunjukkan performa klasifikasi yang sangat stabil di antara kedua kelas. Secara keseluruhan, model KNN menunjukkan tingkat akurasi yang tinggi, yaitu 0.97 atau 97%. Baik macro average maupun weighted average untuk precision, recall, dan F1-score sama-sama berada pada nilai 0.97, yang mengindikasikan bahwa performa model merata dan tidak bias terhadap salah satu kelas, meskipun distribusi jumlah data (support) kelas 0 dan 1 tidak seimbang sempurna. Hasil ini menunjukkan bahwa model KNN memiliki kemampuan yang sangat baik dalam mengklasifikasikan data penjualan menjadi dua kategori yang ditentukan dalam penelitian, sehingga dapat diandalkan sebagai metode prediksi pada dataset yang digunakan.

b. Evaluasi Confusion Matrix Naïve Bayes



Gambar 8. Confusion Matrix Naïve Bayes

Hasil pengujian menggunakan algoritma Naïve bayes pada data penjualan menghasilkan confusion matrix seperti ditunjukkan pada Gambar 8. Confusion matrix ini menggambarkan kinerja model dalam mengklasifikasikan data penjualan produk berdasarkan status laris dan tidak laris. Dari hasil yang diperoleh dapat dijelaskan bahwa:

- a) *True Positive*(TP) = 39 sebanyak 39 data yang sebenarnya laris berhasil diprediksi dengan benar sebagai tidak laris.
- b) *False Negative*(FN) = 4 terdapat 4 data yang sebenarnya tidak laris tapi diprediksi sebagai data laris.
- c) *False Positive*(FP) = 11 sebanyak 11 data yang sebenarnya laris namun diprediksi sebagai tidak laris.
- d) *True Negative*(TN) = 46 sebanyak 46 data laris berhasil diklasifikasikan dengan benar sebagai laris.

Berdasarkan hasil ini, model Naïve bayes mencapai tingkat akurasi 85%, dengan presisi rata-rata 0,86 dan recall 0,85. Nilai-nilai ini menunjukkan bahwa model tersebut berkinerja cukup baik dalam klasifikasi,



meskipun masih terdapat beberapa kesalahan dalam memprediksi produk laris. Kesalahan klasifikasi terutama terjadi pada kategori laris, dengan 11 contoh kesalahan klasifikasi (False Positive). Hal ini disebabkan oleh asumsi Naive Bayes yang mengasumsikan independensi antar fitur, sementara dalam data penjualan furnitur, mungkin terdapat interdependensi antar variabel seperti harga, volume penjualan, dan kategori produk. Akibatnya, model ini kurang optimal dalam mengenali pola kompleks.

Tabel 3. Hasil Pemodelan Naïve bayes

	Precision	Recall	F1-score	Support
0	0.92	0.81	0.86	57
1	0.78	0.92	0.84	43
Accuracy			0.85	100
Macro avg	0.85	0.86	0.85	100
Wigthed avg	0.86	0.85	0.85	100

Tabel 3 hasil pemodelan Naive Bayes menunjukkan kinerja algoritma dalam mengklasifikasikan dua kategori kelas, yaitu kelas 0 dan kelas 1, berdasarkan metrik evaluasi precision, recall, dan F1-score. Untuk kelas 0, model memperoleh nilai precision sebesar 0.92, yang berarti sebagian besar prediksi model untuk kelas ini benar. Nilai recall sebesar 0.81 menunjukkan bahwa model mampu mengenali mayoritas data aktual kelas 0, meskipun masih terdapat sebagian data yang tidak terdeteksi dengan baik. F1-score kelas 0 sebesar 0.86 menunjukkan keseimbangan yang kuat antara precision dan recall. Pada kelas 1, kinerja model juga cukup baik dengan precision 0.78, artinya sekitar 78% prediksi kelas 1 adalah benar. Nilai recall mencapai 0.92, menunjukkan kemampuan model yang sangat baik dalam menangkap data aktual kelas 1. F1-score sebesar 0.84 mencerminkan stabilitas performa model dalam memprediksi kelas ini. Secara keseluruhan, model menghasilkan akurasi 0.85, yang berarti 85 dari 100 data diuji berhasil diprediksi dengan benar. Nilai macro average dan weighted average masing-masing berada pada kisaran 0.85–0.86, menunjukkan bahwa performa model stabil dan konsisten pada kedua kelas meskipun distribusi datanya mungkin tidak seimbang. Dengan hasil tersebut, dapat disimpulkan bahwa model Naive Bayes memiliki performa yang kuat dan layak digunakan untuk klasifikasi dalam konteks dataset ini.

3.8 Pembahasan

Secara visual, heatmap Confusion Matriks menunjukkan bahwa area diagonal, yang mewakili prediksi yang benar, berwarna hijau tua. Ini berarti sebagian besar prediksi yang dibuat oleh model tersebut benar. Hal ini mendukung kesimpulan bahwa model KNN dengan $k=5$ mampu menghasilkan hasil klasifikasi yang akurat dan andal, sedangkan pada model Naïve bayes heatmap Confusion Matriks menunjukkan bahwa sebagian besar nilai berada pada diagna utama, yang berwarna biru tua. Ini berarti model tersebut Sebagian besar membuat prediksi yang benar namun, ada juga beberapa nilai diluar diagonal yang berwarna biru muda yang menunjukkan bahwa model Naïve bayes masih membuat beberapa kesalahan dalam prediksinya. Meskipun demikian, naive bayes tetap menunjukkan kinerja yang cukup baik dengan pelatihan yang cepat dan efisien.

Tabel 4. Hasil evaluasi kedua model

Model	Accuracy	Precision	Recall	F1-score
K-NN	0.97	1.00	0.95	0.97
Naïve Bayes	0.85	0.92	0.81	0.86

Berdasarkan Tabel 4, menunjukkan bahwa algoritma KNN berkinerja lebih baik daripada Naive Bayes dalam setiap ukuran evaluasi. Hal ini menunjukkan bahwa KNN lebih efektif dalam menangani pola kompleks dalam data penjualan furnitur, terutama karena data tersebut mencakup banyak fitur numerik yang terkait. Sementara itu, Naive Bayes tidak selalu menjadi pilihan terbaik karena menggunakan metode yang mengasumsikan setiap fitur terpisah. Dalam data penjualan, hal-hal seperti harga, jumlah penjualan, dan total pendapatan saling terhubung, sehingga asumsi ini kurang tepat.

4. KESIMPULAN

Berdasarkan hasil evaluasi, Penelitian ini menyimpulkan bahwa algoritma K-Nearest Neighbor (KNN) menunjukkan kinerja lebih baik dibandingkan Naive Bayes dalam mengklasifikasikan status penjualan produk furnitur berdasarkan data historis. KNN menghasilkan akurasi sebesar 97%, presisi 1.00, recall 0.95, dan F1-score 0.97, sedangkan Naive Bayes memperoleh akurasi 85%, presisi 0.92, recall 0.81, dan F1-score 0.86. Perbedaan performa ini menunjukkan bahwa KNN lebih efektif dalam menangkap pola kompleks antar fitur penjualan, sementara asumsi independensi pada Naive Bayes tidak sesuai dengan karakteristik data. Model yang dihasilkan dapat digunakan sebagai alat pendukung keputusan dalam pengelolaan stok dan strategi pemasaran berbasis data. Manfaat penggunaan KNN dalam studi ini berasal dari kemampuannya untuk mempelajari bagaimana berbagai fitur saling berhubungan langsung dari data, tanpa perlu mengasumsikan jenis model probabilitas tertentu. KNN



menggunakan metode yang mengamati seberapa dekat suatu titik data baru dengan titik data lain yang sudah memiliki kategori yang diketahui. Pendekatan ini membantu KNN menangani perubahan pola data dengan lebih baik, terutama ketika faktor input memiliki hubungan yang kompleks dan melengkung. Berbeda dengan KNN, metode Naïve Bayes mengasumsikan bahwa setiap fitur terpisah dari fitur lainnya. Saat mengamati data penjualan furnitur, asumsi ini tidak selalu benar karena beberapa fitur, seperti harga, jenis produk, dan volume penjualan, saling berkaitan erat. Oleh karena itu, Naïve Bayes tidak berkinerja sebaik KNN. Dari temuan ini, jelas bahwa algoritma KNN lebih cocok untuk mengklasifikasikan status penjualan produk furnitur di CV Surya Gemilang. Model ini efektif dalam mengidentifikasi pola yang rumit dan memberikan prediksi yang lebih andal untuk kategori “laris” dan “tidak laris”. Dari sudut pandang implementasi bisnis, hasil klasifikasi ini dapat mendukung pengambilan keputusan strategis bagi organisasi. Dengan menggunakan prediksi, bisnis dapat lebih efektif menentukan produk mana yang kemungkinan berkinerja baik di pasar, sehingga memungkinkan mereka untuk meningkatkan area seperti perencanaan produksi, pengendalian inventaris, dan upaya promosi. Pendekatan ini memastikan bahwa keputusan tidak didasarkan pada tebakan atau penilaian pribadi, melainkan pada wawasan berbasis data yang diperoleh melalui analisis pembelajaran mesin. Penggunaan algoritma KNN juga memungkinkan analisis prediktif yang berkelanjutan. Model ini dapat diperbarui secara berkala dengan data penjualan baru, memastikan prediksi tetap selaras dengan perubahan kondisi pasar. Seiring dengan semakin banyaknya data yang tersedia dan penyempurnaan parameter, KNN berpotensi menawarkan wawasan yang sangat akurat untuk keputusan bisnis di masa mendatang.

REFERENCES

- [1] F. Sholehah, A. D. Putri, R. Rahmadden, and L. Efrizoni, “Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom: Comparison of Naive Bayes and K-Nearest Neighbors Algorithms for Metabolic Syndrome Classification,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 507–514, 2024, doi: <https://doi.org/10.57152/malcom.v4i2.1249>.
- [2] H. A. Dwi Fasnua, H. Yuana, and M. T. Chulkamdi, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus,” *Antivirus J. Ilm. Tek. Inform.*, vol. 16, no. 2, pp. 133–142, 2022, doi: [10.35457/antivirus.v16i2.2445](https://doi.org/10.35457/antivirus.v16i2.2445).
- [3] A. Muflih and Z. Fatah, “PREDIKSI PRODUK PENJUALAN DI SUPERMARKET DENGAN METODE ALGORITMA K-NEAREST NEIGHBORS (KNN),” *J. Ilm. SAINS Teknol. DAN Inf.*, vol. 3, no. 1, pp. 75–82, 2024, doi: [DOI:10.59024/jiti.v3i1.1056](https://doi.org/10.59024/jiti.v3i1.1056).
- [4] S. Wirma, “Jurnal Informatika Ekonomi Bisnis Data Mining Dengan Metode Naïves Bayes Classifier dalam Memprediksi Tingkat,” *J. Inform. Ekon. Bisnis*, vol. 4, no. 3, pp. 5–9, 2022, doi: [10.37034/inf.v4i3](https://doi.org/10.37034/inf.v4i3).
- [5] I. Rahmawati and Mustafa, “IMPLEMENTASI ALGORITMA NAÏVE BAYES DALAM PENYEDERHANAAN TEKS UNTUK ANAK DISLEKZIA,” *J. Rekayasa Sist. Inf. dan Teknol.*, vol. 3, no. 1, pp. 86–102, 2025, doi: <https://doi.org/10.70248/jrsit.v3i1.2925>.
- [6] S. R. Pratama and I. N. Fajri, “Comparison of K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) Algorithms in Predicting Customer Satisfaction,” *J. Comput. Sci. Informatics Eng.*, vol. 4, no. 3, pp. 135–146, 2025, doi: <https://doi.org/10.55537/cosie.v4i3.1160>.
- [7] Avinash, A. Widjaja, and O. Karnalim, “Comparative Analysis Of Machine Learning Algorithms For Forecasting Of Basic Goods Products,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 361–378, 2024, doi: <https://doi.org/10.28932/jutisi.v10i2.9357>.
- [8] Dinni Aulia, “Transformasi Komunikasi Pemasaran di Era Artificial Intelligence,” *J. Lensa Mutiara Komun.*, vol. 8, no. 2, pp. 1–16, 2024, doi: <https://dx.doi.org/10.51544/jlmlk.v8i2.5120>.
- [9] V. Listy and I. Ilham, “Revolusi Sistem Informasi Manajemen di Era AI dan Big Data Mengubah Cara Bisnis Bekerja,” *Simpatik J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 27–36, 2025, doi: [10.31294/simpatik.v5i1.7621](https://doi.org/10.31294/simpatik.v5i1.7621).
- [10] A. M. Husein, F. R. Lubis, and M. K. Harahap, “Analisis Prediktif untuk Keputusan Bisnis : Peramalan Penjualan,” *Data Sci. Indones.*, vol. 1, no. 1, pp. 32–40, 2021, doi: [10.47709/dsi.v1i1.1196](https://doi.org/10.47709/dsi.v1i1.1196).
- [11] H. R. Ami, K. Andrean, T. S. Gultom, and A. Hasibuan, “Peran Big Data dan AI (Artificial Intelligence) dalam Mendukung Kewirausahaan Technopreneurship,” *Fact. J. Ind. Manaj. dan Rekayasa Sist. Ind.*, vol. 3, no. 2, pp. 41–46, 2025, doi: [10.56211/factory.v3i2.716](https://doi.org/10.56211/factory.v3i2.716).
- [12] N. M. Putry, “KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELLITUS,” *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, pp. 45–57, 2022, doi: <https://doi.org/10.31294/evolusi.v10i1.12514>.
- [13] A. O. Permana and S. Saepudin, “Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes pada Aplikasi Shopee,” *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 1, pp. 25–32, 2023, doi: <https://doi.org/10.37859/cosictech.v4i1.4474>.
- [14] D. Susanto, B. Erlangga, A. Peryanto, and Y. F. Widodo, “Komparasi Efektifitas Analisis Sentimen pada Ulasan Produk E-Commerce Menggunakan Naive Bayes dan K-Nearest Neighbor,” *J. RESTIKOM Ris. Tek. Inform. dan Komput.*, vol. 7, no. 2, pp. 226–236, 2025, doi: [10.52005/restikom.v7i2.463](https://doi.org/10.52005/restikom.v7i2.463).
- [15] M. K. Al Fatah and A. Wibowo, “Perbandingan Metode Algoritma C4.5 Dan Naive Bayes Untuk Memprediksi Penjualan Kosmetik Pada Toko Jelita,” *J. Mnemon.*, vol. 7, no. 2, pp. 220–225, 2024, doi: [10.36040/mnemonic.v7i2.10730](https://doi.org/10.36040/mnemonic.v7i2.10730).
- [16] Y. Yulhendri, M. Malabay, and K. Kartini, “Correlated Naïve Bayes Algorithm to Determine Healing Rate of Hepatitis Patients,” *Int. J. Sci. Technol. Manag.*, vol. 4, no. 2, pp. 401–410, 2023, doi: [10.46729/ijstm.v4i2.776](https://doi.org/10.46729/ijstm.v4i2.776).
- [17] I. Arfanda, W. Ramdhan, and R. A. Yusda, “Naive Bayes Dalam Menentukan Penerimaan Bantuan Langsung Tunai,” *Digit. Transform. Technol.*, vol. 1, no. 1, pp. 9–16, 2021, doi: <https://doi.org/10.47709/digitech.v1i1.1091>.
- [18] Tresi Aprilia, “Klasifikasi Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor dan Metode Naive Bayes,”



- SATESI J. Sains Teknol. dan Sist. Inf.*, vol. 4, no. 2, pp. 156–163, 2024, doi: 10.54259/satesi.v4i2.3167.
- [19] S. Sathyanarayanan and B. R. Tantri, “Confusion Matrix-Based Performance Evaluation Metrics,” *African J. Biomed. Res.*, vol. 27, no. 4, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [20] B. A. Hamed, O. A. S. Ibrahim, and T. Abd El-Hafeez, “Optimizing classification efficiency with machine learning techniques for pattern matching,” *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00804-6.
- [21] V. Sheth, U. Tripathi, and A. Sharma, “A Comparative Analysis of Machine Learning Algorithms for Classification Purpose,” *Procedia Comput. Sci.*, vol. 2015, pp. 422–431, 2022, doi: <https://doi.org/10.1016/j.procs.2022.12.044>.
- [22] S. A. Ajagbe, J. B. Awotunde, and H. Florez, “Intrusion Detection: A Comparison Study of Machine Learning Models Using Unbalanced Dataset,” *SN Comput. Sci.*, vol. 5, no. 8, 2024, doi: 10.1007/s42979-024-03369-0.