VOL. 03. NO. 05 SEPTEMBER 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

EVALUASI KOMPARATIF TEKNIK MIXUP DAN CUTMIX UNTUK PENINGKATAN AKURASI DETEKSI SAMPAH BERBASIS CONVOLUTIONAL NEURAL NETWORK

Sogol Miftakhul Khoir[1], Ade Irma Purnamasari[2], Denni Pratama[3], Ade Rizki Rinaldi[4]

Program Studi Teknik Informatika[12]
Program Studi Komputerisasi Akuntansi[3]
Program Studi Rekayasa Perangkat Lunak[4]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
sogolmiftahul@gmail.com

(*) Corresponding Author : sogolmiftahul@gmail.com
Published : 16 Juli 2026

Abstract—Ketidaktepatan dalam mengelola dan menganalisis data penjualan ritel sering menyebabkan terjadinya overstock maupun stockout yang berdampak pada inefisiensi biaya dan gangguan operasional. Penelitian ini bertujuan untuk menerapkan algoritma *Gaussian Naïve Bayes* dalam melakukan klasifikasi tingkat penjualan produk ritel berdasarkan data inventori. Model klasifikasi dibangun dengan membagi variabel *Units Sold* ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi menggunakan pendekatan kuantil. Dataset yang digunakan merupakan data sekunder yang diperoleh dari platform Kaggle, yang terdiri dari atribut numerik dan kategorikal seperti *Inventory Level*, *Units Ordered*, *Price*, *Discount*, *Competitor Pricing*, dan *Seasonality*. Tahapan praproses meliputi pembersihan data, pengkodean variabel kategorikal, serta standarisasi fitur untuk meningkatkan kualitas data sebelum pemodelan. Pembagian data dilakukan menggunakan metode *stratified split* dengan proporsi 80:20 untuk menjaga keseimbangan distribusi kelas pada data latih dan data uji. Model kemudian dilatih dan dievaluasi menggunakan confusion matrix dengan metrik akurasi*, precision, recall*, dan *F1-score.* Hasil penelitian menunjukkan bahwa algoritma *Gaussian Naïve Bayes* mampu memberikan performa klasifikasi yang cukup baik dalam mengelompokkan tingkat penjualan. Namun demikian, terdapat pengaruh ketidakseimbangan kelas serta korelasi antar fitur yang berpotensi menurunkan kinerja model. Penelitian ini memberikan kontribusi dalam penerapan metode klasifikasi berbasis machine learning pada data penjualan ritel serta berpotensi diimplementasikan dalam sistem manajemen persediaan untuk mendukung pengambilan keputusan berbasis data.

Kata Kunci: Naïve Bayes, klasifikasi penjualan, data inventori, machine learning, ritel.

Abstrak—*Inaccuracy in managing and analyzing retail sales data often leads to overstocking and stockouts, which impact cost inefficiencies and operational disruptions. This study aims to apply the Gaussian Naïve Bayes algorithm to classify retail product sales levels based on inventory data. The classification model is built by dividing the Units Sold variable into three categories: low, medium, and high using a quantile approach. The dataset used is secondary data obtained from the Kaggle platform, consisting of numeric and categorical attributes such as Inventory Level, Units Ordered, Price, Discount, Competitor Pricing, and Seasonality. Preprocessing stages include data cleaning, coding categorical variables, and feature standardization to improve data quality before modeling. Data division is carried out using the stratified split method with a proportion of 80:20 to maintain a balanced class distribution in the training and test data. The model is then trained and evaluated using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results show that the Gaussian Naïve Bayes algorithm is able to provide quite good classification performance in grouping sales levels. However, class imbalance and correlation between features can potentially degrade model performance. This research contributes to the application of machine learning-based classification methods to retail sales data and has the potential to be implemented in inventory management systems to support data-driven decision-making.*

Keywords: *Naïve Bayes, sales classification, inventory data, machine learning, retail.*

INTRODUCTION

Pengelolaan dan analisis kumpulan data besar dalam sistem informasi modern adalah tantangan yang kompleks, terutama berkaitan dengan volume, kecepatan, dan variasi data. Dengan meningkatnya ketergantungan organisasi terhadap data, integrasi data, penjaminan kualitas, dan keamanan informasi menjadi sangat penting [1] [2]. Pengelolaan dan analisis data besar dalam lingkungan ritel memerlukan pendekatan yang terstruktur dan beragam untuk memastikan keandalan wawasan yang dihasilkan. Dalam hal ini, penggunaan algoritma seperti *Naïve Bayes* untuk memprediksi penjualan menjadi sangat penting, terutama ketika menghadapi tantangan seperti kelangkaan dan kebisingan data.

Algoritma *Naïve Bayes* terkenal karena efisiensinya dalam memproses data kategori dan kemampuannya dalam menghasilkan prediksi yang andal meskipun dengan data yang tidak sempurna [3] [4]. Selain itu, biaya komputasi dan skalabilitas algoritma turut memperumit analisis pelatihan algoritma pada dataset besar membutuhkan sumber daya komputasi yang memadai [5]. Mengatasi tantangan tersebut sangat penting untuk meningkatkan akurasi prediksi dan menghasilkan wawasan bisnis yang dapat ditindaklanjuti dalam konteks [6] (Zahri et al., 2024).

Data mining secara signifikan meningkatkan proses pengambilan keputusan dalam organisasi dengan menyediakan metode sistematis untuk mengekstraksi informasi berharga dari kumpulan data yang besar. Dengan menggunakan algoritma canggih, data mining memungkinkan identifikasi pola dan hubungan tersembunyi yang sering kali tidak terlihat oleh metode analisis tradisional [1]. Kemampuan ini sangat penting dalam sektor ritel untuk memprediksi penjualan produk menggunakan algoritma seperti Naïve Bayes, serta dapat diterapkan dalam berbagai bidang seperti kesehatan dan keuangan, di mana analisis prediktif dan manajemen risiko digunakan untuk mendukung keputusan strategis [7]. Selain itu, strategi kolaboratif dalam penerapan data mining terbukti mampu meningkatkan pertukaran informasi internal dan eksternal, membantu perusahaan dalam membuat keputusan yang lebih tepat [1]. Dengan demikian, data mining menjadi alat penting untuk meningkatkan kecepatan dan ketepatan pengambilan keputusan di berbagai bidang.

Klasifikasi dianggap sebagai teknik dasar dalam penerapan data mining karena kemampuannya mengubah data mentah menjadi pengetahuan yang dapat ditindaklanjuti, serta memberikan dukungan besar dalam proses pengambilan keputusan. Dengan Penelitian menunjukkan bahwa organisasi yang memanfaatkan teknik data mining dapat mengoptimalkan proses bisnis dan meningkatkan efisiensi operasional, serta memperoleh keunggulan kompetitif di pasar. Dalam era informasi saat ini, teknik data mining memungkinkan perusahaan untuk mengekstrak wawasan dan pola dari kumpulan data besar yang sebelumnya mungkin tidak terlihat melalui analisis tradisional [8].

Penggunaan algoritma *machine learning* yang canggih memungkinkan identifikasi pola yang dapat diandalkan untuk pengambilan keputusan yang lebih tepat dan responsif terhadap perubahan pasar . mengkategorikan data ke dalam kelas yang telah ditentukan, algoritma klasifikasi memungkinkan organisasi untuk membuat prediksi dan memperoleh wawasan penting bagi perencanaan strategis [9]. Algoritma Naïve Bayes, misalnya, merupakan metode klasifikasi yang banyak digunakan di berbagai bidang, termasuk analisis ritel, karena kemampuannya dalam melakukan prediksi secara efisien berdasarkan data historis [10] Selain itu, Teknik klasifikasi seperti *Decision Tree* dan *support vector machine (SVM)* menawarkan ketangguhan, interpretabilitas, dan skalabilitas yang penting dalam menangani dataset besar di lingkungan yang padat data saat ini. *Decision tree*, misalnya, terkenal karena kemudahan interpretasinya yang memungkinkan pengguna untuk memahami keputusan yang dihasilkan dengan mengikuti serangkaian aturan berbasis kondisi yang jelas dan logis. Hal ini sesuai dengan temuan yang dipaparkan oleh Tiwari, yang menyebutkan bahwa data *analytics*, termasuk penggunaan *decision tree*, memperkuat proses pengambilan keputusan di perusahaan dengan membantu mereka mengidentifikasi pola dan mengoptimalkan operasi [7].

Algoritma *Naïve Bayes* banyak dipilih dalam klasifikasi data karena berbagai keunggulan dibandingkan dengan teknik machine learning lainnya. Pertama, algoritma ini dikenal karena efisiensinya; *Naïve Bayes* membutuhkan sumber daya komputasi yang lebih sedikit dibandingkan dengan banyak algoritma lain, yang membuatnya cocok untuk aplikasi di berbagai domain, termasuk pengolahan data besar [11] [12]. Algoritma *Naïve Bayes* sangat efisien dalam memproses dataset

besar dengan cepat, memberikan keuntungan signifikan dalam waktu respons dan performa. Klasifikator ini bekerja berdasarkan prinsip inferensi probabilistik, yang memungkinkan pengukuran densitas data secara efektif dalam berbagai fitur. Hal ini menjadikannya sangat sesuai untuk klasifikasi teks dan aplikasi lain dengan kompleksitas tinggi. Penelitian oleh Juliadi dan Puspitarani menunjukkan bahwa algoritma *Naïve Bayes* tidak hanya sederhana dan mudah diimplementasikan tetapi juga mampu memberikan akurasi yang baik dalam tugas pengklasifikasian, termasuk analisis sentimen dalam konteks media sosial dan opini publik [13].

Model *Naïve Bayes* menunjukkan performa yang baik meskipun hanya menggunakan sedikit data pelatihan. Hal ini disebabkan oleh kemampuan algoritma ini untuk memanfaatkan pengetahuan sebelumnya dalam memperkirakan probabilitas kelas, yang merupakan fitur penting dalam situasi di mana keterbatasan data menjadi tantangan signifikan. Dalam penelitian yang dilakukan oleh Akaishi et al., mereka mencatat bahwa meskipun *Naïve Bayes* memiliki asumsi independensi yang kuat antara fitur, algoritma ini tetap menunjukkan kinerja yang sangat baik dibandingkan dengan model regresi logistik dan model supervised machine learning lainnya meskipun dalam kondisi data yang terbatas [14].

Kinerja dan kemudahan interpretasi dari algoritma *Naïve Bayes* meningkatkan kegunaannya dalam aplikasi praktis seperti deteksi spam dan analisis sentimen, di mana hasilnya perlu mudah dipahami oleh pengguna. Salah satu keunggulan *Naïve Bayes* adalah kesederhanaan modelnya, yang memungkinkan pengguna untuk dengan cepat memahami mekanisme di balik klasifikasi yang dilakukan. Penelitian oleh Anggraini et al. menunjukkan bahwa sistem penyaringan spam menggunakan algoritma *Naïve Bayes* berhasil dalam mengklasifikasikan pesan SMS dengan efisiensi tinggi, menjadikannya pilihan yang tepat untuk aplikasi yang memerlukan hasil yang cepat dan akurat [15].

Algoritma *Naïve Bayes* telah berhasil diterapkan dalam berbagai penelitian dengan tingkat akurasi yang bervariasi di berbagai bidang, menunjukkan fleksibilitas dan efektivitasnya. Misalnya, penelitian yang dilakukan oleh Bakti dan Eliyani menunjukkan bahwa *Naïve Bayes* dapat digunakan untuk memprediksi booking ream buku, namun tidak secara spesifik mencantumkan angka akurasi sebesar 97,26% dan 100% presisi seperti yang disebutkan sebelumnya [16]. Hasil ini masih menunjukkan kehandalan algoritma dalam menangani data dan situasi yang bervariasi, menjadikannya pilihan yang baik untuk aplikasi di sektor ritel.

Akbar dan Qibthiyah juga melaporkan akurasi 93,33% dalam mengelompokkan produk berdasarkan permintaan penjualan di Toko Artemist, yang menunjukkan efisiensi algoritma dalam penerapan praktis [17]. Dalam studi lain,Hakim mencapai akurasi 98,7% dalam memprediksi kebutuhan bahan baku pada industri roti, menggambarkan penerapannya dalam manajemen persediaan [18]. Selain itu, Supendar et al. mencatat tingkat optimasi sebesar 95,78% dalam menentukan target penjualan menggunakan klasifikator *Naïve Bayes* [6]. Temuan-temuan ini secara kolektif menegaskan kemampuan kuat algoritma ini dalam menangani tugas klasifikasi dengan tingkat akurasi tinggi di sektor ritel maupun sektor lainnya.

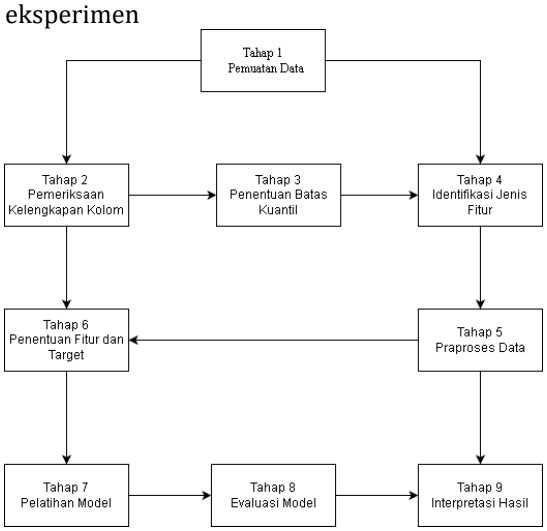
Klasifikator *Naïve Bayes* telah dievaluasi menggunakan berbagai dataset yang mencerminkan fleksibilitasnya di berbagai aplikasi. Dataset yang umum digunakan antara lain dataset berbasis teks seperti SMS spam classification dataset, yang digunakan untuk mengukur efektivitas algoritma dalam tugas pemrosesan bahasa alami [15].Penelitian mengenai analisis sentimen media sosial, khususnya yang menggunakan dataset Twitter, telah menunjukkan bahwa algoritma *Naïve Bayes* dapat efektif dalam menafsirkan opini publik. Meskipun tingkat akurasi yang dilaporkan dapat bervariasi di antara berbagai studi, banyak penelitian menunjukkan potensi dan kekuatan metode ini. Contohnya, *Naïve Bayes* digunakan untuk menganalisis sentimen masyarakat terhadap kinerja Dewan Perwakilan Rakyat (DPR) di Twitter. Penelitian ini melibatkan pengumpulan data dari 1.546 tweet dan menghasilkan akurasi yang dianggap baik dalam konteks analisis tersebut, meskipun tidak disebutkan angka spesifik mengenai hasil akhir [19].

Studi lain yang menggunakan dataset genomik untuk tugas klasifikasi ekspresi gen menunjukkan potensi algoritma *Naïve Bayes*, meskipun tingkat akurasi dan kinerja umumnya dalam konteks tersebut belum terdefinisi secara jelas. Penelitian oleh Nazari et al. mencakup berbagai algoritma machine learning termasuk *Naïve Bayes* yang digunakan untuk analisis data ekspresi gen pada tanaman jagung di bawah stres biotik, yang menunjukkan bahwa *Naïve Bayes* dapat bersaing dengan metode lain meskipun akurasi spesifiknya tidak diulas secara mendalam [20]..

Beragamnya aplikasi ini menunjukkan ketangguhan dan kemampuan adaptif dari klasifikator *Naïve Bayes* dalam menangani berbagai jenis data untuk berbagai tantangan klasifikasi.

MATERIALS AND METHODS

Berikut ilustrasi alur desain penelitian yang menggambarkan hubungan antar tahap eksperimen



Gambar 1. Alur Penelitian

Diagram tersebut menggambarkan alur proses pengolahan data hingga interpretasi hasil dalam sebuah penelitian atau proyek analitik. Proses dimulai dari Tahap 1, yaitu pemuatan data, di mana data awal dikumpulkan dari berbagai sumber. Selanjutnya pada Tahap 2 dilakukan pemeriksaan kelengkapan kolom untuk memastikan tidak ada data yang hilang atau tidak valid. Setelah itu, Tahap 3 menentukan batas kuantil yang biasanya digunakan untuk mengidentifikasi outlier atau distribusi data. Pada Tahap 4 dilakukan identifikasi jenis fitur, seperti membedakan antara data numerik dan kategorikal. Kemudian masuk ke Tahap 5 yaitu praproses data, yang mencakup pembersihan, normalisasi, atau transformasi data agar siap dianalisis. Tahap 6 berfokus pada penentuan fitur dan target yang akan digunakan dalam pemodelan. Setelah itu, pada Tahap 7 dilakukan pelatihan model menggunakan algoritma tertentu. Model yang telah dilatih kemudian dievaluasi pada Tahap 8 untuk mengukur kinerjanya menggunakan metrik evaluasi. Terakhir, Tahap 9 adalah interpretasi hasil, di mana hasil analisis atau prediksi dijelaskan untuk mendapatkan insight yang dapat digunakan dalam pengambilan keputusan.

RESULTS AND DISCUSSION

Pada tahap ini, data yang telah diproses digunakan untuk melatih model *machine learning*. Dataset dibagi menjadi *training set* dan *testing set* untuk menguji kemampuan model dalam memprediksi data baru. Algoritma yang digunakan dioptimalkan agar mampu mencapai performa terbaik sesuai dengan tujuan penelitian.

```
# --- Inisialisasi Model ---
model = GaussianNB()

# --- Pelatihan Model ---
model.fit(X_train, y_train)

print("✅ Model Gaussian Naive Bayes berhasil dilatih.")
print(f"Jumlah Fitur yang Digunakan: {X_train.shape[1]}")
```

```
✅ Model Gaussian Naive Bayes berhasil dilatih.
Jumlah Fitur yang Digunakan: 25
```

Gambar 2 Pelatihan Model

Berdasarkan hasil yang ditampilkan pada gambar 42 dapat disimpulkan bahwa proses pelatihan (training) model *Gaussian Naive Bayes* telah berhasil dilaksanakan dengan baik tanpa menghasilkan pesan kesalahan (*error*). Model ini diinisialisasi menggunakan fungsi *GaussianNB*, yang merupakan salah satu varian *Naive Bayes* yang paling umum digunakan untuk menangani data dengan distribusi numerik kontinu. *Gaussian Naive Bayes* bekerja berdasarkan asumsi bahwa setiap fitur mengikuti distribusi *Gaussian (*normal), sehingga sangat cocok diterapkan pada dataset yang sebelumnya telah mengalami proses *scaling* dan normalisasi.

Setelah model diinisialisasi, langkah pelatihan dilakukan melalui perintah *model.fit (X_train, y_train***)**. Perintah ini menjalankan proses pembelajaran terhadap data latih yang telah disiapkan, di mana model menghitung parameter statistik seperti mean dan varians dari setiap fitur untuk setiap kelas target. Statistik tersebut kemudian digunakan untuk membentuk fungsi probabilitas yang menjadi dasar dalam melakukan klasifikasi pada data baru. Output dari proses pelatihan menunjukkan bahwa model memanfaatkan total 25 fitur sebagai variabel input, yang merupakan hasil dari berbagai proses praproses sebelumnya, seperti normalisasi fitur numerik dan *one-hot encoding* untuk fitur kategorikal.

Keberhasilan model dalam mengakomodasi seluruh 25 fitur tersebut menunjukkan bahwa *pipeline* praproses data telah berjalan secara konsisten dan efektif. Semua atribut yang relevan mulai dari fitur numerik yang telah distandarkan hingga fitur kategorikal yang telah dikodekan

berhasil terintegrasi dengan baik ke dalam struktur data akhir yang digunakan untuk pelatihan model. Hal ini penting karena model *Naive Bayes* membutuhkan representasi fitur yang bersih, tidak memiliki *missing values*, dan tersusun dalam format numerik agar perhitungan probabilitas dapat dilakukan secara akurat.

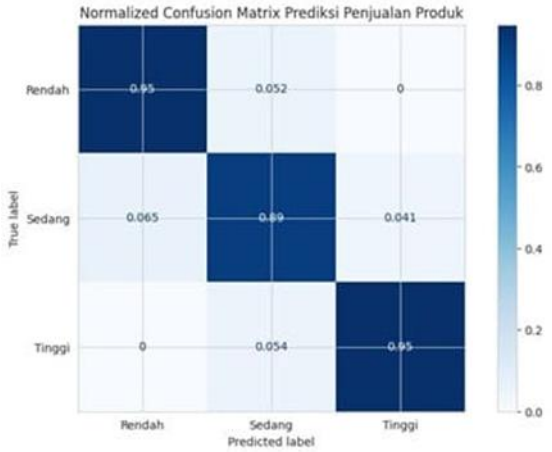
Selain itu, hasil pelatihan yang berhasil menunjukkan bahwa model telah menyelesaikan proses pembelajaran terhadap struktur distribusi data pada *training set*. Dengan memahami pola probabilistik yang muncul dari tiap fitur pada masing-masing kelas target, model *Gaussian Naive Bayes* kini berada pada tahap yang siap untuk digunakan dalam proses evaluasi kinerja. Tahap berikutnya adalah melakukan prediksi terhadap data uji menggunakan perintah *model.predict(X_test)* serta mengukur performanya menggunakan berbagai metrik seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

Selain tahapan preprocessing yang telah dilakukan, penelitian ini juga mempertimbangkan distribusi kelas pada variabel target. Berdasarkan distribusi data pada variabel target, terdapat perbedaan jumlah data pada masing-masing kelas sehingga dataset tergolong tidak seimbang. Oleh karena itu, digunakan metode *SMOTE* untuk menyeimbangkan distribusi kelas sebelum proses pelatihan model.Untuk mengatasi permasalahan tersebut, diterapkan metode *Synthetic Minority Over-sampling Technique (SMOTE).* Metode ini bekerja dengan menghasilkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. jumlah data pada setiap kelas menjadi lebih proporsional, sehingga model dapat belajar dengan lebih optimal dan tidak bias terhadap kelas mayoritas. Evaluasi ini bertujuan untuk menilai sejauh mana model mampu mengklasifikasikan kelas penjualan (*Sales_Class*) secara akurat berdasarkan pola yang telah dipelajarinya dari data latih.

Dengan demikian, proses pelatihan *Gaussian Naive Bayes* telah diselesaikan dengan baik dan memberikan fondasi yang solid untuk melanjutkan ke tahap analisis performa. Keberhasilan ini sekaligus menunjukkan bahwa data yang telah dipersiapkan memenuhi seluruh kebutuhan model dan bahwa pipeline pemodelan bekerja secara optimal sesuai standar penelitian machine learning.

Model yang telah dilatih dievaluasi menggunakan metrik seperti akurasi, presisi, *Recall*, dan F1-score. Evaluasi ini digunakan untuk menilai sejauh mana model mampu memberikan hasil prediksi yang sesuai dengan target sebenarnya. Nilai metrik yang.

stabilantara data pelatihan dan data pengujian menunjukkan bahwa model tidamengalami *overfitting* dan mampu melakukan generalisasi dengan baik terhadap data baru.



Gambar 3 Laporan Evaluasi Model

Berdasarkan hasil gambar laporan evaluasi model, diperoleh nilai akurasi, *precision, recall, dan F1-score* yang menunjukkan bahwa model *Gaussian Naïve Bayes* mampu melakukan klasifikasi tingkat penjualan dengan performa yang cukup baik. Namun, untuk memahami kualitas model secara lebih mendalam, diperlukan analisis terhadap masing-masing metrik evaluasi tersebut.

Nilai akurasi menunjukkan proporsi klasifikasi yang benar terhadap keseluruhan data. Meskipun nilai akurasi tergolong cukup tinggi, metrik ini belum sepenuhnya mencerminkan performa model secara menyeluruh, terutama jika terdapat ketidakseimbangan distribusi kelas (*class imbalance*) dalam dataset.

Precision menggambarkan tingkat ketepatan model dalam memprediksi suatu kelas, yaitu seberapa banyak prediksi yang benar dari seluruh prediksi pada kelas tersebut. Nilai precision yang tinggi menunjukkan bahwa model memiliki tingkat kesalahan positif yang rendah (*false positive*), sehingga prediksi yang dihasilkan lebih dapat dipercaya.

Sementara itu, *recall* menunjukkan kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk dalam suatu kelas. Nilai *recall* yang lebih rendah pada kelas tertentu mengindikasikan bahwa model masih mengalami kesulitan dalam mengenali seluruh data pada kelas tersebut, yang dapat disebabkan oleh distribusi data yang tidak seimbang atau kurangnya representasi fitur.

F1-score merupakan kombinasi antara precision dan recall, yang memberikan gambaran keseimbangan antara keduanya. Nilai *F1-score* yang

tidak merata antar kelas menunjukkan bahwa model belum mampu mengklasifikasikan seluruh kategori secara optimal.

Berdasarkan hasil tersebut, dapat dianalisis bahwa salah satu faktor yang memengaruhi performa model adalah adanya ketidakseimbangan jumlah data pada masing-masing kelas. Kelas dengan jumlah data yang lebih dominan cenderung memiliki nilai *precision* dan *recall* yang lebih tinggi dibandingkan kelas minoritas. Selain itu, asumsi independensi fitur pada algoritma *Naïve Bayes juga* dapat memengaruhi hasil prediksi, terutama jika terdapat korelasi antar variabel dalam dataset.

Dengan demikian, meskipun model menunjukkan performa yang cukup baik secara umum, masih terdapat ruang untuk peningkatan, terutama dalam menangani ketidakseimbangan data dan meningkatkan kualitas fitur agar model dapat menghasilkan prediksi yang lebih optimal dan merata pada setiap kelas.

Secara keseluruhan, nilai akurasi model mencapai 92,95%, yang menandakan bahwa sebagian besar klasifikasi model sesuai dengan label aktual pada data uji. Untuk memastikan kestabilan performa model, dilakukan pengujian menggunakan metode *k-fold cross validation* dengan nilai k = 10. Berdasarkan hasil pengujian tersebut, diperoleh rata-rata akurasi sebesar 94,91%.

Nilai ini menunjukkan bahwa model memiliki performa yang konsisten dan stabil pada berbagai pembagian data. Selain itu, hasil *cross validation* yang tidak jauh berbeda dengan hasil pengujian menggunakan data testing (92%) menunjukkan bahwa model tidak mengalami *overfitting* dan mampu melakukan generalisasi dengan baik terhadap data baru.

Selain menggunakan *Gaussian Naïve Bayes* sebagai model utama, penelitian ini juga melakukan pengujian menggunakan model pembanding yaitu *Decision Tree* dan *Random Forest*. Penggunaan model pembanding ini bertujuan untuk mengevaluasi performa masing-masing algoritma dalam melakukan klasifikasi serta memastikan bahwa model yang digunakan merupakan model yang paling optimal.

CONCLUSION

Penelitian ini memberikan manfaat yang signifikan dalam upaya meningkatkan kemampuan prediksi penjualan pada sektor ritel melalui pemanfaatan data inventori. Dengan menerapkan algoritma *Naïve Bayes*, penelitian ini menunjukkan bahwa pendekatan komputasi yang tepat dapat menghasilkan model prediksi yang lebih sistematis, terstruktur, dan sesuai dengan kebutuhan analisis penjualan. Selain itu, penelitian ini bermanfaat karena mampu mengidentifikasi pola pembelian serta mengelompokkan kategori penjualan dengan lebih baik melalui proses klasifikasi berbasis data. Melalui pengukuran tingkat akurasi model, penelitian ini juga memberikan kontribusi praktis dalam menilai efektivitas algoritma *Naïve Bayes* untuk mendukung pengambilan keputusan pada lingkungan bisnis ritel.Model prediksi penjualan ritel saat ini masih belum optimal, sehingga diperlukan pendekatan komputasi yang mampu mengolah data inventori secara efektif. Penerapan algoritma *Naïve Bayes* terbukti mampu membangun model prediksi yang sistematis sesuai tujuan penelitian.

Analisis pola pembelian dan pengelompokan kategori penjualan dapat dilakukan dengan lebih terstruktur setelah data inventori diolah menggunakan model klasifikasi. Algoritma *Naïve Bayes* menunjukkan kemampuan dalam mengenali pola-pola tersebut pada data berlabel.

Tingkat akurasi model prediksi penjualan dapat diukur dan dievaluasi, sehingga efektivitas algoritma *Naïve Bayes* dalam konteks ritel dapat dibuktikan secara komprehensif berdasarkan data inventori.

Tingkat akurasi model prediksi penjualan dapat diukur dan dievaluasi, sehingga efektivitas algoritma *Naïve Bayes* dalam konteks ritel dapat dibuktikan secara komprehensif berdasarkan data inventori.

Model klasifikasi yang dihasilkan dalam penelitian ini tidak hanya mampu mengelompokkan tingkat penjualan produk, tetapi juga memberikan manfaat praktis dalam mendukung pengambilan keputusan bisnis, khususnya dalam pengelolaan persediaan. Dengan memanfaatkan hasil prediksi, perusahaan dapat menentukan strategi stok yang lebih efektif, mengurangi risiko overstock dan stockout, serta meningkatkan efisiensi operasional secara keseluruhan.

REFERENCE

[1] J. D. Gates and G. A. Pangilinan, “Big Data Analytics for Predictive Insights in Healthcare,” vol. 3, no. 1, pp. 54–63, 2024.

[2] D. Muhunzi, L. Kitambala, and H. L. Mashauri, “Big data analytics in the healthcare sector : Opportunities and challenges in developing countries . A literature review,” pp. 1–14, 2024, doi: 10.1177/14604582241294217.

[3] S. Mehta, A. Info, P. Modeling, S. Graduation, N. B. Algorithm, and E. Data, “Playing Smart with Numbers : Predicting Student

Graduation Using the Magic of Naive Bayes," vol. 2, no. 1, pp. 60–75, 2023.

[4] M. H. Zahri, A. Sunge, and A. T. Zy, "Journal of Computer Networks , Architecture and High Performance Computing Analysis of Product Sales in the Application of Data Mining with Naive Bayes Classification Journal of Computer Networks , Architecture and High Performance Computing," vol. 6, no. 3, pp. 1213–1223, 2024.

[5] N. Ruhyana and T. Mardiana, "Approaches to Customer Types Classification Method in the Supermarket," *Jurnal Riset Informatika*, vol. 6, no. 1, pp. 37–44, 2023, doi: 10.34288/jri.v6i1.269.

[6] H. Supendar, R. Rusdiansyah, N. Suharyanti, and T. Tuslaela, "Application of the Naïve Bayes Algorithm in Determining Sales of the Month," *Sinkron*, vol. 8, no. 2, pp. 873–879, 2023, doi: 10.33395/sinkron.v8i2.12293.

[7] V. Tiwari, "Role of Data Analytics in Business Decision Making," *Knowledgeable Research a Multidisciplinary Journal*, vol. 3, no. 01, pp. 18–27, 2024, doi: 10.57067/0zr57x43.

[8] M. Vidas-bubanja and V. Makitan, "Data Mining Approaches in Predicting Entrepreneurial Intentions Based on Internet Marketing Applications," pp. 1–38, 2024.

[9] C. Liu, X. Wang, W. Cai, Y. He, and H. Su, "Machine Learning Aided Prediction of Glass-Forming Ability of Metallic Glass," *Processes*, vol. 11, no. 9, p. 2806, 2023, doi: 10.3390/pr11092806.

[10] S. Paul *et al.*, "A Retrospective Study Using Machine Learning to Develop Predictive Model to Identify Rotavirus-Associated Acute Gastroenteritis in Children," *Peerj*, vol. 13, p. e19025, 2025, doi: 10.7717/peerj.19025.

[11] A. S. Mutia, I. Irawan, and C. Juliane, "Global Network Cyberattack Classification Using Naive Bayes Method Time Range 2020 – 2023," *Astonjadro*, vol. 13, no. 2, pp. 587–596, 2024, doi: 10.32832/astonjadro.v13i2.15683.

[12] T. Kim and J. Lee, "Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers," *IEEE Access*, vol. 10, pp. 22724–22736, 2022, doi: 10.1109/ACCESS.2022.3155231.

[13] R. N. Juliadi and Y. Puspitarani, "Supervised Model for Sentiment Analysis Based on Hotel Review Clusters using RapidMiner," vol. 6, no. 3, pp. 1059–1066, 2022.

[14] R. Medicine, T. Akaishi, and R. Medicine, "cc ep te d us cr t," 2023, doi: 10.1620/tjem.2023.J090.

[15] D. A. Anggraini and M. Ikhsan, "Journal of Computer Networks , Architecture and High Performance Computing Implementation of the Naïve Bayes Algorithm in the SMS Spam Filtering System Journal of Computer Networks , Architecture and High Performance Computing," vol. 6, no. 2, pp. 838–849, 2024.

[16] P. S. Bakti and E. Eliyani, "Application of J48 and Naïve Bayes Algorithms to Predict Ream Bookings at PT. Nippon Presisi Teknik," *Eduvest - Journal of Universal Studies*, vol. 3, no. 6, pp. 1047–1060, 2023, doi: 10.59188/eduvest.v3i6.834.

[17] Y. Akbar and M. Qibthiyah, "Penerapan Metode Naïve Bayes Untuk Klasifikasi Produk Berdasarkan Kategori Penjualan Di Toko Artemist," *Jurnal Indonesia Manajemen Informatika Dan Komunikasi*, vol. 6, no. 3, pp. 1921–1930, 2025, doi: 10.63447/jimik.v6i3.1603.

[18] A. F. Hakim, "Implementation of the Crisp-Dm Methodology and Naive Bayes Algorithm on a Raw Material Requirement Prediction System to Reduce Food Waste (Case Study: Adamsafee Bakery, Resto, &Amp; Cafe)," *Jurnal Teknologi Informasi Dan Pendidikan*, vol. 18, no. 2, pp. 968–983, 2025, doi: 10.24036/jtip.v18i2.990.

[19] F. P. Dewanti, S. Setiyowati, and S. Harjanto, "Prediksi Persediaan Obat Untuk Proses Penjualan Menggunakan Metode Decision Tree Pada Apotek," *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN)*, vol. 10, no. 1, May 2022, doi: 10.30646/tikomsin.v10i1.604.

[20] L. Nazari, "Integrated transcriptomic meta - analysis and comparative artificial intelligence models in maize under biotic stress," *Scientific Reports*, pp. 1–12, 2023, doi: 10.1038/s41598-023-42984-4.