

Evaluasi Komparatif Algoritma Naïve Bayes, KNN, Logistic Regression, SVM, dan Extra Trees untuk Analisis Sentimen Tokopedia

Indramawan Ciputra*, Amiq Fahmi

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ^{1,*}111202214563@mhs.dinus.ac.id, ²amiq.fahmi@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202214563@mhs.dinus.ac.id

Submitted: 15/10/2025; Accepted: 06/12/2025; Published: 08/12/2025

Abstrak—Pesatnya perkembangan teknologi digital telah mendorong pergeseran pola konsumsi masyarakat, khususnya dalam aktivitas belanja daring melalui platform e-commerce seperti Tokopedia. Ulasan pengguna menghasilkan data berukuran besar yang dapat dianalisis untuk mengidentifikasi sentimen konsumen secara faktual dan sistematis. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa lima algoritma klasifikasi sentimen, yaitu *Naïve Bayes*, *K-Nearest Neighbors* (KNN), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extra Trees Classifier*, berdasarkan data ulasan pengguna Tokopedia. Proses analisis diawali dengan *web crawling*, dilanjutkan dengan pra-pemrosesan teks yang mencakup tokenisasi, normalisasi huruf (*case folding*), dan penghapusan kata-kata umum (*stop-word removal*), serta analisis sentimen menggunakan kelima algoritma klasifikasi tersebut. Evaluasi performa dilakukan dengan menggunakan metrik *accuracy*, *precision*, *recall*, and *F1-score*. Hasil menunjukkan bahwa SVM mencapai akurasi tertinggi sebesar 85%, mengungguli KNN dan *Extra Trees Classifier* (84%), *Logistic Regression* (82%), serta *Naive Bayes* (79%). Keunggulan SVM dalam konteks ini berkaitan dengan kemampuannya menemukan hyperplane optimal yang efektif dalam memisahkan kelas sentimen, terutama pada data berdimensi tinggi. Temuan ini memberikan rekomendasi praktis bagi pengembang sistem analisis sentimen dalam memilih algoritma yang paling optimal, sekaligus memperkuat penerapan teknik pengolahan bahasa alami (*Natural Language Processing/NLP*) dalam sektor *e-commerce* di Indonesia direkomendasikan.

Kata Kunci: Analisis Sentimen; Klasifikasi Teks; E-Commerce; Machine Learning; Extra Trees Classifier

Abstract—The rapid evolution of digital technology has catalyzed a shift in consumer behavior, particularly in online shopping activities facilitated by e-commerce platforms such as Tokopedia. User-generated reviews yield large-scale textual data that can be systematically analyzed to uncover consumer sentiment in a factual and structured manner. This study aims to evaluate and compare the performance of five sentiment classification algorithms Naive Bayes, K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and Extra Trees Classifier based on user review data from Tokopedia. The analytical workflow begins with web crawling, followed by text preprocessing procedures including tokenization, case folding, and stop-word removal, culminating in sentiment classification using the aforementioned algorithms. Performance evaluation was conducted using four standard metrics accuracy, precision, recall, and F1-score. The results reveal that SVM achieved the highest accuracy at 85%, outperforming KNN and Extra Trees Classifier (84%), Logistic Regression (82%), and Naive Bayes (79%). SVM's superior performance is attributed to its ability to identify optimal hyperplanes that effectively separate sentiment classes, particularly in high-dimensional feature spaces. These findings offer practical insights for developers of sentiment analysis systems in selecting the most effective algorithm, while reinforcing the strategic application of Natural Language Processing (NLP) techniques within Indonesia's e-commerce landscape.

Keywords: Sentiment Analysis, Text Classification, E-Commerce, Machine Learning, Extra Trees Classifier

1. PENDAHULUAN

Teknologi informasi dan komunikasi (TIK) telah berkembang dari sekadar alat bantu menjadi kebutuhan esensial dan penggerak utama dalam transformasi sosial-ekonomi masyarakat modern [1]. Perkembangan pesat TIK ini telah mereformulasi aktivitas sehari-hari menjadi lebih efisien, terhubung, dan responsif secara digital. Dampaknya terlihat jelas dalam sektor ekonomi, khususnya bagi usaha kecil dan menengah yang mulai memanfaatkan teknologi untuk memperluas jangkauan pasar secara adaptif. Salah satu manifestasi nyata dari perubahan ini adalah pertumbuhan e-commerce sebagai bentuk perdagangan daring berbasis internet, yang tidak hanya memudahkan transaksi tanpa kehadiran fisik, tetapi juga menawarkan pengalaman belanja yang lebih nyaman dan sesuai dengan keterlibatan lintas generasi (*multigenerational engagement*) [2].

Di Indonesia, sejumlah situs e-commerce seperti Shopee, Tokopedia, Bukalapak, Lazada, dan Blibli telah menjadi platform yang sangat populer di kalangan masyarakat [2]. Kelima layanan ini bersaing secara konstruktif dengan mengadopsi beragam model bisnis serta strategi promosi untuk menarik perhatian konsumen. Persaingan tersebut tidak hanya mencakup aspek harga dan variasi produk, tetapi juga meliputi kualitas pelayanan, kemudahan akses, serta kecepatan dalam proses transaksi. Selain itu, setiap platform berupaya membangun loyalitas pengguna melalui peningkatan pengalaman belanja yang nyaman dan aman. Dengan semakin banyaknya pilihan layanan belanja daring, masyarakat kini memiliki keleluasaan untuk menentukan platform yang dianggap paling sesuai dengan kebutuhan dan preferensi mereka. Persaingan yang ketat ini mendorong setiap penyedia layanan untuk terus berinovasi agar mampu mempertahankan kepercayaan konsumen. Hal tersebut juga menunjukkan bahwa perilaku dan kepuasan pengguna menjadi faktor penting dalam keberlanjutan bisnis e-commerce di Indonesia. Keberhasilan platform kini tidak hanya ditentukan oleh jumlah pengguna, tetapi juga oleh kemampuannya memahami kebutuhan pasar.

Percepatan transformasi digital telah mengubah pola konsumsi masyarakat Indonesia. Tokopedia, sebagai salah satu platform *e-commerce* terkemuka, tidak hanya digunakan untuk bertransaksi, tetapi juga menjadi wadah bagi

pengguna untuk memberikan ulasan terhadap produk yang telah dibeli. Ulasan-ulasan ini menghasilkan data tekstual yang sangat kaya dan dapat dimanfaatkan untuk memahami persepsi konsumen secara lebih sistematis [3]. Dengan memanfaatkan data tersebut, pelaku bisnis dapat meningkatkan kualitas produk dan layanan secara lebih terarah. Analisis sentimen terhadap ulasan pengguna Tokopedia terbukti mampu memberikan gambaran objektif mengenai tingkat kepuasan pelanggan. Penelitian oleh Ernawati [4] menggunakan algoritma *Naïve Bayes* menunjukkan bahwa sebagian besar ulasan bersifat positif, sementara ulasan negatif umumnya terkait dengan kualitas pengemasan dan produk. Temuan ini sangat relevan sebagai dasar evaluasi bagi pelaku usaha dalam meningkatkan mutu layanan.

Berbagai algoritma machine learning seperti *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Logistic Regression* (LR), *Random Forest*, dan *Extra Trees Classifier* (ETC) telah digunakan secara luas dalam klasifikasi teks dan analisis sentimen. Metode-metode ini memungkinkan proses analisis dilakukan secara cepat dan efisien, dengan tingkat akurasi yang tinggi. Sebelum klasifikasi dilakukan, tahapan pra-pemrosesan dan normalisasi teks menjadi krusial. Tahapan ini meliputi pembersihan karakter tidak relevan, penghapusan tanda baca, serta konversi kata ke bentuk yang konsisten. Normalisasi bertujuan menyamakan variasi kata, memperbaiki singkatan, kesalahan penulisan, dan penggunaan bahasa informal. Selanjutnya, data teks dikonversi ke bentuk numerik menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF), yang membantu mengidentifikasi tingkat kepentingan suatu kata dalam dokumen relatif terhadap korpus lainnya.

Sejumlah studi telah menunjukkan efektivitas berbagai algoritma dalam klasifikasi sentimen. Christian Cahyaningtyas et al. [5] menggunakan algoritma *Decision Tree* dengan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data pada ulasan pengguna Shopee, dan mencatat akurasi sebesar 99,91% dengan SMOTE, dibandingkan 99,89% tanpa SMOTE. Sementara itu, Wijianto et al. [6] membandingkan performa KNN dan *Naïve Bayes* dalam klasifikasi ulasan Tokopedia, dengan hasil bahwa KNN lebih unggul (80%) dibandingkan *Naïve Bayes* (71,5%). Muhamad Jodi Setiawan dan Vinna Rahmayanti Setyaning Nastiti [7] melakukan analisis sentimen pada ulasan aplikasi DANA dengan membandingkan tiga algoritma klasifikasi, yaitu *XGBoost*, *Support Vector Machine*, dan *Extra Trees Classifier*. Hasil evaluasi menunjukkan bahwa metode *XGBoost* dan SVM sama-sama memperoleh akurasi sebesar 93%, sementara metode ETC sedikit lebih rendah dengan akurasi 92%. Selain itu, pada uji K-Fold Cross Validation, algoritma ETC mencatat nilai F1-Score tertinggi, yaitu 96% pada K=6. Aditya Rahman dan Riza Adrianti [8] membandingkan performa algoritma *Naïve Bayes* dan SVM pada analisis sentimen ulasan aplikasi e-commerce. Hasil penelitian menunjukkan bahwa algoritma SVM unggul dengan akurasi sebesar 88%, sedangkan *Naïve Bayes* hanya mencapai 82%. Ernianti Hasibuana dan Elmo Allistair Heriyanto [9] melakukan analisis sentimen pada ulasan aplikasi Amazon Shopping di *Google Play Store* menggunakan berbagai varian algoritma *Naïve Bayes*. Hasil penelitian menunjukkan bahwa *Multinomial Naïve Bayes* memberikan performa terbaik dengan *accuracy* 86,74%, *precision* 78,82%, *recall* 85,90%, dan *F1-score* 82,21%. Penerapan metode TF-IDF mampu meningkatkan akurasi hingga 88,37%, menjadikannya lebih unggul dibanding algoritma *Naïve Bayes* lainnya.

Secara umum, algoritma seperti *Decision Tree* (dengan SMOTE), KNN, SVM, dan *Naïve Bayes* memiliki keunggulan masing-masing, yang sangat bergantung pada karakteristik data, ukuran dataset, dan teknik pra-pemrosesan yang digunakan. Teknik penyeimbangan data seperti SMOTE terbukti mampu meningkatkan akurasi secara signifikan pada dataset yang tidak seimbang. Beberapa studi menunjukkan bahwa SVM cenderung memiliki akurasi lebih tinggi dibandingkan *Naïve Bayes* dalam klasifikasi sentimen [10], khususnya pada platform seperti Shopee dan Tokopedia. Tahapan pra-pemrosesan, termasuk normalisasi teks, juga berkontribusi besar terhadap peningkatan performa model, penelitian yang dilakukan Prastyo et al. [11] membuktikan bahwa normalisasi slang berdampak signifikan pada peningkatan performa model dimana akurasi dari *Naïve Bayes* meningkat dari 65% menjadi 73% serta SVM dan *Random Forest* yang awalnya masing-masing mendapatkan akurasi 88% dan 89% naik menjadi 96% setelah diterapkan normalisasi slang. Di sisi lain, *Logistic Regression* dalam beberapa kasus mampu melampaui performa SVM dan *Naïve Bayes* [12], menegaskan bahwa efektivitas algoritma sangat dipengaruhi oleh karakteristik data dan pendekatan pra-pemrosesan yang diterapkan.

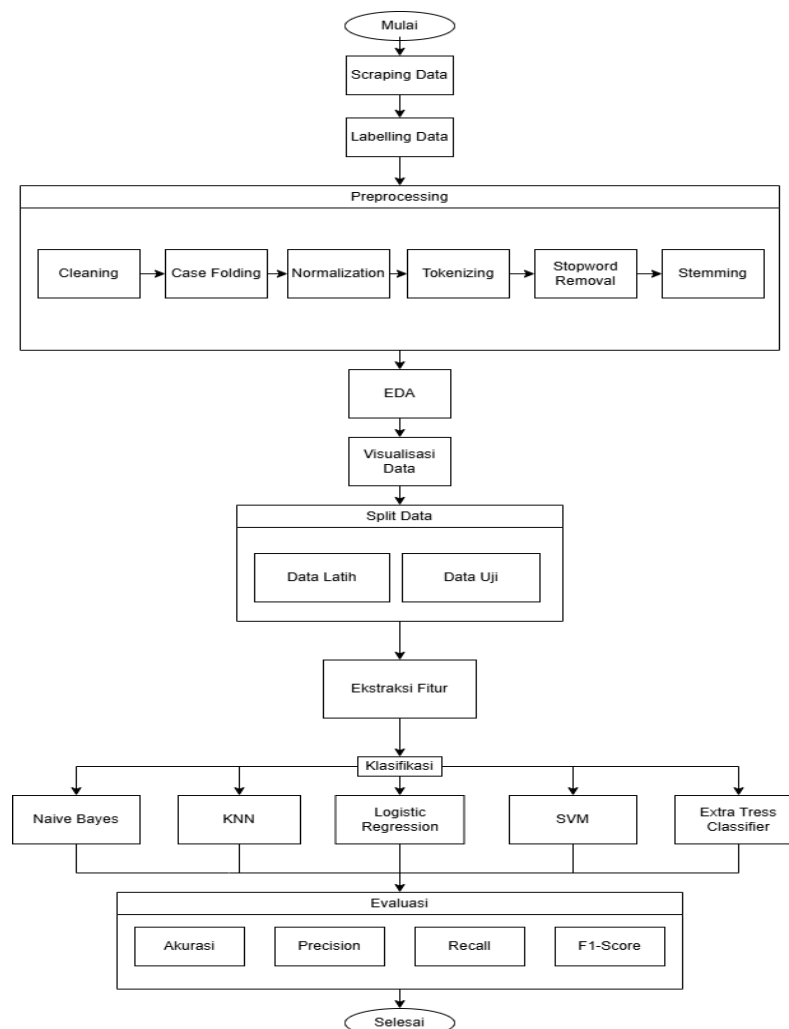
Dalam menilai performa model klasifikasi sentimen, akurasi bukan satu-satunya metrik yang relevan. Evaluasi juga perlu mempertimbangkan metrik lain seperti *precision*, *recall*, dan *F1-score*, terutama ketika data yang digunakan tidak seimbang [13]. Penanganan ketidakseimbangan data dan penerapan metode ensemble menjadi penting untuk menjaga stabilitas dan reliabilitas hasil klasifikasi [14]. Teknik ensemble seperti *Random Forest* dan *Extra Trees Classifier* semakin mendapat perhatian karena kemampuannya dalam meningkatkan akurasi dan stabilitas prediksi [15]. Namun, penggunaan *Extra Trees Classifier* dalam konteks ulasan Tokopedia masih relatif jarang, sehingga membuka peluang eksplorasi lebih lanjut dalam ranah e-commerce Indonesia.

Meskipun telah banyak penelitian dilakukan, sebagian besar hanya membandingkan dua hingga tiga algoritma dengan tahapan pra-pemrosesan yang sederhana. Studi yang menggabungkan normalisasi teks dengan algoritma ensemble seperti *Extra Trees Classifier*, khususnya untuk data ulasan Tokopedia, masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi secara komprehensif lima algoritma klasifikasi sentimen, yaitu KNN, *Naïve Bayes*, *Logistic Regression*, SVM, dan *Extra Trees Classifier*, dengan pendekatan pra-pemrosesan yang diperluas dan metrik evaluasi yang lebih lengkap. Diharapkan hasil penelitian ini dapat memberikan kontribusi signifikan dalam pemilihan algoritma yang paling tepat serta memperkuat penerapan *Natural Language Processing* (NLP) dalam sektor e-commerce di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis yang mencakup pengumpulan data, pengolahan data, eksplorasi data, dan evaluasi model klasifikasi. Alur keseluruhan proses digambarkan dalam bentuk *flowchart* yang mencerminkan langkah-langkah dari proses *Scraping* hingga evaluasi akhir model klasifikasi. Visualisasi alur tersebut disajikan pada Gambar 1.



Gambar 1. Flowchart tahapan penelitian

Gambar 1 menjelaskan tahapan penelitian yang dimulai dengan mengumpulkan data ulasan dari platform Tokopedia menggunakan scraping. Selanjutnya, data dilabelkan sesuai kategori sentiment. Sebelum digunakan dalam pemodelan, data yang telah dilabeli diproses melalui tahap preprocessing. Selanjutnya, dilakukan *Exploratory Data Analysis* (EDA), visualisasi, pembagian data menjadi data uji dan data latih, ekstraksi fitur, pelatihan model klasifikasi, dan evaluasi hasil dilakukan. Berikut adalah penjelasan dari masing-masing tahapan.

2.2 Scraping Data

Scraping data yang terdapat pada aplikasi playstore merupakan sebuah proses untuk pengambilan data secara otomatis pada halaman aplikasi yang terdapat di google playstore [16]. Informasi yang dikumpulkan umumnya meliputi ulasan pengguna, penilaian (rating), jumlah unduhan, deskripsi aplikasi, serta berbagai detail lain yang ditampilkan pada laman aplikasi tersebut. Proses ini dapat digunakan dalam sebuah penelitian atau sebuah analisis data untuk dapat memahami tentang sentimen pengguna, tren pada aplikasi, atau untuk memperoleh insight lainnya dari data yang tersedia.

2.3 Labelling Data

Labeling data merupakan bagian penting dari proses yang melibatkan pemberian label atau kategori pada data mentah yang telah dikumpulkan. Pelabelan dilakukan berdasarkan analisis isi teks, dengan mengklasifikasikan ulasan ke dalam dua kategori utama, yaitu positif, mencakup ulasan yang menggambarkan pengalaman menyenangkan,

kepuasan, atau pujian terhadap produk maupun layanan dan negatif, mencakup ulasan yang menunjukkan ketidakpuasan, keluhan, atau pengalaman buruk yang dialami pengguna. Proses pelabelan dilakukan oleh annotator manusia yang telah diberikan pedoman klasifikasi yang seragam, guna memastikan konsistensi dan menghindari bias persepsi antarpemilai. Kriteria pelabelan disusun secara sistematis agar interpretasi terhadap makna emosional dalam teks dapat dilakukan secara objektif dan replikatif.

2.4 Preprocessing

Preprocessing bertujuan untuk memastikan bahwa data teks yang digunakan dalam pemodelan klasifikasi berada dalam kondisi bersih, konsisten, dan siap untuk diproses secara komputasional. Proses ini terdiri dari beberapa langkah sistematis yang dirancang untuk meningkatkan kualitas representasi data dan efektivitas algoritma klasifikasi. Adapun tahapan yang dilakukan sebagai berikut:

2.4.1 Cleaning

Tahap ini merupakan tahap dimana penghapusan beberapa elemen yang tidak relevan dalam teks, seperti tanda baca, karakter khusus, angka, URL, serta simbol non-alfabetik yang dapat mengganggu analisis [17].

2.4.2 Case Folding

Semua huruf pada teks diubah menjadi huruf kecil (lowercase) untuk menjaga konsistensi dan menghindari redundansi akibat perbedaan kapitalisasi [18].

2.4.3 Normalization

Kata-kata tidak baku, singkatan, dan istilah informal diubah menjadi bentuk standar. Sebagai contoh, kata “gak” dinormalisasi menjadi “tidak” agar makna tetap terjaga dalam konteks analisis [15].

2.4.4 Stopword removal

Kata-kata umum yang tidak memiliki kontribusi signifikan terhadap analisis sentimen, seperti “yang”, “dan”, “di”, dan sejenisnya, dihapus dari korpus data. Penghapusan stopwords bertujuan untuk mengurangi noise dan meningkatkan efisiensi pemrosesan [19].

2.4.5 Tokenizing

Teks ulasan dipisah menjadi unit-unit kata atau token [20]. Proses ini memungkinkan analisis dilakukan pada tingkat kata, yang merupakan unit dasar dalam pemodelan teks.

2.4.6 Stemming

Kata-kata yang memiliki imbuhan diubah menjadi bentuk dasar atau akar kata. Sebagai contoh, kata “berlari” diubah menjadi “lari”. Proses stemming membantu menyederhanakan variasi kata dan memperkuat generalisasi model [21]. Tahapan pra-pemrosesan ini merupakan fondasi penting dalam pipeline analisis sentimen, karena kualitas input data sangat menentukan performa akhir dari model klasifikasi yang digunakan.

2.5 EDA

EDA bertujuan untuk memahami distribusi sentimen berdasarkan kategori rating pengguna serta karakteristik umum dalam teks ulasan. Analisis ini mencakup pemetaan jumlah ulasan positif dan negatif per tingkat rating, serta identifikasi kata-kata yang paling sering muncul. Visualisasi dilakukan menggunakan word cloud dan grafik frekuensi untuk mengungkap tema dominan yang relevan bagi proses klasifikasi.

2.6 Visualisasi Data

Visualisasi data biasanya dilakukan untuk mempermudah interpretasi hasil eksplorasi. Diagram batang digunakan untuk menampilkan frekuensi kemunculan kata-kata dominan dalam ulasan serta distribusi sentimen berdasarkan kategori rating pengguna. Representasi visual ini membantu mengidentifikasi pola umum dan mendukung analisis lanjutan dalam proses klasifikasi.

2.7 Split Data

Proses membagi data menjadi subset pelatihan dan pengujian merupakan teknik dalam pembelajaran mesin untuk memisahkan dataset menjadi dua bagian: satu untuk melatih model dan satu lagi untuk mengevaluasi kinerjanya [22].

- Data Train (Training Data):** Subset data yang dipakai untuk melatih model. Pada tahap ini, model mempelajari pola, hubungan, dan karakteristik penting dari data untuk membangun kemampuan prediktif. Umumnya mencakup sekitar 70%–80% dari total dataset.
- Data Testing (Testing Data):** Subset data yang digunakan untuk menguji performa model setelah pelatihan. Tujuannya untuk mengetahui sejauh mana model mampu menggeneralisasi pada data baru yang belum pernah dilihat. Biasanya terdiri dari 20%–30% dari dataset.

Pembagian ini penting untuk mencegah overfitting, agar model tidak hanya bagus pada data pelatihan tetapi juga tetap efektif saat digunakan pada data baru, sehingga evaluasi kinerjanya lebih objektif

2.8 Ekstraksi Fitur

Ekstraksi fitur dilakukan dengan metode *Term Frequency–Inverse Document Frequency* (TF–IDF), yang mengubah data teks menjadi representasi numerik berdasarkan tingkat kepentingan suatu kata dalam dokumen relatif terhadap keseluruhan korpus [19]. TF–IDF menggabungkan dua elemen pokok, yaitu: *Term Frequency* (TF) yang berarti frekuensi kemunculan kata dalam sebuah dokumen dan *Inverse Document Frequency* (IDF) yang merupakan ukuran seberapa jarang kata tersebut muncul di seluruh dokumen. Dengan pendekatan ini, kata-kata yang sering muncul namun kurang informatif (seperti stopwords) akan memiliki bobot rendah, sedangkan kata-kata yang lebih spesifik dan bermakna akan diberi bobot lebih tinggi, sehingga meningkatkan efektivitas model klasifikasi.

2.8.1 TF

Term Frequency digunakan untuk mengukur seberapa sering suatu kata (token) muncul dalam sebuah dokumen. Nilai TF mencerminkan tingkat kepentingan kata tersebut dalam konteks dokumen tertentu. Formula yang digunakan dalam penelitian ini adalah sebagai berikut:

$$TF(w, d) = 1 + \frac{f_{w,d}}{\sum_{t \in d} f_{t,d}} \quad (1)$$

di mana $f_{w,d}$ adalah jumlah kemunculan kata w dalam dokumen d , sedangkan $\sum_{t \in d} f_{t,d}$ merupakan total seluruh kata dalam dokumen. Penambahan konstanta 1 bertujuan untuk menghindari nilai nol pada kata yang sangat jarang muncul, sehingga tetap memberikan kontribusi dalam proses pembobotan. Nilai TF yang lebih tinggi menunjukkan bahwa kata tersebut lebih dominan dalam dokumen, dan akan berpengaruh dalam pembentukan vektor fitur untuk klasifikasi.

2.8.2 IDF

Inverse Document Frequency digunakan untuk mengukur tingkat kekhasan atau kelangkaan suatu kata dalam korpus dokumen. Semakin jarang sebuah kata muncul di seluruh dokumen, semakin tinggi nilai IDF-nya, yang berarti kata tersebut lebih informatif dan relevan untuk membedakan dokumen satu dengan lainnya.

$$IDF(w) = \log + \frac{N}{DF(w)} \quad (2)$$

Di mana N menyatakan jumlah total dokumen, sedangkan $DF(w)$ menunjukkan banyaknya dokumen yang memuat kata w . Nilai IDF yang tinggi berarti kata tersebut muncul di sedikit dokumen, sehingga memiliki bobot yang lebih besar dalam proses pembobotan fitur. Sebaliknya, kata-kata yang muncul di hampir semua dokumen (misalnya kata umum seperti “dan”, “adalah”, atau “seperti”) akan memiliki nilai IDF yang rendah, karena kurang memberikan informasi diskriminatif.

2.8.3 Perhitungan TF-IDF

Nilai bobot kata dihitung dengan mengalikan TF dan IDF

$$TF - IDF(w, d) = TF(w, d) \times IDF(w) \quad (3)$$

$TF(w, d) \times IDF(w)$ adalah perkalian antara rumus TF dan IDF. Metode ini memberikan bobot yang tinggi pada kata yang sering muncul di dokumen tertentu tetapi jarang muncul di dokumen lain, sehingga fitur yang dihasilkan lebih informatif untuk proses klasifikasi.

2.9 Klasifikasi

Lima algoritma klasifikasi yang berbeda digunakan dalam penelitian ini, yaitu Naïve Bayes, KNN, LR, SVM, dan Extra Trees Classifier. Kelima algoritma ini dipilih karena telah terbukti efektif dalam beberapa analisis sentimen. Selanjutnya adalah pembahasan dari masing-masing algoritma.

2.9.1 Multinomial Naïve Bayes

Multinomial Naïve Bayes (MNB) adalah algoritma klasifikasi berbasis probabilistik yang umum digunakan dalam analisis teks, seperti deteksi spam dan analisis sentimen. MNB mengasumsikan bahwa setiap fitur (kata) bersifat independen secara kondisional terhadap kelas dokumen. Model ini menghitung probabilitas posterior berdasarkan frekuensi kata dalam dokumen, lalu menetapkan dokumen ke kelas dengan probabilitas tertinggi. Representasi dokumen menggunakan vektor frekuensi kata, sehingga cocok untuk data teks berdimensi tinggi. Secara matematis, untuk kelas C_k dan dokumen $x = (x_1, x_2, \dots, x_n)$, Dimana x_i adalah frekuensi kemunculan kata ke- i , algoritma menghitung:

$$P(C_k | X) \propto P(C_k) \cdot \prod_{i=1}^n P(x_i | C_k) \quad (4)$$

Di mana $P(C_k)$ adalah probabilitas prior kelas C_k , dan $P(x_i|C_k)$ adalah probabilitas kondisional kemunculan kata ke- i pada kelas C_k , yang biasanya diestimasi menggunakan metode *Maximum Likelihood* dengan *Laplace smoothing* untuk mengatasi frekuensi nol:

$$P(x_i|C_k) = \frac{N_{i,k} + \alpha}{N_k + \alpha n} \quad (5)$$

Di mana $N_{i,k}$ adalah jumlah kemunculan kata i dalam dokumen kelas C_k , n adalah jumlah total kata dalam kamus, dan α adalah parameter smoothing (biasanya bernilai 1).

2.9.2 KNN

KNN merupakan algoritma berbasis instance yang mengklasifikasikan data uji berdasarkan kedekatan jarak ke data pelatihan. Jarak Euclidean antara titik data x dan titik data pelatihan x_i didefinisikan sebagai:

$$d(x, x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2} \quad (6)$$

dengan n jumlah fitur. Klasifikasi dilakukan dengan menentukan kelas mayoritas dari k tetangga terdekat [23]. KNN tidak melakukan proses pelatihan model secara *eksplisit*, sehingga sangat bergantung pada struktur data dan ukuran dataset.

2.9.3 Logistic Regression

LR adalah model statistik yang digunakan untuk klasifikasi biner dengan memperkirakan probabilitas suatu kejadian melalui fungsi sigmoid:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta^T x)}} \quad (7)$$

Di mana β adalah vektor koefisien, β_0 adalah *intercept*, dan x adalah vektor fitur. Probabilitas ini kemudian digunakan untuk memutuskan kelas berdasarkan ambang batas tertentu (biasanya 0.5). LR memiliki interpretasi probabilistik yang jelas dan mampu memberikan bobot fitur yang dapat diinterpretasi, membuatnya populer dalam analisis teks dan klasifikasi biner.

2.9.4 SVM

SVM mencari hyperplane optimal yang memaksimalkan margin antar kelas. Fungsi objektif SVM untuk data linear separable dirumuskan sebagai

$$\min_{w,b} \frac{1}{2} \|w\|^2 = \quad (8)$$

Dengan kendala

$$y_i(w^T x_i + b) \geq 1, \quad i = 1, \dots, n \quad (9)$$

Di mana y_i adalah label kelas, x_i adalah fitur data, w adalah bobot, dan b adalah bias. SVM juga dapat menggunakan kernel untuk mengatasi data yang tidak dapat dipisahkan secara linear. Algoritma ini dikenal unggul dalam klasifikasi teks dan dataset berdimensi tinggi dengan performa yang stabil.

2.9.5 Extra Trees Classifier

Extra Trees Classifier merupakan algoritma ensemble yang membangun sejumlah pohon keputusan secara acak dengan cara memilih split secara acak pada fitur dan *threshold* [24]. Setiap pohon memberikan prediksi, dan hasil akhir diperoleh melalui voting mayoritas.

2.10 Evaluasi

Evaluasi model dilakukan untuk mengukur keakuratan model dalam memprediksi data baru. Klasifikasi dinilai dengan metrik seperti *akurasi*, *presisi*, *recall*, dan *F1-Score*. Tujuan utama proses adalah untuk memastikan bahwa model dapat mengenali pola dengan baik dan dapat digunakan dengan andal pada data baru.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$Presisi = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F1 - Score = 2x \frac{Presisi \times Recall}{Presisi+Recall} \quad (13)$$

True Positive (TP) merupakan total data yang sebenarnya berada dalam kelas positif yang diprediksi secara tepat oleh model, *False Positive* (FP) menggambarkan kesalahan model ketika memprediksi data sebagai positif, padahal

sebenarnya data tersebut negatif; *True Negative* (TN) mengacu pada jumlah data yang sebenarnya berasal dari kelas negatif yang diprediksi secara tepat oleh model dan *False Negative* (FN) menggambarkan kesalahan di mana model memprediksi data secara tidak akurat di mana model memprediksi data sebagai negatif tetapi kenyataannya data tersebut merupakan bagian dari kelas positif.

3. HASIL DAN PEMBAHASAN

3.1 Scraping Data

Data ulasan pengguna yang digunakan pada penelitian ini yaitu ulasan pengguna aplikasi Tokopedia yang diperoleh melalui teknik web scraping. Proses pengambilan data dilakukan dalam rentang waktu mulai dari 3 Mei 2025 pukul 00.15 hingga 3 Agustus 2025 pukul 14.10. Total ulasan yang berhasil dikumpulkan dalam periode tersebut berjumlah 3.160 entri, mencakup atribut seperti reviewID, userName, content, score, dan at, yang ditampilkan pada Tabel 1.

Tabel 1. Data Hasil Proses Scraping

reviewID	userName	content	score	at
c097d407-6c24-4dba-afd7-9c306c5167de	Pengguna Google	Aplikasi pada dasarnya bagus, tetapi ...	3.0	20/07/2025 10:28
9b921835-a04c-4f70-82d5-22ec98defd5f	Pengguna Google	transaksi di tokopedia cukup simple ...	5.0	01/07/2025 11:07:00
7e2cc691-959a-4baa-abd3-398a9700ca3b	Pengguna Google	saya kecewa, cek out barang ...	1.0	25/07/2025 19:49:00

Tabel 1 merupakan hasil proses pengumpulan data yang dilakukan menggunakan web scraping. Dari pengumpulan data ini, diperoleh data reviewID, userName, content, score, dan at dari aplikasi Tokopedia.

3.2 Labelling Data

Tahap pelabelan data berfungsi untuk mengkategorikan apakah komentar tersebut termasuk dalam kelas sentimen negatif atau positif. Tahap pelabelan ini dilakukan dengan cara membagi data ulasan ke dalam dua kelas sentimen yaitu sentimen positif dengan rating 4 dan 5 dan sentimen negatif dengan rating 1, 2, dan 3. Dilakukan juga tahap penyaringan variabel untuk menghilangkan atribut yang tidak relevan.

Tabel 2. Data Hasil Pelabelan

Content	Score	Label
Aplikasi pada dasarnya bagus, terakhir ...	3.0	Negatif
transaksi di tokopedia cukup simple	5.0	Positif
saya kecewa, cek out barang ...	1.0	Negatif

Tabel 2 memperlihatkan hasil dari proses pemberian label pada setiap ulasan pengguna, di mana ulasan tersebut dikategorikan ke dalam dua kelompok, yaitu positif dan negatif. Pemberian label ini dilakukan berdasarkan skor yang diperoleh dari masing-masing ulasan, sehingga setiap ulasan dapat diidentifikasi sesuai dengan sentimen yang diungkapkan oleh pengguna. Tabel tersebut juga sudah menampilkan hasil penyaringan variabel pada dataset ulasan pengguna Tokopedia, di mana hanya tiga atribut utama yang dipertahankan, yaitu *content*, *score*, dan *label*. Penyaringan ini berfungsi untuk menghapus kolom yang tidak relevan supaya data lebih terfokus dan siap digunakan pada tahap *preprocessing* serta analisis klasifikasi sentimen selanjutnya.

3.3 Preprocessing

Tahap *preprocessing* bertujuan untuk memastikan kualitas dan konsistensinya sebelum digunakan dalam analisis klasifikasi sentimen. Tahapan ini terdiri dari langkah penting yaitu *cleaning*, untuk menghapus karakter non-alfabet dan simbol yang tidak relevan, dilanjutkan dengan *case folding* untuk menyeragamkan huruf menjadi format kecil. Selanjutnya dilakukan *normalization* guna menyamakan bentuk kata yang bervariasi namun bermakna sama, diikuti oleh *stopword removal* untuk menghapus kata-kata umum yang tidak memberikan kontribusi penting terhadap makna. Proses *tokenizing* kemudian memecah teks menjadi unit kata, dan ditutup dengan *stemming* untuk mengubah kata ke bentuk dasarnya. Melalui serangkaian tahap ini, data teks menjadi lebih bersih, terstruktur, dan siap digunakan dalam proses analisis berikutnya.

Tabel 3. Hasil Preprocessing

Tahapan	Content
Data awal	Aplikasi pada dasarnya bagus, tetapi terakhir ini ada banyak gangguan teknis di Tokped, baik di aplikasi maupun di situs web. Seperti tidak bisa loading halaman, gambar tidak muncul, pembayaran macet dan pada umumnya lambat, sehingga pembelian harus ditunda dulu. Terima kasih.



Tahapan	Content
<i>Cleaning</i>	Aplikasi pada dasarnya bagus tetapi terakhir ini ada banyak gangguan teknis di Tokped baik di aplikasi maupun di situs web Seperti tidak bisa loading halaman gambar tidak muncul pembayaran macet dan pada umumnya lambat sehingga pembelian harus ditunda dulu Terima kasih
<i>Case Folding</i>	aplikasi pada dasarnya bagus tetapi terakhir ini ada banyak gangguan teknis di tokped baik di aplikasi maupun di situs web seperti tidak bisa loading halaman gambar tidak muncul pembayaran macet dan pada umumnya lambat sehingga pembelian harus ditunda dulu terima kasih
<i>Normalization</i>	aplikasi pada dasarnya bagus tetapi terakhir ini ada banyak gangguan teknis di tokopedia baik di aplikasi maupun di situs web seperti tidak bisa loading halaman gambar tidak muncul pembayaran macet dan pada umumnya lambat sehingga pembelian harus ditunda dulu terima kasih
<i>Stopword Removal</i>	Aplikasi pada dasarnya bagus tetapi terakhir ini ada banyak gangguan teknis di Tokped baik di aplikasi maupun di situs web Seperti tidak bisa loading halaman gambar tidak muncul pembayaran macet dan pada umumnya lambat sehingga pembelian harus ditunda dulu Terima kasih
<i>Tokenizing</i>	Aplikasi pada dasarnya bagus tetapi terakhir ini ada banyak gangguan teknis di Tokped baik di aplikasi maupun di situs web Seperti tidak bisa loading halaman gambar tidak muncul pembayaran macet dan pada umumnya lambat sehingga pembelian harus ditunda dulu Terima kasih
<i>Stemming</i>	Aplikasi pada dasarnya bagus tetapi terakhir ini ada banyak gangguan teknis di Tokped baik di aplikasi maupun di situs web Seperti tidak bisa loading halaman gambar tidak muncul pembayaran macet dan pada umumnya lambat sehingga pembelian harus ditunda dulu Terima kasih

Tabel 3 menampilkan hasil penerapan setiap tahapan *preprocessing* pada data ulasan pengguna Tokopedia. Proses diawali dengan tahap *cleaning* yaitu penghapusan karakter non-alfabet, tanda baca, angka, dan simbol yang kurang relevan untuk memastikan teks bebas dari elemen yang dapat mengganggu analisis. Tahap berikutnya adalah *case folding* dengan mengubah seluruh huruf menjadi bentuk kecil agar format penulisan lebih seragam. Selanjutnya dilakukan *normalization* guna penyeragaman kata tidak baku atau singkatan menjadi bentuk yang benar dan *stopword removal* yang melakukan penghapusan kata umum seperti “ke”, “dan”, atau “di” yang tidak memiliki makna penting pada konteks sentimen. Proses kemudian dilanjutkan dengan *tokenizing* yaitu pemecahan teks menjadi kata-kata terpisah dan ditutup dengan *stemming* yang mengganti setiap kata ke bentuk dasarnya. Setiap tahap tersebut secara bertahap membuat teks menjadi lebih bersih, konsisten, dan siap digunakan dalam tahap analisis sentimen.

3.4 EDA

Tahapan *Exploratory Data Analysis* dilakukan untuk memperoleh gambaran awal mengenai struktur dan karakteristik data. Analisis ini mencakup identifikasi jumlah total data, serta pemeriksaan distribusi label sentimen

Tabel 4. Jumlah Ulasan Tiap Rating

Rating	Jumlah
1.0	1734
2.0	328
3.0	275
4.0	158
5.0	665
Total	3160

Berdasarkan Tabel 4, ulasan dengan rating 1 memperoleh ulasan terbanyak yaitu 1.734 ulasan. Rating 5 menempati posisi kedua dengan 665 ulasan, rating 2 menempati posisi ketiga dengan 328 ulasan, disusul rating 3 berada d posisi keempat dengan 275, sedangkan jumlah ulasan terendah terdapat pada rating 4, yaitu sebanyak 158 ulasan.

3.5 Visualisasi Data

Visualisasi dilakukan menggunakan teknik *wordcloud* untuk mengidentifikasi kata-kata yang paling sering digunakan dalam ulasan pengguna. Wordcloud ini dibagi berdasarkan label sentimen, di mana warna cerah digunakan untuk mewakili kata-kata dari ulasan berlabel positif, dan warna gelap untuk kata-kata dari ulasan berlabel negatif. Tampilan visual ini membantu memperjelas perbedaan pola bahasa antara kedua kategori sentimen.



Gambar 2. Wordcloud Sentimen Positif

Gambar 2 memperlihatkan visualisasi untuk label positif dimana menampilkan kata-kata yang paling sering muncul dalam ulasan pengguna berupa tokopedia, barang, aplikasi, bagus, lengkap, dan mudah. Kata-kata tersebut mencerminkan pengalaman pengguna yang cenderung memuaskan. Temuan ini menunjukkan bahwa sebagian besar pengguna Tokopedia menilai kinerja aplikasi dan kualitas produk secara positif, sehingga memperkuat validitas hasil analisis sentimen yang mengindikasikan dominasi ulasan bernada positif pada dataset.



Gambar 3. Wordcloud Sentimen Negatif

Sementara itu, Gambar 3 memperlihatkan visualisasi wordcloud untuk label negatif memperlihatkan kata-kata yang paling sering muncul pada ulasan bernada keluhan atau ketidakpuasan, seperti ribet, susah, kecewa, tolong, transaksi, tipu, komplain, dan paket. Kata-kata tersebut mencerminkan isu-isu yang sering dikeluhkan pengguna, seperti masalah pengiriman, pelayanan penjual, dan kendala transaksi. Dari visualisasi diatas, dapat disimpulkan dari kedua sentimen tersebut memiliki beberapa kesamaan kata yang serupa, seperti aplikasi, tokopedia, barang, jual, beli yang terdapat pada visualisasi wordcloud sentimen positif dan negatif. Visualisasi ini memperkuat pemahaman terhadap pola bahasa yang khas dalam ulasan negatif.

3.6 Split Data

Tahap selanjutnya adalah membagi data latih (*training data*) dan data uji (*testing data*) adalah langkah selanjutnya dalam penelitian ini. 80% data untuk pelatihan dan 20% untuk pengujian . Data uji digunakan untuk menguji kinerja model terhadap data yang belum pernah diproses sebelumnya. Sementara itu, data latih digunakan untuk membuat model dengan menyesuaikan parameter untuk memungkinkan prediksi atau klasifikasi yang akurat.

3.7 Ekstraksi Fitur

Dalam tahap Ekstraksi Fitur, penelitian ini menggunakan pendekatan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengkonversi ulasan pengguna terkait aplikasi Tokopedia menjadi bentuk numerik yang supaya dapat dianalisis oleh algoritma *machine learning*. TF-IDF menilai relevansi suatu kata dalam sebuah komentar berdasarkan kemunculannya di dokumen tersebut serta di seluruh korpus data. Kata - kata yang bersifat umum dan tidak memberikan banyak informasi akan diberi bobo rendah, sedangkan kata-kata yang jarang muncul tetapi memiliki informasi penting misalnya kata-kata yang mengindikasikan keluhan, fitur, atau pengalaman layanan, akan diberi bobot lebih tinggi. Proses ini menghasilkan sebuah sparse matrix, di mana setiap kolom.

Tabel 5. Hasil TF-IDF

Term	TFIDF SUM
tokopedia	41.579699
barang	30.612780
aplikasi	29.483556
kirim	27.233901
toko	25.134570

Tabel 5 memperlihatkan hasil perhitungan TF-IDF, di mana kata-kata seperti "tokopedia," "barang," dan "aplikasi" memiliki nilai TF-IDF tertinggi, menunjukan dominasi dan relevansinya dalam ulasan pengguna Tokopedia.

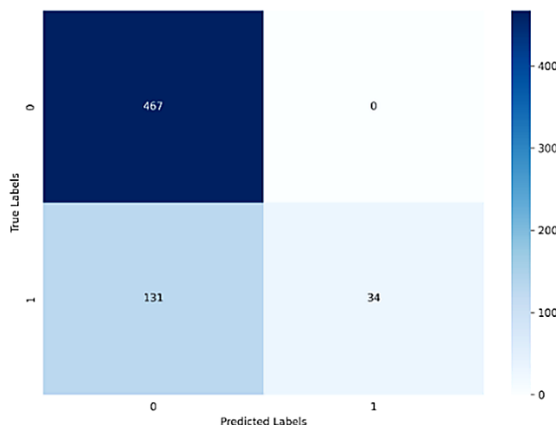
3.8 Klasifikasi dan Evaluasi

Tahap klasifikasi dan evaluasi dilakukan menggunakan lima algoritma pembelajaran mesin, yaitu *Naïve Bayes*, KNN, *Logistic Regression*, SVM, dan *Extra Trees Classifier*. Nilai evaluasi model diperoleh dari hasil perbandingan antara prediksi yang dihasilkan algoritma dengan label sebenarnya melalui *confusion matrix*. *Precision* digunakan untuk mengukur seberapa tepat model dalam memprediksi kelas tertentu, yaitu rasio antara jumlah prediksi benar pada suatu kelas dengan total prediksi yang dihasilkan untuk kelas tersebut. *Recall* menjelaskan sampai sejauh mana mode dapat mengenali seluruh data yang sebenarnya termasuk dalam kelas yang sama. Sementara itu, *f1-score* dihitung sebagai rata-rata harmonik antara *precision* dan *recall* yang berfungsi untuk memberikan gambaran keseimbangan antara keduanya. Adapun *support* merepresentasikan jumlah data aktual pada masing-masing kelas yang digunakan untuk evaluasi. Setelah metrik per kelas diperoleh, *macro average* dihitung sebagai rata-rata sederhana antar kelas tanpa bobot, sedangkan *weighted average* merupakan rata-rata berbobot berdasarkan *support* sehingga kelas dengan jumlah data lebih besar memberikan kontribusi lebih besar pada nilai akhir. Dengan demikian, hasil evaluasi dapat memberikan gambaran umum tentang sejauh mana model bekerja baik pada keseluruhan data maupun pada tiap kelas secara terpisah.

Tabel 6. Hasil Evaluasi Kinerja *Naïve Bayes*

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Negatif	0.78	1.00	0.87	467
Positif	1.00	0.20	0.34	165
Accuracy			0.79	632
<i>Macro average</i>	0.89	0.60	0.60	632
<i>Weighted average</i>	0.83	0.79	0.73	632

Pada Tabel 6 diperlihatkan hasil evaluasi menggunakan algoritma *Naïve Bayes* menunjukkan akurasi sebesar 79%. Untuk label negatif, diperoleh nilai presisi sebesar 78%, *recall* 100%, dan *f1-score* 87%. Sementara itu, untuk label positif, presisi mencapai 100%, namun *recall* hanya 20%, dengan *f1-score* sebesar 34%. Hasil ini mengindikasikan bawasanya *Naïve Bayes* cukup baik ketika mengenali komentar negatif, namun kurang optimal dalam mendeteksi komentar positif, kemungkinan disebabkan oleh ketidakseimbangan distribusi data. Secara keseluruhan, model *Naïve Bayes* menunjukkan karakteristik prediksi yang konservatif dan cenderung fokus pada kelas dengan jumlah data lebih besar.



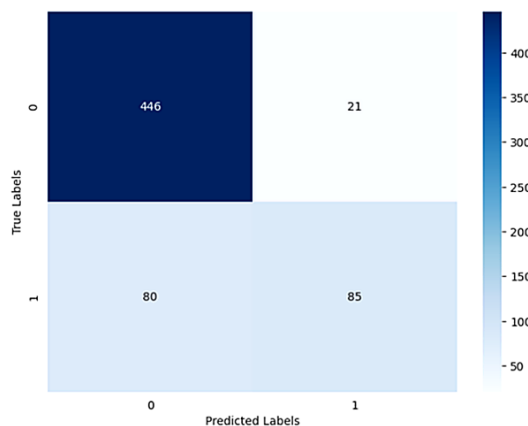
Gambar 4. Hasil Confusion Matrix *Naïve Bayes*

Berdasarkan Gambar 4, diperlihatkan *confusion matrix* dari *Naïve Bayes* menghasilkan nilai sebagai berikut: *true negative* sebanyak 467, *false positive* 0, *false negative* 131, dan *true positive* 34. Kesimpulan ini menjelaskan bahwa model dapat mengklasifikasikan ulasan negatif dengan sangat baik (tanpa kesalahan FP), namun masih menghasilkan cukup banyak kesalahan klasifikasi pada ulasan positif (FN yang cukup tinggi), yang berdampak pada rendahnya *recall* untuk kelas positif. Secara keseluruhan Hal ini menegaskan bahwa performa *Naïve Bayes* lebih dominan pada kelas negatif dengan akurasi tinggi namun rentan melewatkan sentimen positif.

Tabel 7. Hasil Evaluasi Kinerja KNN

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.84	0.95	0.89	467
Positif	0.80	0.51	0.62	165
Accuracy			0.84	632
<i>Macro average</i>	0.82	0.73	0.76	632
<i>Weighted average</i>	0.83	0.84	0.82	632

Pada Tabel 7 menunjukkan bahwa klasifikasi menggunakan KNN menghasilkan akurasi sebesar 84%, dengan presisi komentar negatif sebesar 84%, *recall* 95%, dan *f1-score* 89%. Sementara itu, untuk komentar positif, presisinya sebesar 80%, *recall* 51%, dan *f1-score* 62%. Hasil tersebut menunjukkan bahwa KNN mampu mengidentifikasi komentar negatif secara akurat sekaligus mempertahankan tingkat presisi yang cukup baik pada kelas positif. Secara keseluruhan, model ini menunjukkan performa yang relatif seimbang dengan kemampuan klasifikasi yang stabil meskipun masih dipengaruhi oleh ketimpangan data.



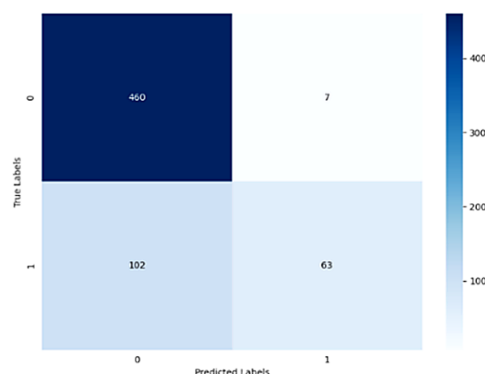
Gambar 5. Hasil Confusion Matrix KNN

Pada Gambar 5 diperlihatkan *confusion matrix* dari KNN, yang menghasilkan nilai *true negative* sebesar 446, dan *true positive* 85. Namun, model juga melakukan kesalahan klasifikasi sebanyak 21 sampel negatif yang terdeteksi sebagai positif (*false positive*) serta 80 sampel positif yang salah diklasifikasikan sebagai negatif (*false negative*). Model menunjukkan keahlian yang cukup baik ketika mengenali ulasan bernada negatif, tetapi masih memiliki keterbatasan dalam mendeteksi ulasan positif. Jumlah *true positive* yang lebih tinggi dibanding model *Naïve Bayes* juga menunjukkan bahwa KNN cukup responsif terhadap data positif, meski kesalahan pada *false negative* masih terlihat. Dengan demikian, KNN mampu menampilkan keseimbangan performa yang cukup baik antara kedua kelas sentimen. Secara umum, hasil ini menunjukkan bahwa KNN memiliki stabilitas performa yang baik pada kelas dominan, meskipun akurasinya masih dipengaruhi oleh ketimpangan data antar kelas.

Tabel 8. Hasil Evaluasi Kinerja LR

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.81	0.98	0.89	467
Positif	0.90	0.38	0.53	165
<i>Accuracy</i>			0.82	632
<i>Macro average</i>	0.85	0.68	0.71	632
<i>Weighted average</i>	0.83	0.82	0.80	632

Tabel 8 memperlihatkan menunjukkan akurasi untuk *Logistic Regression* sebesar 82%. Untuk komentar negatif, model menghasilkan presisi sebesar 81%, *recall* 98%, dan *f1-score* 89%, menandakan kemampuan tinggi dalam mendeteksi komentar negatif. Namun, untuk komentar positif, presisi mencapai 90%, tetapi *recall* hanya 38%, dengan *f1-score* sebesar 53%, yang mengindikasikan ketidakseimbangan klasifikasi antar kelas. Temuan ini mengindikasikan bahwa *Logistic Regression* cenderung lebih fokus pada deteksi sentimen negatif dibandingkan positif. Secara keseluruhan, model bekerja efektif untuk mengenali ulasan bernada negatif, namun masih memerlukan peningkatan dalam mendeteksi ulasan positif agar hasil klasifikasi menjadi lebih seimbang.



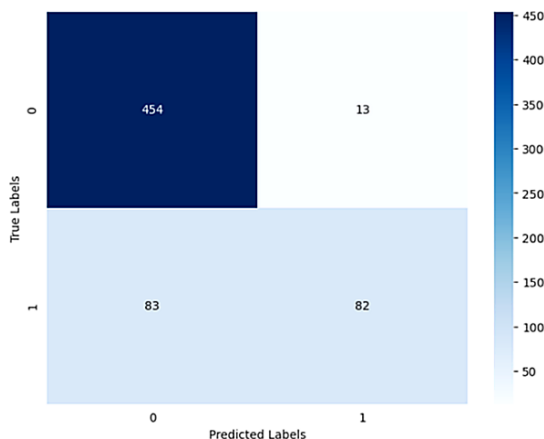
Gambar 6. Hasil Confusion Matrix LR

Gambar 6 menyajikan *confusion matrix* dari kinerja *Logistic Regression* dengan nilai *true negative* sebesar 460, *false positive* sebesar 7, *false negative* 102, dan *true positive* 63. Hasil ini menunjukkan bahwa *Logistic Regression* mampu mengklasifikasikan komentar negatif dengan sangat baik karena tingkat kesalahan prediksinya rendah. Namun demikian, pada kelas positif masih terdapat cukup banyak kesalahan klasifikasi, di mana sebagian data positif terdeteksi sebagai negatif. Kondisi ini mengindikasikan adanya ketidakseimbangan data yang menyebabkan model lebih berfokus pada kelas negatif selama proses pelatihan. Secara keseluruhan, model ini cenderung bias terhadap kelas negatif, yang menandakan bahwa peningkatan diperlukan pada tahap penyeimbangan data atau penyesuaian parameter agar performa terhadap kelas positif dapat meningkat.

Tabel 9. Hasil Evaluasi Kinerja SVM

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Negatif	0.84	0.97	0.90	467
Positif	0.86	0.49	0.63	165
<i>Accuracy</i>			0.85	632
<i>Macro avg</i>	0.85	0.73	0.76	632
<i>Weighted avg</i>	0.85	0.84	0.83	632

Tabel 9 menunjukkan akurasi dari algoritma SVM sebesar 85%. Untuk komentar negatif, model menghasilkan presisi sebesar 84%, *recall* 97%, dan *f1-score* 90%, menandakan kemampuan yang sangat baik dalam mengenali komentar negatif, performa ini menunjukkan bahwa model memiliki kecenderungan fokus pada kelas dengan jumlah data yang lebih dominan selama proses pelatihan. Sementara itu, untuk komentar positif, presisi mencapai 86%, *recall* 49%, dan *f1-score* 63%, SVM masih kesulitan mendeteksi seluruh ulasan positif dengan akurat. Namun nilai-nilai ini menunjukkan bahwa *SVM* mampu memberikan keseimbangan yang lebih baik antar kelas dibanding model sebelumnya. Secara Keseluruhan *SVM* menampilkan kinerja yang kuat dengan kemampuan pemisahan data yang efisien serta hasil klasifikasi yang stabil pada kedua kelas sentimen seimbang.



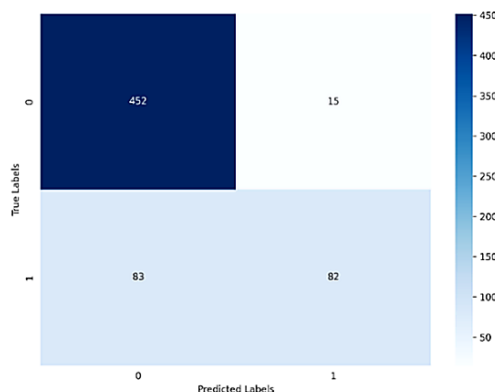
Gambar 7. Hasil Confusion Matrix SVM

Pada Gambar 7 ditampilkan hasil evaluasi algoritma SVM menggunakan *confusion matrix*, dengan nilai *true negative* sebesar 454, *false positive* 13, *false negative* 83, dan *true positive* 82 menunjukkan bahwa meskipun performa terhadap komentar positif belum optimal, SVM tetap lebih seimbang dibandingkan algoritma lainnya dalam menangani kedua kelas. Penelitian ini menunjukkan bahwa model SVM memiliki performa yang cukup stabil dan relatif seimbang dalam mendeteksi kedua kelas dibandingkan algoritma lainnya. Meskipun begitu, akurasi terhadap ulasan positif masih perlu ditingkatkan agar model dapat menghasilkan klasifikasi yang lebih proporsional dan tidak bias terhadap kelas negatif.

Tabel 10. Hasil Evaluasi Kinerja Extra Trees Classifier

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.84	0.96	0.90	467
Positif	0.84	0.49	0.62	165
<i>Accuracy</i>			0.84	632
<i>Macro avg</i>	0.84	0.73	0.76	632
<i>Weighted avg</i>	0.84	0.84	0.83	632

Tabel 10 menunjukkan bahwa model ini menghasilkan akurasi sebesar 84%. Untuk komentar negatif, algoritma ETC memperlihatkan performa yang dapat dikatakan sangat baik dengan presisi sebesar 84%, *recall* 96%, dan *f1-score* 90%. Sementara itu, untuk komentar positif, presisi juga mencapai 84%, namun *recall* hanya 49%, sehingga *f1-score* berada di angka 62%. Secara keseluruhan, Extra Trees Classifier menunjukkan kinerja yang stabil dan seimbang dalam mengenali kedua kelas, meskipun sensitivitas terhadap komentar positif masih dapat ditingkatkan.



Gambar 8. Hasil Confusion Matrix Extra Trees Classifier

Pada Gambar 8 ditampilkan hasil evaluasi algoritma Extra Trees Classifier menggunakan confusion matrix, dengan nilai *true negative* sebesar 452, *false positif* 15, *false negative* 83, dan *true positive* 82. Secara keseluruhan, model ini menampilkan kinerja yang cukup stabil dan seimbang di antara kedua kelas, dengan kecenderungan sedikit lebih baik dalam menangani sentimen negatif dibandingkan positif. Hasil *Confusion Matrix* dari ETC terlihat hampir sama dengan model SVM pada Gambar 7 karena keduanya menampilkan jumlah kesalahan klasifikasi yang seimbang antara kelas negatif dan positif. Namun, akurasi ETC sedikit lebih rendah dibandingkan SVM karena perbedaan cara kerja kedua algoritma. SVM membentuk bidang pemisah (*hyperplane*) optimal untuk memaksimalkan jarak antar kelas, sedangkan ETC menggunakan pendekatan *ensemble* pohon keputusan yang menggabungkan hasil dari banyak pohon acak. Pendekatan *ensemble* ini membuat model lebih stabil terhadap *noise*, tetapi kurang presisi dalam menentukan batas antar kelas, sehingga hasilnya tetap seimbang namun dengan akurasi sedikit di bawah SVM.

Tabel 11. Perbandingan Berdasarkan Metode Terdahulu

Peneliti	Dataset	Model	Hasil
Ernawati et al. (2023)	Ulasan Tokopedia (produk Kesehatan)	NB	Akurasi 88%, mayoritas ulasan positif
Setiawan et al. (2024)	Ulasan aplikasi DANA	XGBoost, SVM, ETC	XGBoost dan SVM: 93%, ETC: 92% (F1-Score: 96% pada K=6)
Pratmanto et al. (2025)	Ulasan Tokopedia	KNN vs NB	KNN: 80%, NB: 71,5%
Saputra et al. (2025)	Kaggle	NB vs SVM	SVM: 88%, NB: 72%
Penelitian ini (2025)	Ulasan Tokopedia	NB, KNN, LR, SVM, ETC	SVM tertinggi (85%, KNN dan ETC : 84%, LR: 82%, NB: 79%.

Pada tabel 11 ini memperlihatkan bahwa sebagian besar penelitian sebelumnya hanya membandingkan model dengan cakupan yang terbatas yaitu hanya dua atau tiga algoritma, penelitian yang dilakukan pada tahun 2023 oleh Ernawati et al. Dengan menggunakan model NB menunjukkan hasil akurasi yang cukup tinggi yaitu 88% dengan mayoritas ulasan positif, lalu di tahun berikutnya Setiawan et al. melakukan penelitian dengan menggunakan dataset ulasan aplikasi DANA pada ruang lingkup yang lebih luas dengan menggunakan 3 model yang menunjukkan performa tinggi pada algoritma SVM, XGBoost, dan ETC dengan F1-Score mencapai 96% untuk ETC. Penelitian menggunakan ulasan tokopedia ini juga dilakukan Pratmanto et al. pada tahun 2025 dengan hasil KNN memperoleh akurasi lebih unggul sebesar 80% dibanding NB yang hanya memperoleh 71,5%. Penelitian selanjutnya yang dilakukan oleh Saputra et al. dengan memadukan teknik seperti Word2Vec dan SMOTE menunjukkan konsistensi SVM yang mendapatkan akurasi lebih tinggi dibanding NB. Kontribusi penelitian ini terletak pada luas cakupan evaluasinya dengan menguji 5 algoritma sekaligus, sehingga memberikan gambaran yang lebih komprehensif.

4. KESIMPULAN

Penelitian ini memberikan pemahaman yang cukup luas dan mendalam mengenai perbandingan kinerja dari lima model klasifikasi yang telah dijabarkan. Hasil evaluasi dari penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memberikan performa terbaik dengan akurasi tertinggi sebesar 85%, diikuti oleh *K-Nearest Neighbors* (KNN) dan *Extra Trees Classifier* yang masing-masing mencapai 84%. *Logistic Regression* memperoleh akurasi sebesar 82%, sementara *Naïve Bayes*(NB) mencatatkan akurasi terendah sebesar 79%. Meskipun terdapat perbedaan akurasi, seluruh algoritma menunjukkan kemampuan yang sangat baik dalam mendeteksi komentar negatif, dengan nilai recall minimal 95%. Bahkan, *Naïve Bayes* mencapai recall sempurna sebesar 100% untuk kelas negatif. Namun, performa dalam mengidentifikasi komentar positif cenderung lebih rendah, terutama pada *Naïve Bayes* yang

menunjukkan penurunan recall secara signifikan. Temuan ini menegaskan bahwa SVM memiliki keunggulan untuk data teks berdimensi tinggi sekaligus menunjukkan stabilitas *Extra Trees Classifier* (ETC) pada skenario ulasan yang tidak seimbang. Untuk penelitian selanjutnya, disarankan untuk menerapkan penyeimbangan data (misalkan dengan memberikan bobot kelas, *oversampling/undersampling*) atau melakukan penyesuaian parameter model guna meningkatkan kinerja klasifikasi pada kelas positif. Selain itu, dapat mencoba dengan penambahan fitur sederhana seperti n-gram atau karakter serta dengan penggunaan model bahasa indonesia pralatih (seperti *IndoBert*) yang dapat lebih peka terhadap konteks. Upaya ini diharapkan dapat menghasilkan distribusi performa yang lebih seimbang antar kelas, serta meningkatkan akurasi keseluruhan dalam sistem klasifikasi komentar.

REFERENCES

- [1] A. Simanungkalit, J. P. P. Naibaho, and A. De Kweldju, “Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Shopee Menggunakan Algoritma Naïve Bayes” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 3, pp. 659–670, Mar. 2023. doi: 10.35889/jutisi.v13i1.1826.
- [2] A. Wardhana, M. Pradana, H. Shabira, D. M. A. Buana, D. W. Nugraha, and K. S. Sandi, “The Influence of Consumer Behavior on Purchasing Decision Process of Tokopedia E-Commerce Customers in Indonesia,” *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management (IEOM)*, vol. 11 no. 1, pp. 5929–5934, Mar. 2021, doi: 10.46254/an11.20210998.
- [3] N. M. Al Ghazali and Y. Sibaroni, “Sentiment Classification in E-Commerce Using Naïve Bayes and Combined Lexicon - N-Gram Features,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1257–1271, Mar. 2025, doi: 10.29100/jupi.v10i2.6157.
- [4] A. Ernawati, A. Ofa Sari, S. Nurhaliza Sofyan, M. Iqbal, and R. Farta Wijaya, “Implementasi Algoritma Naïve Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan,” *Bulletin of Information Technology (BIT)*, vol. 4, no. 4, pp. 533–543, Dec. 2023, doi: 10.47065/bit.v3i1.1090.
- [5] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiasari, “Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE,” *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [6] D. Pratmanto, R. Wijianto, A. Widayanto, and Ubaidilah, “Komparasi K-Nearest Neighbors (KNN) dan Naïve Bayes pada Klasifikasi Sentimen Ulasan Aplikasi Tokopedia di Google Play Store,” *Informatics and Computer Engineering Journal*, vol. 5, no. 2, pp. 75–86, Aug. 2025, doi: 10.31294/icej.v5i2.8939.
- [7] M. J. Setiawan and V. R. S. Nastiti, “DANA App Sentiment Analysis: Comparison of XGBoost, SVM, and Extra Trees,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 3, pp. 337–345, Nov. 2024, doi: 10.32736/sisfokom.v13i3.2239.
- [8] J. Saputra, L. Maryani, R. Rahmaddeni, D. Wulandari, and W. Eka, “Analisis Performa Naive Bayes dan SVM terhadap Sentimen Teks Media Sosial dengan Word2Vec dan SMOTE,” *INTEK: Jurnal Informatika dan Komputer*, vol. 10, no. 1, pp. 143–155, Apr. 2025, doi: 10.24252/instek.v10i1.54889.
- [9] E. Hasibuan and E. A. Heriyanto “Analisis Sentimen pada Ulasan Aplikasi Amazon Shopping di Google Play Store Menggunakan Naive Bayes Classifier,” *Jurnal Teknologi dan Sistem (JTS)*, vol. 1, no. 3, pp. 13–24, Oct. 2022, doi: 10.56127/jts.v1i3.434.
- [10] H. Aprilianti, H. Mustofa, K. Umam, and M. R. Handayani, “Comparative Study of SVM, KNN, and Naïve Bayes for Sentiment Analysis of Religious Application Reviews” *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 3, pp. 920–927, Jun. 2025, doi: 10.30871/jaic.v9i3.9482
- [11] P. A. Prastyo, Berlilana, and I. Tahyudin, “Sentiment Analysis on Slang Enriched Texts Using Machine Learning Approaches,” *Journal of Applied Data Sciences*, vol. 6, no. 2, pp. 1076–1087, May 2025, doi: 10.47738/jads.v6i2.626.
- [12] P. Assiroj, A. Kurnia, and S. Alam, “The performance of Naïve Bayes, support vector machine, and logistic regression on Indonesia immigration sentiment analysis,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3843–3852, Dec. 2023, doi: 10.11591/eei.v12i6.5688.
- [13] S. Riyanto, I. S. Sitanggang, T. Djabatna, and T. D. Atikah, “Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification.” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 6, pp. 1081–1090, Jun. 2023, doi: 10.14569/ijacsa.2023.01406116.
- [14] Z. Xiao, L. Wang, and J. Y. Du, “Improving the performance of sentiment classification on imbalanced datasets with transfer learning,” *IEEE Access*, vol. 7, pp. 28281–28290, Jan. 2019, doi: 10.1109/ACCESS.2019.2892094.
- [15] E. Wibowo and I. Pratama, “Analisis Sentimen Terhadap Ulasan Hotel Melalui Platform Google Review Menggunakan Metode Stacking,” *Jurnal Teknologi Dan Sistem Informasi Bisnis (JTEKSIS)*, vol. 6, no. 4, pp. 774–784, Oct. 2024, doi: 10.47233/jteksis.v6i4.1475.
- [16] N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, “The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [17] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 406–414, 2021, doi: 10.30865/mib.v5i2.2835.
- [18] N. S. Ramadan, D. Darwis, “Perbandingan Metode Naïve Bayes dan SVM untuk Sentimen Analisis Masyarakat terhadap Serangan Ransomware pada Data KIP-K,” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, pp. 12–23, 2025, doi: 10.47080/simika.v8i1.3621.
- [19] M. F. As Shidiq and D. Alita, “Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data Dari Media Sosial X Pendekatan Naive Bayes Dan Svm,” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, pp. 24–35, 2025, doi: 10.47080/simika.v8i1.3624.



- [20] N. A. Ramdhania, A. C. Siregar, and Sucipto, “Analisis sentimen pengguna X terhadap Istana Garuda IKN menggunakan algoritma Naïve Bayes,” *Jurnal Pendidikan Informatika dan Sains*, vol. 14, no. 1, pp. 36–53, Jun. 2025, doi: 10.31571/saintek.v14i1.8735.
- [21] D. N. Agustia dan R. R. Suryono, “Comparison of Naïve Bayes, Random Forest, and Logistic Regression Algorithms for Sentiment Analysis Online Gambling,” *Jurnal Inovtek Polbeng – Seri Informatika*, vol. 10, no. 1, pp. 284–295, Mar. 2025, doi: 10.35314/prk93630.
- [22] Sudriyanto, F. Syahro, and N. Fitriani, “Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, dan K-Nearest Neighbors dalam Prediksi Harga Saham,” *Jar’s: Jurnal Advance Research – Informatika dan Sistem Informasi*, vol. 2, no. 1, pp. 13–21, Dec. 2023, doi: 10.24929/jars.v2i1.2983.
- [23] F. Fazrin, O. N. Pratiwi, and R. Andreswari, “Perbandingan Algoritma K-Nearest Neighbor dan Logistic Regression pada Analisis Sentimen terhadap Vaksinasi Covid-19 pada Media Sosial Twitter dengan Pelabelan Vader dan Textblob,” *Journal e-Proceeding of Engineering*, vol. 10, no. 2, pp. 1596–1604, Apr. 2023, doi: 10.1109/icsintesa56431.2022.10041609.
- [24] D. Al Mahkya, K. A. Notodiputro, and B. Sartono, “Extra Trees Method for Stock Price Forecasting with Rolling Origin Accuracy Evaluation,” *MEDIA STATISTIKA*, vol. 15, no. 1, pp. 36–47, Jul. 2022, doi: 10.14710/medstat.15.1.36-47.