

Perbandingan Metode Naïve Bayes Dengan SVM Pada Analisis Sentimen Aplikasi Pemesanan Tiket Kapal Ferizy

Muhammad Sulhan, Erizal*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹muhammadsulhannst@gmail.com, ^{2,*}erizal@uhamka.ac.id

Email Penulis Korespondensi: erizal@uhamka.ac.id

Submitted: 13/01/2025; Accepted: 26/02/2025; Published: 01/03/2025

Abstrak—Dalam era digital, ulasan pengguna pada platform aplikasi memainkan peran penting dalam mengevaluasi kualitas layanan dan kepuasan pelanggan. Penelitian ini bertujuan membandingkan dua metode analisis sentimen, yaitu Naive Bayes dan Support Vector Machine (SVM), dalam mengklasifikasikan sentimen ulasan aplikasi Ferizy di PlayStore ke dalam kategori positif, negatif, dan netral. Naive Bayes, dengan keunggulan sederhana, efisien pada dataset kecil, dan cepat dalam pelatihan, dibandingkan dengan SVM, yang dikenal memiliki performa tinggi pada data kompleks dengan distribusi non-linear serta fleksibilitas dalam penggunaan kernel. Penelitian ini juga mengevaluasi performa kedua metode berdasarkan metrik akurasi, presisi, *recall*, dan F1-score, khususnya dalam menghadapi ketidakseimbangan kelas dan noise dalam data. Dataset yang digunakan terdiri dari ulasan pengguna aplikasi Ferizy, yang dianalisis untuk mengidentifikasi pola dan tren sentimen. Hasil implementasi menunjukkan bahwa Naive Bayes memiliki nilai akurasi sebesar 79,27%, sedangkan SVM mencapai nilai akurasi sebesar 82,62%. Perbedaan ini mengindikasikan bahwa SVM lebih unggul dalam menangani pola yang lebih kompleks pada data ulasan, meskipun selisihnya relatif kecil. Hasil penelitian ini juga menunjukkan adanya perbedaan signifikan antara kedua metode, terutama dalam keakuratan klasifikasi sentimen. Faktor-faktor seperti kompleksitas bahasa, distribusi kelas yang tidak seimbang, dan pemilihan parameter algoritma ditemukan memengaruhi performa metode. Penelitian ini memberikan wawasan yang berharga bagi pengembang aplikasi untuk meningkatkan kualitas layanan berdasarkan analisis sentimen pengguna. Selain itu, hasilnya diharapkan dapat berkontribusi pada pengembangan strategi analisis sentimen yang lebih canggih dan tepat sasaran, khususnya dalam domain transportasi digital.

Kata Kunci: Analisis Sentimen; Naïve Bayes; Support Vector Machine; Ferizy; Ulasan

Abstract—In the digital era, user reviews on application platforms play a crucial role in evaluating service quality and customer satisfaction. This study aims to compare two sentiment analysis methods, namely Naive Bayes and Support Vector Machine (SVM), in classifying the sentiment of Ferizy app reviews on PlayStore into positive, negative, and neutral categories. Naive Bayes, known for its simplicity, efficiency on small datasets, and fast training, is compared to SVM, which is recognized for its high performance on complex data with non-linear distributions and its flexibility in kernel usage. This study also evaluates the performance of both methods based on accuracy, precision, recall, and F1-score metrics, particularly in handling class imbalance and noise in the data. The dataset consists of user reviews of the Ferizy application, which are analyzed to identify sentiment patterns and trends. The implementation results show that Naive Bayes achieves an accuracy of 79.27%, while SVM reaches an accuracy of 82.62%. This difference indicates that SVM is superior in handling more complex patterns in review data, although the margin is relatively small. The findings also reveal significant differences between the two methods, particularly in sentiment classification accuracy. Factors such as language complexity, class imbalance, and algorithm parameter selection are found to influence the performance of each method. This study provides valuable insights for application developers to improve service quality based on user sentiment analysis. Additionally, the results are expected to contribute to the development of more advanced and targeted sentiment analysis strategies, particularly in the digital transportation domain. **Keyword:** Analisis Sentimen; Naïve Bayes; Support Vector Machine; Ferizy; Ulasan

Keyword: Sentiment Analysis; Naïve Bayes; Support Vector Machine; Ferizy; Review

1. PENDAHULUAN

Dalam era digital yang semakin berkembang, aplikasi pemesanan tiket kapal telah menjadi salah satu solusi utama bagi para pelancong yang ingin merencanakan perjalanan laut mereka dengan mudah dan efisien. Ferizy, sebagai salah satu aplikasi pemesanan tiket kapal yang tersedia di platform PlayStore, memberikan layanan yang penting bagi pengguna untuk merencanakan dan mengatur perjalanan mereka. Ulasan pengguna di platform PlayStore memberikan pandangan langsung dari pengguna tentang pengalaman mereka dalam menggunakan aplikasi, termasuk kepuasan mereka, masalah yang dihadapi, dan saran untuk perbaikan [1].

Analisis sentimen ulasan pengguna menjadi penting untuk memahami persepsi dan tanggapan pengguna terhadap aplikasi Ferizy. Dalam konteks ini, dua metode yang umum digunakan untuk analisis sentimen, yaitu Naive Bayes dan Support Vector Machine (SVM), menjadi fokus utama penelitian ini. Kedua metode ini menawarkan pendekatan yang berbeda dalam mengklasifikasikan sentimen dari ulasan teks, dan perbandingan antara keduanya akan memberikan wawasan yang berharga bagi pengembang aplikasi untuk memperbaiki kualitas layanan mereka.

Naive Bayes adalah metode klasifikasi yang berbasis pada teorema Bayes, yang mengasumsikan independensi antara fitur-fitur dalam data. Support Vector Machine atau SVM adalah metode klasifikasi yang bertujuan untuk menemukan hyperplane terbaik yang memisahkan dua kelas sentimen dalam ruang fitur yang diberikan [2].

Penelitian sebelumnya yang dilakukan oleh Mujaddid Izzul Fikri, dll, perbandingan analisis sentimen dilakukan menggunakan metode yang sama yaitu metode naïve bayes dan SVM. Namun hasil dari perbandingan tersebut menunjukkan bahwa metode naïve bayes lebih unggul daripada SVM[3]. Hal tersebut telah menunjukkan

bahwa performa Naive Bayes dan SVM dalam analisis sentimen dapat bervariasi tergantung pada berbagai faktor, termasuk sifat data, ukuran dataset, dan kompleksitas bahasa yang digunakan dalam teks ulasan. Oleh karena itu, penelitian ini bertujuan untuk membandingkan efektivitas dan keakuratan kedua metode ini dalam menganalisis sentimen ulasan aplikasi Ferizy di PlayStore.

Metode Naïve Bayes memiliki kelebihan, yaitu: Sederhana dan cepat [4], sehingga cocok untuk dataset yang besar; Efisien pada data kecil sehingga sering memberikan performa yang baik meskipun pada dataset yang relatif kecil; Scalabilitas, dimana dapat diimplementasikan dengan mudah dan di-skalakan untuk dataset yang sangat besar; Tidak membutuhkan banyak data pelatihan sehingga tidak memerlukan banyak data pelatihan untuk menghasilkan model yang baik.

Selain kelebihan, metode naïve bayes juga memiliki kekurangan, yaitu: Asumsi Independen, asumsi independensi antar fitur seringkali tidak realistis dalam data dunia nyata, yang dapat mempengaruhi akurasi; Keterbatasan dalam Klasifikasi Multimodal, sehingga sulit mengklasifikasikan data dengan distribusi multimodal dengan baik; Over-simplicity, kesederhanaannya bisa menjadi kelemahan ketika dihadapkan dengan data yang kompleks [5].

Algoritma Support vector machine juga memiliki kelebihan, yaitu: Efektivitas dalam ruang fitur tinggi, sehingga cocok untuk analisis teks; Margin Maksimal, mencari hyperplane dengan margin maksimal antara kelas, yang seringkali meningkatkan generalisasi model; Flexibilitas dengan Kernel, dapat menggunakan kernel yang berbeda untuk menangani data yang tidak dapat dipisahkan secara linear, baik untuk Data Kompleks, karena dapat menangani data yang kompleks dan data dengan distribusi non-linear menggunakan kernel trick.

Sama seperti metode Naïve Bayes, metode Support Vector Machine juga memiliki kekurangan, yaitu: Kompleksitas komputasi, memiliki kompleksitas komputasi yang lebih tinggi dibandingkan naïve bayes, terutama pada dataset yang sangat besar; Pemilihan hyperparameter, memerlukan pemilihan hyperparameter yang tepat, yang dapat menjadi proses yang rumit dan memakan waktu; Overfitting pada data noise, rentan terhadap overfitting jika terdapat banyak noise dalam data; Waktu pelatihan yang lama, waktu pelatihan SVM bisa sangat lama untuk dataset yang besar atau data dengan banyak fitur.

Dengan memahami kelebihan dan kekurangan masing-masing metode, diharapkan penelitian ini dapat memberikan wawasan yang berharga bagi pengembang aplikasi untuk meningkatkan kualitas layanan mereka, meningkatkan kepuasan pengguna, dan memperkuat citra merek aplikasi Ferizy di pasar yang semakin kompetitif [6].

Penelitian ini berfokus pada perbandingan antara metode Naive Bayes dan Support Vector Machine (SVM) dalam analisis sentimen ulasan pengguna aplikasi Ferizy di PlayStore. Rumusan masalah mencakup perbandingan keakuratan kedua metode, perbedaan signifikan dalam klasifikasi sentimen, faktor yang mempengaruhi kinerja masing-masing metode, serta efektivitas kedua metode dalam mengklasifikasikan sentimen positif, negatif, dan netral.

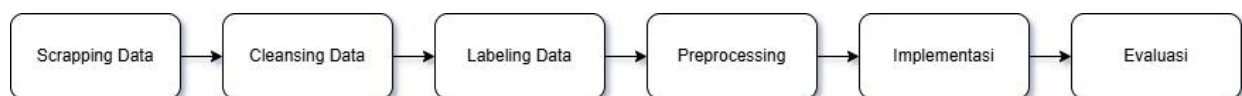
Penelitian ini juga hanya membandingkan Naive Bayes dan SVM serta membatasi klasifikasi sentimen menjadi tiga kategori (positif, negatif, netral). Data yang digunakan terbatas pada ulasan aplikasi Ferizy di PlayStore.

Tujuan utama penelitian ini adalah untuk membandingkan efektivitas Naive Bayes dan SVM dalam analisis sentimen ulasan aplikasi Ferizy. Penelitian ini juga bertujuan mengevaluasi performa kedua metode dalam menghadapi noise dan ketidakseimbangan kelas dalam dataset, serta mengidentifikasi faktor-faktor yang mempengaruhi kinerja kedua metode.

Manfaat dari penelitian ini mencakup memberikan wawasan kepada pengembang aplikasi Ferizy untuk meningkatkan layanan berdasarkan analisis sentimen, memberikan pemahaman lebih mendalam tentang kinerja Naive Bayes dan SVM dalam konteks ini, dan membantu pengembang lebih responsif terhadap kebutuhan pengguna untuk meningkatkan kepuasan. Selain itu, hasil penelitian ini dapat menjadi kontribusi penting bagi penelitian lanjutan dalam bidang analisis sentimen, terutama dalam aplikasi pemesanan tiket kapal dan bidang transportasi lainnya.

2. METODOLOGI PENELITIAN

Berikut ini alur yang dapat menjelaskan untuk visualisasi yang jelas tentang hubungan antara berbagai elemen seperti Gambar 1 :



Gambar 1. Alur Penelitian

2.1 Scrapping Data

Scraping data adalah proses ekstraksi data dari situs web secara otomatis. Teknik ini sering digunakan untuk mengumpulkan data yang tidak tersedia dalam format yang mudah diakses atau diunduh [7]. Web scraping merujuk pada teknik otomatis untuk mengambil data atau informasi dari situs web. Tujuan dari web scraping adalah mengekstraksi informasi dari berbagai situs web yang bersifat tidak terstruktur, kemudian mentransformasikannya menjadi format yang lebih terstruktur dan teratur.

2.2 Cleansing Data

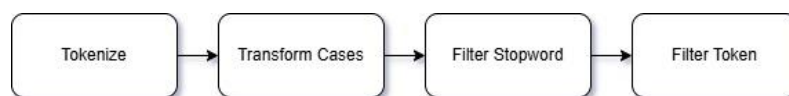
Cleaning data adalah proses mengidentifikasi dan mengoreksi (atau menghapus) data yang korup, tidak akurat, atau tidak relevan dari dataset, untuk menghilangkan noise atau symbol [7]. Langkah ini penting untuk memastikan kualitas dan keakuratan analisis data. Proses cleaning meliputi penghapusan duplikat, penanganan nilai yang hilang, dan koreksi kesalahan ejaan atau format.

2.3 Labeling Data

Labeling adalah proses memberikan label atau kategori pada data. Dalam konteks analisis sentimen, labeling berarti menentukan apakah sebuah teks (misalnya, ulasan atau tweet) memiliki sentimen positif, negatif, atau netral. Tujuan dari tahap ini adalah untuk melatih model yang akan digunakan pada tahap pelatihan data dan mempermudah pengidentifikasian sentiment dalam dataset [8]. Proses ini dapat dilakukan secara manual oleh manusia atau otomatis oleh algoritma pembelajaran mesin yang telah dilatih.

2.4 Preprocessing

Preprocessing adalah tahap awal dalam pemrosesan data yang bertujuan untuk mengubah data mentah menjadi format yang lebih sesuai untuk analisis atau pemodelan [8]. Langkah ini ditempuh dengan maksud supaya dataset bisa diolah secara efektif selama proses klasifikasi, dengan mempersiapkan teks yang akan diolah selanjutnya melalui transformasi data dari bentuk yang belum atau tidak teratur menjadi data yang teratur. [5]. Proses ini penting karena data yang diperoleh dari berbagai sumber sering kali tidak bersih, tidak lengkap, atau tidak sesuai dengan format yang dibutuhkan untuk analisis lebih lanjut. Gambar 2 merupakan langkah dalam preprocessing yang dilakukan dalam penelitian ini adalah:



Gambar 2. Alur Preprocessing

2.4.1 Tokenize

Tokenization adalah proses memecah teks menjadi unit-unit yang lebih kecil [9] biasanya kata-kata atau frasa, yang disebut token. Ini adalah langkah awal dalam banyak tugas pemrosesan bahasa alami (NLP) [10] yang memungkinkan analisis lebih lanjut seperti pencocokan kata, analisis frekuensi, atau pembuatan model teks.

2.4.2 Transform Cases

Transform cases atau case folding adalah proses mengonversi [11] semua karakter dalam teks menjadi huruf kecil. Langkah ini dilakukan untuk memastikan bahwa perbedaan antara huruf besar dan kecil tidak memengaruhi analisis teks, seperti saat menghitung frekuensi kata atau melakukan pencocokan kata kunci [5].

2.4.3 Stopword

Stopwords adalah suatu prosedur yang dilakukan untuk menghilangkan setiap kata penghubung dan istilah yang tidak perlu dari kumpulan data. Penggunaan operator filter stopwords pada alat RapidMiner memungkinkan proses eliminasi kata-kata tidak relevan berdasarkan kamus Bahasa Indonesia yang diunduh dari situs Kaggle [5].

2.4.4 Filter Token

Filter token berdasarkan panjangnya suatu huruf. Tujuan dari proses Filter token by length adalah untuk membatasi jumlah karakter dalam kalimat data yang sedang diproses dengan cara menyaring token berdasarkan panjangnya [5].

2.5 Implementasi

Implementasi adalah proses penerapan model analisis sentimen menggunakan metode Naïve Bayes dan Support Vector Machine (SVM) pada data yang telah diproses.

2.5.1 Naïve Bayes

Naive Bayes adalah metode klasifikasi yang didasarkan pada Teorema Bayes dengan asumsi independensi [2] antar fitur. Ini berarti bahwa dalam konteks klasifikasi teks, setiap kata dianggap independen dari kata lainnya. Metode ini sangat populer untuk analisis sentimen karena kesederhanaannya dan efektivitasnya dalam banyak kasus. Ada beberapa varian dari algoritma Naive Bayes, seperti Gaussian Naive Bayes, Multinomial Naive Bayes, dan Bernoulli Naive Bayes, yang masing-masing sesuai untuk jenis data yang berbeda. Rumus untuk menghitung *confusion matrix* adalah sebagai berikut [12]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

2.5.2 Support Vector Machine (SVM)

Support Vector Machine adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan mencari hyperplane optimal yang memisahkan kelas-kelas dalam ruang fitur dengan margin maksimum. SVM sangat efektif [13] dalam ruang berdimensi tinggi dan sangat berguna dalam kasus di mana jumlah fitur lebih besar daripada jumlah sampel. SVM dapat menggunakan kernel [14] yang berbeda (seperti linear, polynomial, dan RBF) untuk menangani data yang tidak dapat dipisahkan secara linear [15]. Rumus untuk menghitung *confusion matrix* adalah sebagai berikut [12]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

2.2 Evaluasi

Evaluasi adalah proses mengukur kinerja model yang telah dibangun. Dalam penelitian ini, evaluasi dilakukan untuk membandingkan kinerja metode Naïve Bayes dan SVM dalam menganalisis sentimen. Menggunakan *confusion matrix* untuk menilai kinerja metode klasifikasi sebagai bagian dari evaluasi (4) Rumus untuk menghitung *confusion matrix* adalah sebagai berikut [12]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (7)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (8)$$

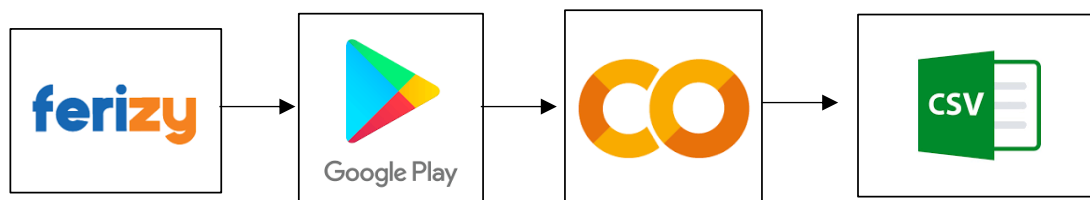
$$Recall = \frac{TP}{(TP + FN)} \quad (9)$$

Penjelasan dari rumus tersebut mencakup True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN). TP menunjukkan jumlah data yang benar-benar positif dan diprediksi dengan benar sebagai positif, sedangkan FP adalah jumlah data yang sebenarnya negatif tetapi salah diprediksi sebagai positif. Sebaliknya, FN mengacu pada jumlah data yang sebenarnya positif tetapi diprediksi dengan salah sebagai negatif, sedangkan TN menunjukkan jumlah data yang benar-benar negatif dan diprediksi dengan benar sebagai negatif [16].

3. HASIL DAN PEMBAHASAN

3.1 Pengambilan Data

Langkah awal dari analisis sentimen ialah pengumpulan data, teknik pengumpulan data yang digunakan yaitu Scrapping data. Scrapping data adalah proses pengumpulan data dari situs web dengan menggunakan google colab, data yang akan diambil merupakan ulasan dari aplikasi ferizy yang berada pada google playstore. Kemudian akan disimpan dalam format .csv, Gambar 3 menunjukkan alur dari pengambilan data



Gambar 3. Proses Scrapping Data

Dataset yang akan diambil sebanyak 1100 data, dengan ketentuan data terbaru yang akan diambil, Gambar 4 menunjukkan source code jumlah serta penyesuaian yang diinginkan.

```

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.nutech_integrasi.asdpmobile',
    lang='id',
    country='id',
    sort=Sort.MOST_NEWEST, # defaults to Sort.MOST_RELEVANT you can use Sort.NEWEST to get newst reviews
    count=1100,
    filter_score_with=None
)

```

Gambar 4. Source Code Scrapping Data

Selanjutnya data yang sudah diambil akan disimpan dan diubah ke format .csv dengan jumlah data yang telah disesuaikan sebelumnya, Gambar 5 menunjukkan source code untuk menyimpan data.

```

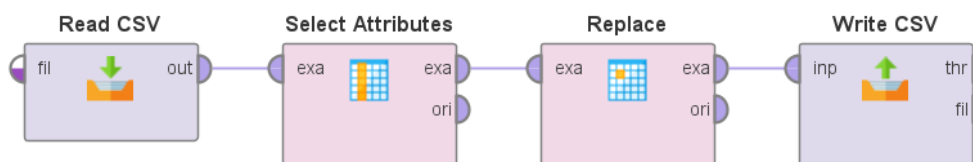
[ ] my_df.to_csv("scrapped_data_ferizy.csv", index = False)

```

Gambar 5. Source Code Menyimpan Data

3.2 Cleansing Data

Setelah data terkumpul, langkah selanjutnya adalah proses cleansing data, yang bertujuan untuk membersihkan data mentah dari berbagai elemen yang tidak relevan atau dapat mengganggu analisis. Proses ini dilakukan dengan langkah-langkah pada gambar 6:

**Gambar 6.** Cleansing Data

Select Attributes untuk memilih kolom 'content' yang berisi ulasan dari aplikasi ferizy yang akan digunakan. Replace untuk menghapus punctuation character: [-!"#\$%&'()*+,-./:;<=>?@\[\]_`{|}~].

3.3 Labeling

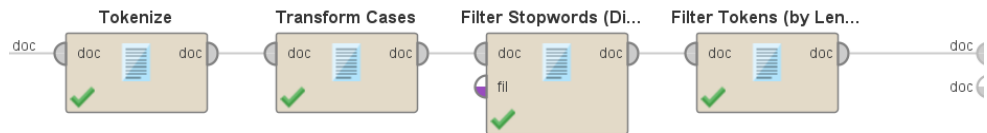
Tahap selanjutnya setelah cleansing data mentah, data yang sudah disimpan akan dilabeli dengan sentimen positif atau negatif, dengan menggunakan aplikasi excel atau dengan menggunakan google sheet. Tahap labeling ini dilakukan dengan melabeli data atau ulasan pada aplikasi ferizy yang sebelumnya telah di cleansing, dengan menganalisis data atau ulasan tersebut dan kemudian menambahkan kolom sentiment. Tabel 1 menunjukkan hasil labeling data:

Tabel 1. Hasil Label

CONTENT	SENTIMEN
Sangat baik	Positif
Aplikasi paling ga jelas,	Negatif
Tidak bisa akses lokasi asal pelabuhan	Negatif
Tidak jelas	Negatif
Aplikasi tidak berguna. Gak jelas	Negatif
Melatih rakyat untuk tidak nyuap	Positif

3.4 Preprocessing

Preprocessing merupakan tahap penting dalam analisis sentimen yang bertujuan untuk mempersiapkan data teks agar dapat dianalisis lebih lanjut menggunakan algoritma pembelajaran mesin. Proses ini melibatkan serangkaian langkah untuk mengubah data teks mentah menjadi format yang lebih terstruktur dan berguna, menghilangkan elemen yang tidak relevan, serta memperbaiki kualitas data agar hasil analisis sentimen menjadi lebih akurat. Gambar 7 menunjukkan isi dari proses preprocessing:



Gambar 7. Preprocessing

Langkah pertama dalam preprocessing adalah tokenisasi, yaitu memecah teks menjadi unit-unit terkecil seperti kata atau frasa yang disebut token. Tokenisasi membantu dalam memisahkan kata-kata penting dari struktur kalimat yang lebih besar. Tabel 2 menunjukkan hasil dari tokenize.

Tabel 2. Hasil Tokenize

Sebelum	Sesudah
aplikasi sdah bagus tapi belum lengkap hal penting yg tercantu di E-tiket nya, mohon di cantumkan harga dalam E-tiket secara terperinci juga. agar mempermudah bagi orang yg melakukan perjalanan dinas, apalagi tiket yang dicetak mudah terhapus hanya dalam tiga hari. selain itu harga tiket yang tercantum tidak sesuai dengan angka pembayaran aktualnya (biaya lain-lain seperti biaya admin selain harga tiket utama),	['aplikasi', 'sdah', 'bagus', 'tapi', 'belum', 'lengkap', 'hal', 'penting', 'yg', 'tercantu', 'di', 'E', 'tiket', 'nya', 'mohon', 'di', 'cantumkan', 'harga', 'dalam', 'E', 'tiket', 'secara', 'terperinci', 'juga', 'agar', 'mempermudah', 'bagi', 'orang', 'yg', 'melakukan', 'perjalanan', 'dinas', 'apalagi', 'tiket', 'yang', 'dicetak', 'mudah', 'terhapus', 'hanya', 'dalam', 'tiga', 'hari', 'selain', 'itu', 'harga', 'tiket', 'yang', 'tercantum', 'tidak', 'sesuai', 'dengan', 'angka', 'pembayaran', 'aktualnya', 'biaya', 'lain', 'lain', 'seperti', 'biaya', 'admin', 'selain', 'harga', 'tiket', 'utama']

Langkah kedua adalah transform cases, yaitu mengubah semua huruf menjadi huruf kecil (lowercase) untuk memastikan konsistensi data dan menghindari perbedaan akibat huruf besar atau kecil. Tabel 3 menunjukkan hasil dari transform cases.

Tabel 3. Hasil Transform Cases

Sebelum	Sesudah
['aplikasi', 'sdah', 'bagus', 'tapi', 'belum', 'lengkap', 'hal', 'penting', 'yg', 'tercantu', 'di', 'E', 'tiket', 'nya', 'mohon', 'di', 'cantumkan', 'harga', 'dalam', 'E', 'tiket', 'secara', 'terperinci', 'juga', 'agar', 'mempermudah', 'bagi', 'orang', 'yg', 'melakukan', 'perjalanan', 'dinas', 'apalagi', 'tiket', 'yang', 'dicetak', 'mudah', 'terhapus', 'hanya', 'dalam', 'tiga', 'hari', 'selain', 'itu', 'harga', 'tiket', 'yang', 'tercantum', 'tidak', 'sesuai', 'dengan', 'angka', 'pembayaran', 'aktualnya', 'biaya', 'lain', 'lain', 'seperti', 'biaya', 'admin', 'selain', 'harga', 'tiket', 'utama']	['aplikasi', 'sdah', 'bagus', 'tapi', 'belum', 'lengkap', 'hal', 'penting', 'yg', 'tercantu', 'di', 'e', 'tiket', 'nya', 'mohon', 'di', 'cantumkan', 'harga', 'dalam', 'e', 'tiket', 'secara', 'terperinci', 'juga', 'agar', 'mempermudah', 'bagi', 'orang', 'yg', 'melakukan', 'perjalanan', 'dinas', 'apalagi', 'tiket', 'yang', 'dicetak', 'mudah', 'terhapus', 'hanya', 'dalam', 'tiga', 'hari', 'selain', 'itu', 'harga', 'tiket', 'yang', 'tercantum', 'tidak', 'sesuai', 'dengan', 'angka', 'pembayaran', 'aktualnya', 'biaya', 'lain', 'lain', 'seperti', 'biaya', 'admin', 'selain', 'harga', 'tiket', 'utama']

Langkah ketiga adalah filter stopwords, yaitu kata-kata yang tidak memberikan kontribusi signifikan terhadap makna sentimen, seperti "dan", "yang", "untuk", atau "dengan". Selain itu, teks juga akan diperiksa untuk menghapus karakter khusus, angka, dan tanda baca yang tidak diperlukan. Tabel 4 menunjukkan hasil dari stopwords.

Tabel 4. Hasil Stopword

Sebelum	Sesudah
---------	---------

['aplikasi', 'sdah', 'bagus', 'tapi', 'belum', 'lengkap', 'hal', 'penting', 'yg', 'tercantu', 'di', 'e', 'tiket', 'nya', 'mohon', 'di', 'cantumkan', 'harga', 'dalam', 'e', 'tiket', 'secara', 'terperinci', 'juga', 'agar', 'mempermudah', 'bagi', 'orang', 'yg', 'melakukan', 'perjalanan', 'dinas', 'apalagi', 'tiket', 'yang', 'dicetak', 'mudah', 'terhapus', 'hanya', 'dalam', 'tiga', 'hari', 'selain', 'itu', 'harga', 'tiket', 'yang', 'tercantum', 'tidak', 'sesuai', 'dengan', 'angka', 'pembayaran', 'aktualnya', 'biaya', 'lain', 'lain', 'seperti', 'biaya', 'admin', 'selain', 'harga', 'tiket', 'utama']

['aplikasi', 'sdah', 'bagus', 'lengkap', 'yg', 'tercantu', 'e', 'tiket', 'nya', 'mohon', 'cantumkan', 'harga', 'e', 'tiket', 'terperinci', 'mempermudah', 'orang', 'yg', 'perjalanan', 'dinas', 'tiket', 'dicetak', 'mudah', 'terhapus', 'harga', 'tiket', 'tercantum', 'sesuai', 'angka', 'pembayaran', 'aktualnya', 'biaya', 'biaya', 'admin', 'harga', 'tiket', 'utama']

Langkah keempat adalah filter token by length, yaitu penyaringan kata berdasarkan Panjang kata guna untuk meningkatkan kualitas data, token yang terlalu pendek (misalnya, hanya terdiri dari satu huruf) dihapus karena biasanya tidak memberikan informasi yang berarti. Tabel 5 menunjukkan hasil dari filter token.

Tabel 5. Hasil Filter Token

Sebelum	Sesudah
['aplikasi', 'sdah', 'bagus', 'lengkap', 'yg', 'tercantu', 'e', 'tiket', 'nya', 'mohon', 'cantumkan', 'harga', 'e', 'tiket', 'terperinci', 'mempermudah', 'orang', 'yg', 'perjalanan', 'dinas', 'tiket', 'dicetak', 'mudah', 'terhapus', 'harga', 'tiket', 'tercantum', 'sesuai', 'angka', 'pembayaran', 'aktualnya', 'biaya', 'lain', 'lain', 'seperti', 'biaya', 'admin', 'selain', 'harga', 'tiket', 'utama']	['aplikasi', 'sdah', 'bagus', 'lengkap', 'tercantu', 'tiket', 'nya', 'mohon', 'cantumkan', 'harga', 'tiket', 'terperinci', 'mempermudah', 'orang', 'perjalanan', 'dinas', 'tiket', 'dicetak', 'mudah', 'terhapus', 'harga', 'tiket', 'tercantum', 'sesuai', 'angka', 'pembayaran', 'aktualnya', 'biaya', 'biaya', 'admin', 'harga', 'tiket', 'utama']

Proses preprocessing ini sangat penting karena membantu mengurangi kompleksitas data, meningkatkan efisiensi algoritma, dan memastikan bahwa model analisis sentimen hanya memproses informasi yang relevan untuk menentukan sentimen dari teks. Hasil akhirnya adalah data yang siap digunakan dalam implementasi model pembelajaran mesin seperti Naive Bayes dan SVM.

3.5 Implementasi

Setelah melewati proses preprocessing, data kemudian akan diimplementasikan dengan algoritma naïve bayes dan svm. Proses ini dibagi menjadi dua, pertama implementasi dengan metode naïve bayes dan yang kedua implementasi dengan metode svm. [17]

Naïve Bayes

Hasil dari implementasi analisis dengan menggunakan metode naïve bayes diatas menghasilkan nilai akurasi 79,27%. Confusion matrix menunjukkan bahwa terdapat 33 True Positive (TP), 227 True Negative (TN), 24 False Positive (FP), 44 False Negative (FN), gambar 8 menunjukkan hasil performa dari metode naïve bayes.

accuracy: 79.27%

	true Positif	true Negatif	class precision
pred. Positif	33	44	42.86%
pred. Negatif	24	227	90.44%
class recall	57.89%	83.76%	

Gambar 8. Nilai Akurasi Naïve Bayes

Untuk menghitung *accuracy*, *precision*, dan *recall* pada metode naïve bayes, dilakukan menggunakan rumus [18][19]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(33 + 227)}{(33 + 227 + 24 + 44)} = 0.79268 = 79.27\%$$

$$Precision = \frac{TP}{(TP + FP)} = \frac{33}{(33 + 44)} = 0.42857 = 42.86\%$$

$$Recall = \frac{TP}{(TP + FN)} = \frac{33}{(33 + 24)} = 0.57894 = 57.89\%$$

Support Vector Machine

Hasil dari implementasi analisis dengan menggunakan metode naïve bayes diatas menghasilkan nilai akurasi 82,62%. Confusion matrix menunjukkan bahwa terdapat 2 True Positive (TP), 269 True Negative (TN), 55 False Positive (FP), 2 False Negative (FN), gambar 9 menunjukkan hasil performa dari metode svm.

accuracy: 82.62%

	true Positif	true Negatif	class precision
pred. Positif	2	2	50.00%
pred. Negatif	55	269	83.02%
class recall	3.51%	99.26%	

Gambar 9. Nilai Akurasi SVM

Untuk menghitung *accuracy*, *precision*, dan *recall* pada metode naïve bayes, dilakukan menggunakan rumus [18][19]:

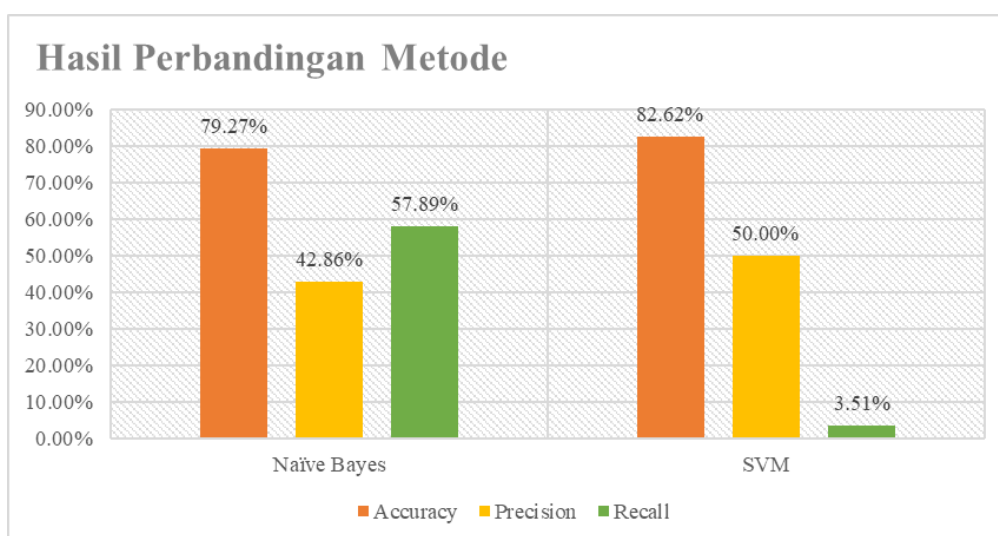
$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(2 + 269)}{(2 + 269 + 2 + 55)} = 0.8262 = 82.62\%$$

$$Precision = \frac{TP}{(TP + FP)} = \frac{2}{(2 + 55)} = 0.0351 = 3.51\%$$

$$Recall = \frac{TP}{(TP + FN)} = \frac{2}{(2 + 2)} = 0.5 = 50.00\%$$

3.6 Pembahasan

Evaluasi hasil perbandingan analisis sentimen antara metode Naive Bayes dan SVM menunjukkan bahwa SVM memiliki kinerja yang lebih baik daripada Naive Bayes dalam klasifikasi sentimen. Dengan nilai akurasi 82,26%, SVM mengungguli Naive Bayes dengan akurasi 79,27%. Perbedaan akurasi ini menunjukkan bahwa SVM lebih efektif [13] dalam menangani pola kompleks dalam data sentimen, terutama saat menghadapi kebisingan atau ketidakseimbangan kelas dalam kumpulan data. Akan tetapi, Naive Bayes masih mempunyai kelebihan dalam hal efisiensi komputasi, sehingga menjadikannya pilihan yang layak untuk kasus-kasus dengan keterbatasan sumber daya. Hasil ini menekankan pentingnya mempertimbangkan karakteristik kumpulan data dan kebutuhan analisis sebelum memilih metode yang paling cocok untuk analisis sentimen. Gambar 10 menunjukkan hasil perbandingan antara metode naïve bayes dan metode svm.



Gambar 10. Perbandingan Hasil Metode

Visualisasi kata ditampilkan menggunakan *Wordcloud*, sebuah representasi visual dari teks di mana kata-kata yang sering muncul memiliki ukuran lebih besar, sementara kata-kata yang jarang muncul memiliki ukuran lebih kecil [20]. Kata 'aplikasi' adalah kata yang paling banyak digunakan dengan jumlah kemunculan sebanyak 626 kali. Kemudian diurutkan kedua kata yang paling banyak digunakan dengan jumlah kemunculan sebanyak 363 kali adalah kata 'tiket', dan diurutkan ketiga kata yang paling banyak digunakan dengan jumlah kemunculan sebanyak 213 kali adalah kata 'pelabuhan'. Gambar 12 menunjukkan representasi visual dalam bentuk wordcloud.



- [15] Friska Aedita Indriyani, Ahmad Fauzi, dan Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, hlm. 176–184, Jul 2023, doi: 10.37373/tekno.v10i2.419.
- [16] A. Komarudin, A. Meutia Hilda, dan C. Author, “Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Pada Play Store Menggunakan Metode Naïve Bayes,” *Computer Science (CO-SCIENCE)*, vol. 4, no. 1, 2024, doi: 10.31294/coscience.v4i1.2955.
- [17] L. B. Ilmawan dan M. A. Mude, “Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store,” *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, hlm. 154–161, Agu 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [18] E. Apriani, F. Oktavianalisti, L. D. H. Monasari, I. Winarni, dan I. F. Hanif, “Analisis Sentimen Penggunaan TikTok Sebagai Media Pembelajaran Menggunakan Algoritma Naïve Bayes Classifier,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, hlm. 1160–1168, Jul 2024, doi: 10.57152/malcom.v4i3.1482.
- [19] N. Wijaya dan E. Setiawan Panjaitan, “Analisis Sentimen Ulasan Aplikasi Instagram di Google Play Store: Pendekatan Multinomial Naive Bayes dan Berbasis Leksikon,” *Technology and Science (BITS)*, vol. 6, no. 2, 2024, doi: 10.47065/bits.v6i2.5615.
- [20] I. Oktavia dan A. R. Isnain, “Analisis Sentimen Opini Terhadap Tools Artificial Intelligence (AI) Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, hlm. 777, Apr 2024, doi: 10.30865/mib.v8i2.7524.