



# ANALISIS SENTIMEN ULASAN PENGUNJUNG MENGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE PADA HOTEL NEO PONTIANAK

Stevin Tandra<sup>1</sup>, Genrawan Hoendarto<sup>2</sup>, Alvin Lesmana<sup>3</sup>.

<sup>1,3</sup>Bisnis Digital Fakultas Teknologi Informasi, Universitas Widya Dharma Pontianak,

<sup>2</sup>Informatika Fakultas Teknologi Inforamasi, Universitas Widya Dharma Pontianak, Kalimantan Barat, Indonesia

Correspondence: E-mail: [22430186@widyadharm.ac.id](mailto:22430186@widyadharm.ac.id), [genrawan@widyadharm.ac.id](mailto:genrawan@widyadharm.ac.id),  
[alvin\\_lesmana@widyadharm.ac.id](mailto:alvin_lesmana@widyadharm.ac.id).

## ABSTRACT

Perkembangan sektor pariwisata seperti hotel dipengaruhi oleh kepuasan dan kenyamanan pelanggan, dengan adanya teknologi yang canggih pada saat ini dapat membantu manajemen hotel untuk meningkatkan kepuasan dan kenyamanan pelanggan melalui ulasan yang diberikan oleh pelanggan pada hotel melalui Google Maps. Penelitian ini mengambil 1387 data ulasan, dengan menggunakan subset data pelatihan (80%) dan pengujian (20%). Algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) digunakan sebagai perbandingan untuk menentukan algoritma yang memiliki akurasi lebih tinggi. Algoritma SVM memiliki tingkat akurasi sebesar 90,71% yang lebih tinggi dari algoritma *Naïve Bayes* yang mendapatkan akurasi sebesar 86,43%. Kemudian dilakukan algoritma *Latend Dirichlet Allocation* (LDA) untuk memodelkan topik untuk memahami kata-kata pada sentimen positif dan negatif yang diberikan oleh pelanggan. Pada sentimen positif didapatkan bahwa fasilitas kamar bersih, pelayanan yang ramah, dan juga lokasi yang strategis. Sedangkan pada sentimen negatif didapatkan terkait kebersihan kamar, kendala pelayanan check-in, gangguan suara pada kamar, dan juga tempat parkir dan akses taksi online.

## ARTICLE INFO

### Article History:

Submitted/Received 7 May 2026

First Revised 8 May 2026

Accepted 12 May 2026

First Available online 12 May 2026

Publication Date 12 May 2026

### Kata Kunci:

*Analisis Sentimen, Latend Dirichlet Allocation, Naïve Baiyes, Support Vector Machine, Ulasan Pengunjung Hotel.*

## 1. PENDAHULUAN

Sektor pariwisata merupakan sektor yang perkembangannya dipengaruhi oleh kepuasan dan kenyamanan pelanggan terutama pada perhotelan. Salah satu cara untuk mengetahui kenyamanan dan kepuasan pelanggan terhadap hotel adalah ulasan online dari pelanggan yang sudah pernah menginap di hotel tersebut (Assiva, 2024). Semakin berkembangnya teknologi di era sekarang memudahkan wisatawan untuk mendapatkan informasi mengenai kualitas dan pelayanan hotel melalui internet (Susilawati & Wahyuni, 2024). Salah satu platform yang dapat digunakan sebagai referensi untuk mencari informasi mengenai kenyamanan dan kepuasan pelanggan adalah Google Maps yang menyediakan fitur ulasan yang diberikan oleh pelanggan terhadap hotel yang pernah ditempati (Azizah et al., 2024).

Seiring dengan pesatnya perkembangan teknologi menyebabkan sektor pariwisata terutama perhotelan memiliki reputasi melalui ulasan pelanggan terkait kepuasan dan kenyamanan dari pelanggan. Dengan banyaknya jumlah data ulasan yang diberikan oleh pelanggan menyebabkan analisis secara manual akan menjadi sulit dan tidak efisien (Apriliansyah et al., 2025). Solusi yang dapat ditawarkan agar analisis dapat efisien adalah dengan menggunakan pembelajaran mesin (*Machine Learning*) untuk membantu melakukan klasifikasi data secara otomatis dan cepat untuk menganalisis sentimen (Slam et al., 2025).

Analisis sentimen digunakan untuk mengklasifikasikan opini pelanggan dalam kategori positif atau negatif. Terdapat beberapa metode pembelajaran mesin yang dapat digunakan dalam analisis sentimen seperti *Naïve Bayes*, *Support Vector Machine*, *Random Forest*, dan *Deep Learning* (Laia & Barus, 2025). *Naïve Bayes* menunjukkan performa yang konsisten dalam menangani permasalahan klasifikasi teks, algoritma ini memiliki efisiensi dalam menangani dataset dengan dimensi tinggi serta kemampuannya menghasilkan prediksi yang akurat meskipun dengan asumsi independensi fitur yang sederhana (Hilmawan, 2025). Metode ini telah menunjukkan akurasi tinggi dalam menentukan teks bersifat positif atau negatif, dan telah banyak digunakan dalam berbagai studi untuk analisis sentiment (Haq et al., 2024).

Algoritma lain yang dapat digunakan untuk menganalisis sentimen ulasan pelanggan adalah *Support Vector Machine* (SVM), SVM merupakan algoritma yang sangat reliabel dan sering digunakan untuk menganalisis sentimen ulasan pelanggan (Junaedi et al., 2024). SVM terbukti sangat efektif dalam menangani data yang kompleks tanpa mengalami penurunan performa yang signifikan dibandingkan algoritma klasifikasi tradisional lainnya (Nadhifah et al., 2024).

Beberapa penelitian terdahulu menunjukkan bahwa algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) mampu menangani analisis sentimen pada data ulasan dalam jumlah besar secara efektif. Beberapa penelitian yang menggunakan metode ini yang digunakan sebagai referensi dalam penelitian ini sebagai berikut. (Alexander Romian Simarmata, 2023) Pada penelitian ini menggunakan algoritma *Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine* (SVM) untuk menentukan algoritma yang memiliki akurasi lebih tinggi diantara tiga algoritma ini. Data dikumpulkan dari web Kaggle yang berjumlah 20.491 data ulasan sebuah hotel, data diberi label dan diproses melalui tahapan *preprocessing* sebelum diuji menggunakan ketiga algoritma. Hasil penelitian yang didapatkan menunjukkan bahwa SVM mendapatkan nilai akurasi yang lebih tinggi sebesar 94%, untuk algoritma *Naïve Bayes* mendapatkan nilai akurasi sebesar 91%, dan untuk algoritma *Logistic Regression* mendapatkan nilai akurasi sebesar 93%. (Syahputra et al., 2022) Penelitian ini membandingkan dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), untuk memetakan analisis sentimen pengguna Twitter terhadap aplikasi PeduliLindungi. Data yang digunakan dalam penelitian ini adalah cuitan dari pengguna Twitter dengan total 4.782 cuitan. Akurasi yang diperoleh menggunakan metode pengujian *k-fold*

untuk *Naïve Bayes* adalah 85% dan untuk SVM adalah 86%. Sementara itu, dengan menggunakan teknik pembagian 80:20, akurasi yang didapatkan untuk *Naïve Bayes* adalah 80% dan untuk SVM adalah 79%. (Suparyati & Fathurrahman, 2022) Penelitian ini menggunakan metode klasifikasi *Naïve Bayes* untuk menganalisis ulasan hotel, data diperoleh dari ulasan TripAdvisor dari Kaggle yang berjumlah 20491 data. Data terdiri dari ulasan positif sebesar 86,74% dan ulasan negatif sebesar 13,26%, teknik pembagian yang digunakan adalah 80:20, akurasi yang dihasilkan menggunakan metode klasifikasi *Naïve Bayes* sebesar 95,6%. (Singgalen, 2023) Penelitian ini menerapkan klasifikasi *Support Vector Machine* (SVM) dalam kerangka kerja *CRISP-DM* untuk menganalisis sentimen tamu hotel, dengan memanfaatkan WebHarvy dan RapidMiner untuk pemrosesan data. Hasil penelitian ini menunjukkan bahwa SVM tanpa menggunakan operator SMOTE menghasilkan akurasi sebesar 95,82%, *precision* 95,80%, *recall* 100%, dan *Area Under the Curve* (AUC) 79,80%. Sebaliknya, performa SVM dengan menggunakan operator SMOTE menghasilkan akurasi sebesar 92,05%, *precision* 100%, *recall* 84,08%, dan AUC sebesar 99,9%. (Sudirman et al., 2025) Penelitian ini menggunakan analisis sentimen untuk mengklasifikasikan opini pelanggan pada aspek fasilitas, kamar, kebersihan, dan pelayanan dengan menggunakan Hotel XYZ di Kota Tangerang sebagai objek. Data ulasan dikumpulkan dari platform Google Maps, Tiket.com, dan Traveloka yang memperoleh 3.609 kata positif dan 6.609 kata negatif. Algoritma yang digunakan adalah *Support Vector Machine* dengan membagi data latih 80% dan data uji 20% klasifikasi dilakukan dengan pengujian parameter  $C = 1 - 5$ , Hasil yang diperoleh menunjukkan *F1-Score* pada aspek fasilitas 90%, kebersihan 83%, kamar 79%, dan pelayanan 67%, dengan parameter  $C = 3$  yang memberikan performa optimal pada sebagian besar aspek.

Berdasarkan hasil penelitian terdahulu menunjukkan bahwa algoritma *Naïve Bayes* dan SVM cocok digunakan pada penelitian ini dengan objek Hotel Neo Pontianak untuk menganalisis ulasan pelanggan dan mendukung pengambilan keputusan. Data yang digunakan pada penelitian ini hanya terbatas pada analisis teks ulasan tanpa mempertimbangkan penilaian bintang dari pengguna, dan juga teks ulasan yang diambil terbatas pada Bahasa Indonesia.

Pada penelitian ini untuk menemukan topik-topik tersembunyi dalam sekumpulan data ulasan, menggunakan pendekatan *Topic Modeling* dengan algoritma *Latent Dirichlet Allocation* (LDA) (Saputra et al., 2025). Metode ini bekerja secara otomatis untuk mengelompokkan kata-kata yang sering muncul secara bersamaan yang dapat menghasilkan topik-topik tertentu tanpa perlu ditentukan secara manual di awal (Astuti & Cahyono, 2023). Dengan menggabungkan analisis sentimen dan pemodelan topik bertujuan agar manajemen hotel dapat memahami sentimen pada aspek-aspek secara spesifik, seperti : aspek fasilitas, aspek lokasi, dan aspek pelayanan.

Fokus dalam penelitian ini bertujuan untuk membandingkan performa algoritma *Naïve Bayes* dan SVM dengan melakukan analisis sentimen ulasan pelanggan berbasis Google Maps. Perbandingan dilakukan guna menentukan algoritma mana yang paling optimal dan stabil dalam menangani data ulasan dengan jumlah data yang banyak dan tingkat ambiguitas bahasa yang tinggi. Dan juga memiliki tujuan untuk menganalisis aspek-aspek apa saja yang sering diberikan ulasan bagus dan aspek yang sering diberikan ulasan buruk yang dapat digunakan sebagai bahan pertimbangan evaluasi. Kontribusi dari penelitian ini diharapkan dapat memberikan wawasan bagi pihak manajemen hotel dalam memilih algoritma yang efektif untuk menganalisis sentimen ulasan pelanggan. Hasil penelitian ini juga diharapkan dapat memberi kontribusi bagi manajemen hotel dalam memahami aspek-aspek apa saja yang perlu ditingkatkan melalui ulasan pelanggan.

## 2. METODE

Pada bagian ini akan membahas mengenai metode yang digunakan pada penelitian ini, langkah-langkah yang dilakukan dimulai dari pengumpulan data hingga yang terakhir adalah evaluasi.

### 2.1. Pengumpulan Data

Penelitian ini mengumpulkan data ulasan pelanggan dari Google Maps melalui proses *scraping*, data ulasan yang dikumpulkan dari hotel Neo yang terletak di kota Pontianak, Kalimantan Barat, Indonesia. Tanggal dan rating tidak digunakan karena pelabelan positif dan negatif dilakukan secara manual, sehingga data yang digunakan hanya teks ulasan pengguna dengan sampel berjumlah sebanyak 1387 data.

### 2.2 Pelabelan

Pelabelan data dalam penelitian ini dilakukan pada setiap ulasan pengguna, dengan membagi dua kategori yakni positif dan negatif. Pelabelan dilakukan secara manual untuk mengelompokkan ulasan yang positif dan negatif, untuk ulasan yang bersifat netral tidak diambil ulasannya karena tidak berperan dalam tujuan penelitian ini, sehingga data yang digunakan berjumlah sebanyak 1356 data.

### 2.3 Pra-pemrosesan data

Pra-pemrosesan data merupakan tahap yang penting dalam analisis sentimen karena data teks mentah dari ulasan tidak semua bersih dan terstruktur, sehingga dapat mempengaruhi kinerja klasifikasi. Langkah-langkah proses ini meliputi *cleaning*, *case folding*, normalisasi, tokenisasi, stopword, dan stemming. *Cleaning* adalah proses penghapusan karakter non-abjad seperti tanda baca, angka, emoji, dan simbol, bertujuan agar teks yang dihasilkan hanya karakter abjad. *Case folding* adalah proses mengubah semua huruf dalam teks menjadi huruf kecil, bertujuan agar data dapat tetap konsisten. Normalisasi adalah proses untuk mengubah kata singkatan atau tidak baku menjadi bentuk dasar sesuai dengan pedoman bahasa(kamus) dan menghilangkan huruf berulang, bertujuan untuk mengurangi variasi kata dan meningkatkan akurasi dari model. Tokenisasi adalah proses memecah sekumpulan kalimat menjadi kata-kata individu, bertujuan untuk memudahkan dalam pemrosesan data. *Stopword* adalah proses menghilangkan kata penghubung, kata depan, kata ganti, dan kata bantu pada kalimat ulasan, ini bertujuan agar model dapat lebih fokus pada kata-kata yang dapat memberikan makna positif atau negatif untuk mengurangi *noise* dalam data. *Stemming* adalah proses untuk menghilangkan imbuhan, awalan, akhiran, dan sisipan, bertujuan untuk mengubah kata menjadi kata dasar untuk mengurangi kata-kata yang memiliki makna yang sama. Dengan melakukan langkah-langkah ini memastikan data yang digunakan tetap bersih, terstruktur, dan konsisten untuk meningkatkan efisiensi dalam pemrosesan data.

### 2.4. Term Frequency Inverse Document Frequency (TF-IDF)

Pada penelitian ini dilakukan tahap pembobotan dengan menggunakan metode *Term Frequency Inverse Document Frequency* (TF-IDF). Proses ini digunakan untuk mengubah data teks yang telah melalui tahap *preprocessing* menjadi bentuk numerik agar dapat dikenali oleh

algoritma klasifikasi. Cara kerja metode ini adalah dengan melihat seberapa sering suatu kata muncul di satu ulasan, kemudian membandingkannya dengan seberapa sering kata itu muncul di semua ulasan yang ada. Kata-kata yang sering muncul namun tidak memiliki makna spesifik akan diberikan bobot yang rendah, sedangkan kata-kata yang mengandung sentimen kuat akan mendapatkan bobot yang lebih tinggi. Hasil dari pembobotan TF-IDF ini dapat mendukung model *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam proses klasifikasi sentimen.

## 2.5. Metode Naïve Bayes

Algoritma *Naïve Bayes* didasarkan pada Teorema Bayes, untuk membuat prediksi data dengan menghitung probabilitas dengan asumsi independensi yang kuat ini membuatnya “*naïve*”. Kelebihan utama dari *Naïve Bayes* dibandingkan dengan algoritma klasifikasi lain adalah efisiensi waktu pelatihan yang hanya melibatkan perhitungan frekuensi dan probabilitas sederhana, metode ini cocok untuk data jenis apa pun (kuantitatif atau kualitatif), efektif walaupun digunakan pada jumlah data dengan skala yang terbatas, serta mudah diimplementasikan. Persamaan dasar yang digunakan dalam penentuan kelas sebagai berikut (1)(Julkarnain & Yustiardin, 2024):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

X	= Data dengan kelas yang belum diketahui
H	= Hipotesis data X merupakan suatu kelas spesifik
P(H X)	= Probabilitas hipotesis H berdasarkan kondisi X
P(H)	= Probabilitas hipotesis H

## 2.6. Metode Support Vector Machine

*Support Vector Machine* (SVM) bekerja dengan cara memetakan kata-kata atau kalimat ke dalam ruang koordinat, kemudian menarik garis pemisah (*hyperplane*) untuk membedakan komentar yang bersifat positif dan komentar yang bersifat negatif. Beberapa kelebihan yang dimiliki oleh SVM, SVM dapat menangani data dengan jumlah fitur yang banyak tanpa mengalami penurunan performa yang drastis, SVM memiliki tingkat akurasi yang lebih stabil dan tinggi untuk klasifikasi teks yang kompleks. Persamaan dasar *hyperplane* ditunjukkan dalam persamaan (2):

$$w \times x + b = 0 \quad (2)$$

w	= bobot margin
x	= titik data
b	= bias

Data nonlinear perlu dibantu dengan kernel SVM untuk memperoleh *hyperplane* yang tepat. Fungsi kernel bekerja dengan mentransformasikan data berdimensi rendah ke dimensi yang lebih tinggi agar pola non-linear di dalamnya dapat dipisahkan secara linear. Proses ini

bertujuan untuk menemukan *hyperplane* optimal yang mampu membedakan kategori dalam dataset secara lebih akurat. Berikut jenis kernel pada SVM:

1. Kernel Linear

Kernel linear diimplementasikan pada dataset yang telah memiliki distribusi data linear. Fungsi ini bekerja secara optimal ketika data yang dianalisis sudah dapat dipisahkan secara langsung dalam ruang fitur aslinya, tanpa memerlukan transformasi ke dimensi yang lebih kompleks. Persamaan untuk fungsi kernel linear adalah (3):

$$K(x, xi) = \text{sum}(x \times xi) \tag{3}$$

2. Kernel Polinomial

Kernel polinomial merupakan solusi efektif dalam menangani dataset yang tidak dapat dipisahkan secara linear. Fungsi ini sering diterapkan dalam algoritma SVM, karena kemampuannya dalam memproses klasifikasi pada data latih yang telah melalui tahap normalisasi. Persamaan untuk fungsi kernel polinomial adalah (4):

$$K(x, xi) = 1 + \text{sum}(x \times xi)^d \tag{4}$$

3. Kernel RBF (*Radial Basic Function*)

Kernel *Radial Basis Function* (RBF), atau yang sering disebut sebagai kernel Gaussian, merupakan pendekatan utama dalam menangani klasifikasi data non-linear. Fungsi ini dikenal memiliki fleksibilitas tinggi serta performa yang unggul melalui optimasi parameter yang tepat. Keunggulan utamanya terletak pada kemampuan menghasilkan nilai *error* yang lebih rendah selama proses pelatihan dibandingkan dengan jenis kernel lainnya. Persamaan untuk fungsi kernel RBF adalah (5):

$$K(x, xi) = \exp(-\text{gamma} \times \text{sum}((x - xi)^2)) \tag{5}$$

## 2.7. Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan model dalam melakukan prediksi dengan menggunakan *Confusion Matrix*. *Confusion Matrix* membandingkan apa yang diprediksi oleh model dan apa yang sebenarnya terjadi di lapangan. Terdapat empat komponen yang digunakan dalam melakukan evaluasi, yaitu:

- a. True Positive (TP) : Model memprediksi benar dan kenyataannya memang benar.
- b. True Negative (TN) : Model memprediksi salah dan kenyataannya memang salah.
- c. False Positive (FP) : Model memprediksi benar tetapi kenyataannya salah.
- d. False Negative (FN) : Model memprediksi salah tetapi kenyataannya benar.

*Accuracy* menentukan seberapa sering model menebak dengan benar secara keseluruhan, menggunakan rumus dalam persamaan (6) (Hidayat et al., 2024):

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

*Precision* mengukur tingkat keandalan model dalam mengidentifikasi instans positif, seperti yang dijelaskan dalam persamaan (7) (Hidayat et al., 2024):

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

*Recall* mengukur kemampuan model untuk mengidentifikasi seluruh instans positif yang tersedia dalam populasi data, seperti yang dijelaskan dalam persamaan (8) (Hidayat et al., 2024):

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

*F1-Score* merupakan rata-rata harmonik dari *Precision* dan *Recall*. *F1-Score* bertujuan untuk memberikan penalti yang signifikan jika salah satu dari kedua metrik tersebut memiliki nilai yang rendah, sebagaimana dijelaskan dalam persamaan (9) (Hidayat et al., 2024):

$$2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

## 2.8. Evaluasi Latent Dirichlet Allocation (LDA) Coherene

*Latent Dirichlet Allocation* (LDA) adalah metode dalam pembelajaran mesin yang berbasis *unsupervised learning* yang digunakan untuk menentukan topik-topik tersembunyi. Algoritma ini bekerja dengan mengelompokkan kata-kata berdasarkan kemunculannya secara bersamaan untuk membentuk kategori topik tertentu. Metode LDA bekerja dengan prinsip bahwa setiap dokumen mengandung proporsi topik yang beragam, dan setiap topik terdiri dari kumpulan kata yang saling berkaitan. LDA dilakukan agar isi dari data yang telah diproses dapat lebih dipahami.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Data

Pada tahap ini dilakukan proses *scraping* pada ulasan Hotel Neo Pontianak melalui *Google Maps*, mendapatkan data ulasan sebanyak 1387 dengan data sentimen positif sebanyak 983 dan data sentimen negatif sebanyak 373. Pelabelan sentimen positif dan sentimen negatif dilakukan secara manual, proses ini dilakukan secara manual karena ada kemungkinan pengguna memberikan rating 4 atau 5 tetapi terdapat komentar negatif pada ulasan yang mereka berikan. Contoh pelabelan manual ditunjukkan pada tabel 1.

**Tabel 1.** Hasil Label Manual 4 Sampel


Komentar	Label
Pelayanan nya bagus, tempat nyaman banget	Positif
Enak tempatnya mau cari makan tinggal jln kaki, mau cari oleh oleh dekat.. recommended deh ni hotel	Positif
Sayang sekali attitude pegawainya yg melayani sy kurang baik	Negatif
Staf tidak profesional...sangat kecewa	Negatif

### 3.2. Text Processing

Setelah melakukan pelabelan, dilakukan beberapa tahap untuk menghasilkan data yang dapat digunakan untuk melakukan pemrosesan yakni: *cleaning*, *case folding*, normalisasi, tokenisasi, *stopword*, *stemming* (Komalla et al., 2025). Tahap-tahap pra-pemrosesan data ditunjukkan dalam tabel 2, tabel 3, tabel 4, tabel 5, tabel 6, dan tabel 7.

1. *Cleaning* : Proses ini menghapus karakter non-abjad seperti tanda baca, angka, emoji, dan simbol dari ulasan pengguna.

**Tabel 2.** Proses *Cleaning*

<b>Sebelum <i>Cleaning</i></b>	<b>Setelah <i>Cleaning</i></b>
PELAYANAN NYAA BAGUS, TEMPAT NYAMAN BANGET	PELAYANAN NYAA BAGUS TEMPAT NYAMAN BANGET
Pelayanannya ramahh bgtt, makanannya juga wuenakk poll, pemandangannya apalagi kerennnn abiss 	Pelayanannya ramahh bgtt makanannya juga wuenakk poll pemandangannya apalagi kerennnn abiss

2. *Case Folding* : Pada proses ini mengubah semua ulasan huruf besar menjadi huruf kecil agar data yang diproses tetap konsisten.

**Tabel 3.** Proses *Case Folding*

<b>Sebelum <i>Case Folding</i></b>	<b>Setelah <i>Case Folding</i></b>
PELAYANAN NYAA BAGUS TEMPAT NYAMAN BANGET	pelayanan nyaa bagus tempat nyaman banget
Pelayanannya ramahh bgtt makanannya juga wuenakk poll pemandangannya apalagi kerennnn abiss	pelayanannya ramahh bgtt makanannya juga wuenakk poll pemandangannya apalagi kerennnn abiss

3. Normalisasi : Proses ini dilakukan untuk mengubah kata singkatan atau tidak baku menjadi bentuk dasar sesuai dengan pedoman bahasa(kamus) dan menghilangkan huruf berulang, bertujuan untuk mengurangi variasi kata.

**Tabel 4.** Proses Normalisasi

<b>Sebelum Normalisasi</b>	<b>Setelah Normalisasi</b>
pelayanan nyaa bagus tempat nyaman banget	pelayanan nya bagus tempat nyaman banget
pelayanannya ramahh bgtt makanannya juga wuenakk poll pemandangannya apalagi kerennnn abiss	pelayanannya ramah banget makanannya juga enak pol pemandangannya apalagi keren abis

4. Tokenisasi : Pada proses ini memecah sekumpulan kalimat menjadi kata-kata individu, bertujuan untuk memudahkan dalam pemrosesan data.

**Tabel 5.** Proses Tokenisasi

Sebelum Tokenisasi	Setelah Tokenisasi
pelayanan nya bagus tempat nyaman banget	"pelayanan" "nya" "bagus" "tempat" "nyaman" "banget"
pelayanannya ramah banget makanannya juga enak pol pemandangannya apalagi keren abis	"pelayanan" "nya" "ramah" "banget" "makanan" "nya" "juga" "enak" "pol" "pemandangan" "nya" "apa" "lagi" "keren" "abis"

5. *Stopword* : Pada proses ini menghilangkan kata depan, kata ganti, kata bantu, dan kata penghubung, bertujuan untuk mengurangi jumlah kata.

**Tabel 6.** Proses *Stopword*

Sebelum <i>Stopword</i>	Setelah <i>Stopword</i>
"pelayanan" "nya" "bagus" "tempat" "nyaman" "banget"	"pelayanan" "bagus" "nyaman" "banget"
"pelayanan" "nya" "ramah" "banget" "makanan" "nya" "juga" "enak" "pol" "pemandangan" "nya" "apa" "lagi" "keren" "abis"	"pelayanan" "ramah" "banget" "makanan" "enak" "pemandangan" "keren"

6. *Stemming* : Pada proses ini mengubah bentuk kata menjadi kata dasar dengan menghilangkan imbuhan, awalan, akhiran, dan sisipan, bertujuan untuk mengurangi kata yang memiliki makna yang sama.

**Tabel 7.** Proses *Stemming*

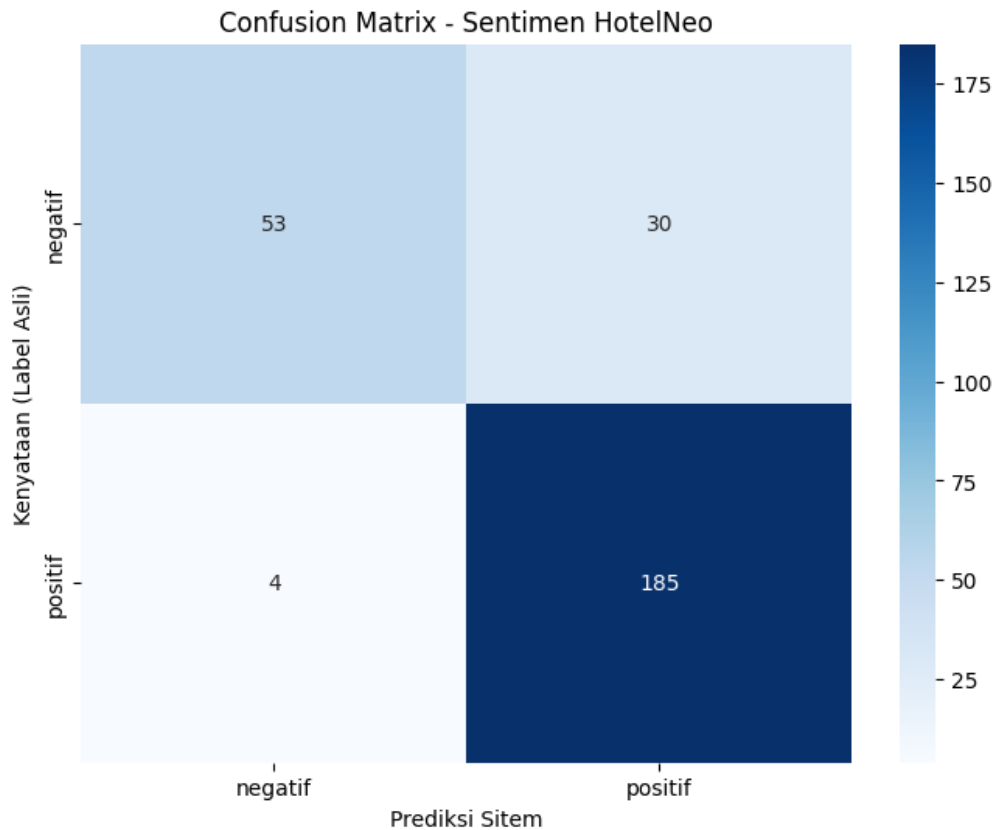
Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
"pelayanan" "bagus" "nyaman" "banget"	"layan" "bagus" "nyaman" "banget"
"pelayanan" "ramah" "banget" "makanan" "enak" "pemandangan" "keren"	"layan" "ramah" "banget" "makan" "enak" "pandang" "keren"

### 3.3. Pembagian Dataset

Pada tahap ini dilakukan proses klasifikasi pada data yang telah melewati tahap pemrosesan, dengan membagi menjadi dua bagian yaitu dataset pelatihan dan dataset pengujian. Proses klasifikasi menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) menggunakan rasio 80% : 20%. Setelah proses klasifikasi telah dilakukan selanjutnya akan dilakukan perbandingan tingkat *accuracy*, *precision*, *recall*, dan *f1-score* antara algoritma *Naïve Bayes* dan SVM untuk menentukan algoritma yang lebih baik (Prastyo et al., 2024).

### 3.4. Hasil Evaluasi Model

Performa algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) diuji menggunakan *confusion matrix* untuk mendapatkan hasil *accuracy*, *precision*, *recall*, dan *f1-score* yang optimal. Hasil pengujian pada algoritma *Naïve Bayes* ditunjukkan pada gambar 1.

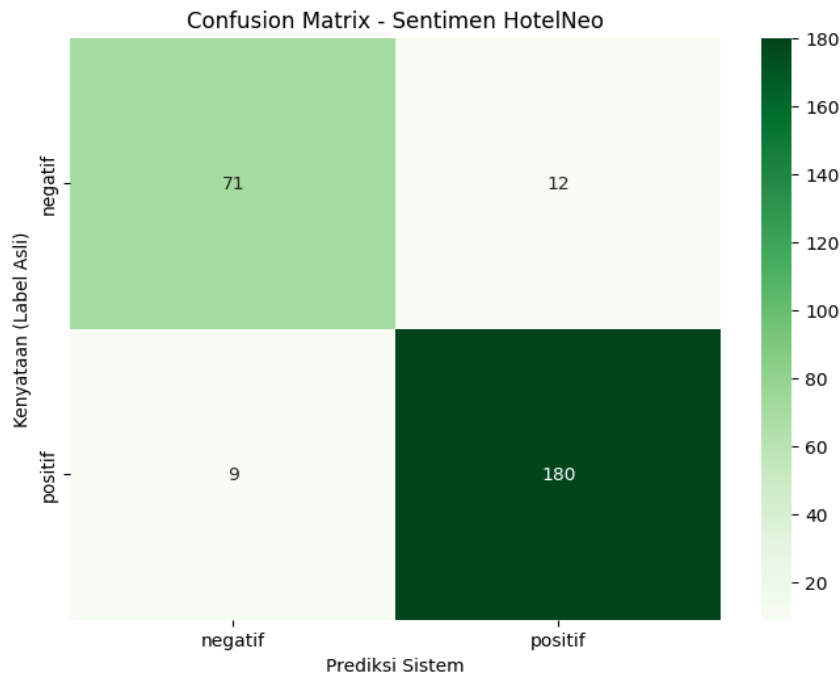


**Gambar 1.** *Confusion Matrix* Model *Naïve Bayes*

Gambar 1 mengilustrasikan hasil pengujian pada model *Naïve Bayes* yang menghasilkan empat kombinasi klasifikasi untuk menentukan algoritma dengan kinerja yang lebih baik pada penelitian ini dengan menggunakan 272 data ulasan.

1. *True Positive* : Diantara 189 komentar positif, model memprediksi sebanyak 185 data komentar positif dengan benar.
2. *False Negative* : Diantara 189 komentar positif, model memprediksi sebanyak 4 data komentar positif tetapi diprediksi negatif.
3. *True Negative* : Diantara 83 komentar negatif, model memprediksi sebanyak 53 data komentar negatif dengan benar.
4. *False Positive* : Diantara 83 komentar negatif, model memprediksi sebanyak 30 data komentar negatif tetapi diprediksi positif.

Selanjutnya dilakukan pengujian pada algoritma *Support Vector Machine* (SVM), hasil pengujian ditunjukkan pada gambar 2.



**Gambar 2.** Confusion Matrix Model Support Vector Machine

Gambar 2 mengilustrasikan hasil pengujian pada model (SVM) yang menghasilkan empat kombinasi klasifikasi untuk menentukan algoritma dengan kinerja yang lebih baik pada penelitian ini dengan menggunakan 272 data ulasan.

1. *True Positive* : Diantara 189 komentar positif, model memprediksi sebanyak 180 data komentar positif dengan benar.
2. *False Negative* : Diantara 189 komentar positif, model memprediksi sebanyak 9 data komentar positif tetapi diprediksi negatif.
3. *True Negative* : Diantara 83 komentar negatif, model memprediksi sebanyak 71 data komentar negatif dengan benar.
4. *False Positive* : Diantara 83 komentar negatif, model memprediksi sebanyak 12 data komentar negatif tetapi diprediksi positif.

Hasil perbandingan akurasi, *precision*, *recall*, dan *f1-score* algoritma *Naïve Bayes* dan algoritma SVM ditunjukkan pada tabel 6.

**Tabel 6.** Perbandingan Hasil *Accuracy*, *Precision*, *Recall*, dan *F1-Score*

Algoritma	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	87,50%	89,50%	81%	84%
SVM	92,28%	91,50%	90,50%	90,50%

Berdasarkan tabel 6, hasil pengukuran dari dua algoritma yang digunakan, algoritma *Naïve Bayes* mendapatkan hasil *accuracy* sebesar 87,50% sementara algoritma SVM mendapatkan hasil *accuracy* sebesar 92,28%. Algoritma *Naïve Bayes* mendapatkan hasil *precision* sebesar 89,50% sementara algoritma SVM mendapatkan hasil *precision* sebesar 91,50%. Algoritma *Naïve Bayes* mendapatkan hasil *recall* sebesar 81% sementara algoritma SVM mendapatkan hasil *recall* sebesar 90,50%. Algoritma *Naïve Bayes* mendapatkan *f1-score* sebesar 84% sementara algoritma SVM mendapatkan *f1-score* sebesar 90,50%. Berdasarkan hasil yang telah didapatkan maka dapat disimpulkan bahwa algoritma SVM lebih unggul dibandingkan algoritma *Naïve Bayes* pada penelitian ini (Firdaus & Suryono, 2025).

### 3.5. Hasil Latend Dirichlet Allocation

Setelah dilakukan perbandingan antara algoritma *Naïve Bayes* dan algoritma *Support Vector Machine (SVM)*, selanjutnya dilakukan pemodelan topik menggunakan *Latend Dirichlet Allocation (LDA)*. Proses ini dilakukan untuk membantu memahami isi dari data yang telah diproses. Hasil *Latend Dirichlet Allocation (LDA)* kata positif dan kata negatif ditunjukkan pada gambar 3 dan gambar 4.

	Topik 1	Topik 2	Topik 3
0	layan	hotel	layan
1	hotel	kamar	ramah
2	lokasi	layan	bagus
3	ramah	bersih	kasih
4	strategis	enak	terima
5	makan	ramah	hotel
6	enak	nyaman	bantu
7	kamar	makan	mas
8	kota	pontianak	resepsionis
9	bersih	bagus	mbak

**Gambar 3.** *Latend Dirichlet Allocation* Kata Positif

Berdasarkan hasil dari *Latend Dirichlet Allocation (LDA)* pada gambar 3, alasan memilih 3 topik dikarenakan 3 topik memiliki *coherence score* tertinggi diantara jumlah topik yang lain (Raudhatul Eni et al., 2025). Berdasarkan kata-kata tersebut dapat disimpulkan bahwa topik 1 membahas mengenai kenyamanan dan lokasi dari hotel Neo yang memiliki letak strategis di kota Pontianak. Pada topik 2 membahas mengenai kualitas kamar hotel Neo yang bersih, nyaman, kemudian juga membahas mengenai makanan yang enak. Kemudian topik 3 membahas mengenai pelayanan dari hotel Neo yang memiliki pelayanan yang bagus dan ramah, resepsionis yang membantu, mbak dan mas yang ramah, serta ucapan terima kasih.

	Topik 1	Topik 2	Topik 3	Topik 4	Topik 5	Topik 6	Topik 7	Topik 8	Topik 9	Topik 10
0	kamar	kamar	kamar	parkir	kamar	hotel	kamar	parkir	taksi	makan
1	sprei	hotel	hotel	susah	parkir	layan	bersih	hotel	hotel	sarap
2	kotor	air	asap	toilet	suara	jam	mandi	bagus	online	menu
3	parkir	reception	ya	mobil	mobil	kamar	hotel	area	kamar	ya
4	kuning	malam	in	bawa	sempit	buruk	kotor	buruk	tamu	hotel
5	bau	sopan	bau	pas	pakai	check	air	ramah	sedia	kurang
6	bersih	kondisi	remang	batas	hotel	bersih	makan	security	pakai	enak
7	bed	tamu	room	kamar	jalan	in	jam	mobil	lift	air
8	bantal	makan	tamu	satpam	banget	ramah	minum	sulit	bintang	lampu
9	ruang	kasur	masuk	depan	tidur	rapi	tanggal	kembali	tugas	kasih

**Gambar 4.** *Latend Dirichlet Allocation* kata Negatif

Berdasarkan hasil dari *Latent Dirichlet Allocation* (LDA) pada gambar 4, alasan memilih 10 topik dikarenakan 10 topik memiliki *coherence score* tertinggi diantara jumlah topik yang lain (Raudhatul Eni et al., 2025). Dengan memilih 10 topik pada kata-kata negatif menemukan beberapa aspek yang perlu diperbaiki, pada topik 1,3, dan 5, membahas terkait kebersihan ruang kamar hotel yang kotor dan bau pada spreng, bantal dan juga gangguan suara yang dialami. Kemudian topik 2 dan 6 membahas mengenai pelayanan dari pihak resepsionis dan juga proses check-in. Pada topik 4,8, dan 9 membahas mengenai akses parkir yang susah, petugas keamanan, dan juga akses transportasi taksi online. Topik 7 membahas mengenai kamar mandi dan kualitas air. Dan juga topik 10 yang membahas mengenai menu sarapan yang kurang enak.

#### 4. KESIMPULAN

Pada penelitian ini, menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) sebagai perbandingan dari kedua algoritma ini mana yang memiliki akurasi lebih tinggi untuk menganalisis sentimen ulasan pelanggan pada Hotel Neo Pontianak, kemudian dilanjutkan menggunakan algoritma *Latent Dirichlet Allocation* (LDA) yang digunakan untuk memodeling topik. Berdasarkan hasil penelitian yang dilakukan hasil akurasi SVM lebih tinggi dibandingkan hasil akurasi *Naïve Bayes* yakni SVM mendapatkan akurasi sebesar 90,45% sementara *Naïve Bayes* mendapatkan akurasi sebesar 85,20%. Berdasarkan perbandingan akurasi yang didapatkan maka disarankan kepada manajemen hotel untuk menggunakan algoritma SVM untuk melakukan analisis sentimen pada ulasan pelanggan. Kemudian didapatkan hasil topik model dari LDA yakni kata kunci positif dan kata kunci negatif, pada kata kunci positif didominasi oleh kualitas pelayanan yang bagus dan ramah, kualitas kamar yang nyaman dan bersih, dan juga lokasi hotel Neo Pontianak yang strategis. Namun pada kata kunci negatif terdapat beberapa aspek yang perlu diperbaiki oleh manajemen hotel yakni: kualitas makanan, kebersihan kamar, kendala pada pelayanan waktu check-in, gangguan suara pada kamar, kerusakan fasilitas kamar, dan permasalahan pada tempat parkir dan akses taksi online. Maka disarankan kepada manajemen hotel untuk memperbaiki beberapa aspek terutama pada kebersihan ruang kamar tidur dan kamar mandi, karyawan perlu memastikan ruang kamar sudah bersih sebelum ditempati oleh tamu, kemudian pelayanan pada waktu check-in agar pelanggan tidak menunggu lama untuk melakukan proses check-in. Selanjutnya ada gangguan pada suara kamar, karena letak hotel Neo yang berada di tengah kota Pontianak sehingga banyak kendaraan yang melewati hotel Neo, maka disarankan kepada manajemen hotel untuk memberikan peredam suara pada kamar agar pelanggan dapat tidur dengan lebih nyaman dan tenang. Untuk akses parkir apabila kesulitan dapat dibantu oleh petugas untuk memarkirkan kendaraan tamu. Pelatihan bagi satpam agar ramah kepada tamu dan menjaga keamanan hotel agar tamu dapat beristirahat dengan tenang, dan juga memperbaiki kualitas makanan.

#### 5. REFERENSI

- Abel Laia, N., & Barus, S. P. (2025). Analisis Sentimen YouTube: “Di Balik Ambisi Jokowi dalam IKN.” *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, 5(1), 07–12. <https://doi.org/10.55382/jurnalpustakaai.v5i1.891>
- Alexander Romian Simarmata, M. Z. (2023). *Sentiment Analysis of Hotel Reviews Using Support Vector Machine*. <https://doi.org/10.33022/ijcs.v12i5.3405>

- Apriliansyah, R. D. R., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN PENGUNJUNG DI PANTAI KEJAWANAN. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5774>
- Azizah, H., Syuhada, F., & Sa'adati, Y. (2024). SENTIMEN ANALISIS TEMPAT WISATA BERDASARKAN ULASAN GOOGLE MAPS MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus Bukit Merese) Sentiment Analysis of Tourist Destinations Based on Google Maps Reviews Using the Naïve Bayes Method (Case Study of Bukit Merese). In *SIJ* (Vol. 7, Number 2). <https://doi.org/10.37824/sij.v7i2.2024.753>
- Firdaus, N. D., & Suryono, R. R. (2025). Perbandingan Algoritma Naive Bayes dan SVM dalam Analisis Sentimen Pengguna AI di Platform X. *Technology and Science (BITS)*, 6(4). <https://doi.org/10.47065/bits.v6i4.7081>
- Gilang, O. :, & Hilmawan, H. (2025). PT. Media Akademik Publisher LITERATUR REVIEW : EFEKTIFITAS PENERAPAN METODE NAÏVE BAYES UNTUK PREDIKSI KELULUSAN. *JMA*, 3(6), 3031–5220. <https://doi.org/10.62281>
- Haq, M. Z., Octiva, C. S., Ayuliana, A., Nuryanto, U. W., & Suryadi, D. (2024). Algoritma Naive Bayes untuk Mengidentifikasi Hoaks di Media Sosial. *Jurnal Minfo Polgan*, 13(1), 1079–1084. <https://doi.org/10.33395/jmp.v13i1.13937>
- Haris Tri Saputra, Alfath Damanik, M. Mayo Shaquille, Marini Adibah Rusydi, Muhammad Habib Riziq, & Naya Septia Zulva. (2025). Analisis Modelling Pada Reviewes Lazada Indonesia Menggunakan Latent Dirichlet Allocation (LDA) Untuk Optimalisasi Strategi Bisnis. *JEKIN - Jurnal Teknik Informatika*, 5(1), 361–371. <https://doi.org/10.58794/jekin.v5i1.1325>
- Hidayat, R., Sy, Y. S., Sujana, T., Husnah, M., Saputra, H. T., & Okmayura, F. (2024). Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine. *BIOS : Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 5(2), 161–168. <https://doi.org/10.37148/bios.v5i2.152>
- Junaedi, Hendra Gunawan, A., Kuswanto, V., & Jonathan. (2024). Tinjauan Support Vector Machine dalam Text-Mining untuk Analisis Sentimen di Sektor Pariwisata. *Bit-Tech*, 7(2), 323–330. <https://doi.org/10.32877/bt.v7i2.1810>
- Komalla, D., Guntur Alam, R., & Wijaya, A. (2025). Analisis Performa Algoritma Naive Bayes dan SVM Menggunakan Python Pada Ulasan Sentimen Game Roblox. *Jurnal Riset Komputer*, 12(6), 2407–389. <https://doi.org/10.30865/jurikom.v12i6.9396>
- M. Julkarnain, M. Y. (2024). Penerapan Algoritma Naive Bayes dalam Memprediksi Lulus Tepat Waktu Mahasiswa. *Digital Transformation Technology (Digitech)*. <https://doi.org/10.47709/digitech.v4i2.4963>
- Mochamad Amry Assiva. (2024). Analisis Sentimen Terhadap Pariwisata di Kabupaten Grobogan Berbasis Orange Menggunakan Naive Bayes. *INNOVATIVE: Journal Of Social Science Research*. <https://doi.org/10.31004/innovative.v4i6.16469>

- Nadhifah, S., Aini, F. N., Kusumawardhani, H. H., & Febrianto, M. Y. (2024). *Analisis Sentiment Ulasan Aplikasi Gopay Pada Google Play Store Menggunakan Algoritma Support Vector Machine*. 14(1). <https://doi.org/10.48144/suryainformatika.v14i1.1787>
- Prastyo, P. A., Berlilana, B., & Tahyudin, I. (2024). Analisis Sentimen dan Pemodelan Topik pada Ulasan Pengguna Aplikasi myIM3 Menggunakan Support Vector Machine dan Latent Dirichlet Allocation. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1618–1626. <https://doi.org/10.47065/bits.v6i3.6268>
- Raudhatul Eni, A., Christien Arisona, D., Ngurah Adhi Wibawa, G., Yahya, I., & Abapihi, B. (2025). PEMODELAN TOPIK DAN ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MOBILE JKN MENGGUNAKAN ALGORITMA LATENT DIRICHLET ALLOCATION DAN RANDOM FOREST. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Number 6). <https://doi.org/10.36040/jati.v9i6.15559>
- Reni Dwi Astuti, A., Cahyono, N., & Amikom Yogyakarta, U. (n.d.). Analisis Topic Modelling Persepsi Pengguna Internet Menggunakan Metode Latent Dirichlet Allocation. *Indonesian Journal of Computer Science Attribution*, 12(1), 2023–2326. <https://doi.org/10.33022/ijcs.v12i1.3155>
- Singgalen, Y. A. (2023). Extract Sentiment and Support Vector Machine (SVM) Performance of Hotel Guest Review Classification. *Building of Informatics, Technology and Science (BITS)*, 5(3). <https://doi.org/10.47065/bits.v5i3.4737>
- Slam, B. E., Irawan, F., Efranda, N., & Herikson, R. (n.d.). *JIP (Jurnal Informatika Polinema) IMPLEMENTASI MACHINE LEARNING UNTUK KLASIFIKASI BUKU OTOMATIS PADA PERPUSTAKAAN DIGITAL*. <https://doi.org/10.33795/jip.v11i3.7298>
- Sudirman, Y., Dwi, ), Simatupang, S., & Artikel, R. (2025). *ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN HOTEL XYZ DI KOTA TANGERANG DENGAN ALGORITMA SVM 1) INFO ARTIKEL ABSTRAK*. 4(4), 370–377. <https://doi.org/10.55123>
- Suparyati Suparyati, A. F. (2022). *Analisis Sentimen Dengan Klasifikasi Naïve Bayes pada Review HotelTripadvisor*. <https://doi.org/10.33884/jif.v10i01.4524>
- Syahputra, R., Yanris, G. J., & Irmayani, D. (2022). SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter. *Sinkron*, 7(2), 671–678. <https://doi.org/10.33395/sinkron.v7i2.11430>
- Wahyuni, S. (2024). *Bulletin of Information Technology (BIT) Analisis Sentimen Publik Terhadap Pariwisata Aceh di Media Sosial X Menggunakan Algoritma Naive Bayes Classifier*. 5(4), 269–278. <https://doi.org/10.47065/bit.v5i2.1700>