



# Perbandingan Naïve Bayes dan Support Vector Machine Dalam Analisis Sentimen Google Maps Pusat Perbelanjaan

Eliza Cahyaningrum, Astrid Novita Putri\*

Fakultas Teknik Informasi dan Komunikasi, Program Studi Teknik Informatika, Universitas Semarang, Semarang, Indonesia

Email: <sup>1</sup>elizacahya14@gmail.com, <sup>2,\*</sup>astrid@usm.ac.id

Email Penulis Korespondensi: astrid@usm.ac.id

Submitted: 26/03/2026; Accepted: 02/06/2026; Published: 05/06/2026

**Abstrak**—Pesatnya pertumbuhan ulasan pengguna pada *Google Maps* tidak selalu diiringi dengan kemudahan dalam memahami sentimen yang terkandung di dalamnya, sehingga wisatawan dan masyarakat sering mengalami kesulitan dalam menentukan pusat perbelanjaan dengan reputasi dan kualitas layanan yang baik. Kurangnya informasi mengenai tingkat kepuasan pengunjung serta berbagai permasalahan fasilitas, seperti kepadatan pengunjung, keterbatasan tempat parkir, dan kenyamanan fasilitas umum, yang disertai dengan banyaknya ulasan yang bersifat subjektif dan tidak sistematis, mengakibatkan analisis sentimen secara manual menjadi tidak efektif dan berpotensi menghasilkan kesimpulan yang kurang akurat. Penelitian ini ditujukan guna menganalisis sentimen ulasan *Google Maps* terhadap pusat perbelanjaan di Kota Semarang menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes*. Data diambil dari lima pusat perbelanjaan dengan jumlah ulasan terbanyak di Kota Semarang, yaitu Paragon Mall, Mall Ciputra, Java Mall, DP Mall, dan Queen City Mall. Metode penelitian meliputi *text preprocessing*, pembobotan TF-IDF, serta klasifikasi sentimen ke dalam tiga kelas, diantaranya negatif, netral, dan positif. Data dibagi dengan rasio 80:20 antara data latih dan data uji. Temuan penelitian mengindikasikan, metode *Naïve Bayes* memperoleh akurasi mencapai 85,56%, sedangkan metode *Support Vector Machine* (SVM) memperoleh akurasi mencapai 89,20%. Mengacu temuan tersebut, metode SVM memperlihatkan performa yang lebih baik dalam klasifikasi sentimen ulasan *Google Maps* terhadap Pusat Perbelanjaan di Kota Semarang.

**Kata Kunci:** Analisis Sentimen; Google Maps; Support Vector Machine; Naïve Bayes; Pusat Perbelanjaan

**Abstract**—The rapid growth of user reviews on Google Maps is not always accompanied by ease in understanding the sentiment contained within them, causing tourists and the general public to face difficulties in determining shopping centers with good reputation and service quality. The lack of information regarding visitor satisfaction levels, along with various facility-related issues such as crowd density, limited parking space, and the comfort of public facilities, combined with the large number of subjective and unstructured reviews, makes manual sentiment analysis ineffective and potentially leads to less accurate conclusions. This investigation aims to analyze sentiment from Google Maps reviews of shopping centers in the city of Semarang utilizing the Support Vector Machine (SVM) and Naïve Bayes methods. The data were collected from five shopping centers with the highest number of reviews in Semarang, namely Paragon Mall, Mall Ciputra, Java Mall, DP Mall, and Queen City Mall. The investigation method includes text preprocessing, TF-IDF weighting, and sentiment classification into three classes: negative, neutral, and positive. The dataset was divided into training and testing data with a ratio of 80:20. The outcomes reveal that the Naïve Bayes method achieved an accuracy of 85.56%, while the Support Vector Machine (SVM) method achieved an accuracy of 89.20%. Considering the outcomes, the SVM method performs better in classifying sentiment from Google Maps reviews of shopping centers in Semarang.

**Keywords:** Sentiment Analysis; Google Maps; Support Vector Machine; Naïve Bayes; Shopping Centers

## 1. PENDAHULUAN

Perkembangan teknologi informatika di era transformasi digital telah mengubah paradigma masyarakat dalam menyampaikan aspirasi. *Google Maps* kini menjadi instrumen *big data* yang krusial bagi publik untuk memberikan ulasan dan *rating* terhadap fasilitas umum, khususnya pusat perbelanjaan. Data ulasan yang didapat pada *Google Maps*, terdapat lima pusat perbelanjaan dengan jumlah ulasan terbanyak di Kota Semarang, yaitu Paragon Mall, Mall Ciputra, Java Mall, DP Mall, dan Queen City Mall. Namun, penilaian yang diberikan oleh pengguna pada *Google Maps* tersebut sering kali hanya menjadi tumpukan data teks tidak terstruktur yang sulit dipahami secara cepat. Kondisi ini menyebabkan informasi yang terkandung di dalam ulasan tidak dapat dimanfaatkan secara optimal oleh masyarakat maupun pengelola layanan. Guna mengetahui sentimen dari tumpukan komentar yang sangat beragam tersebut, diperlukan alat bantu komputasi atau analisis sentimen guna mengekstraksi informasi yang bermakna bagi pengelola layanan [1]. Permasalahan yang muncul adalah kurangnya informasi mengenai tingkat kepuasan pengunjung serta berbagai permasalahan fasilitas, seperti kepadatan pengunjung, keterbatasan tempat parkir, dan kenyamanan fasilitas umum. Kondisi ini menyulitkan pengunjung maupun pengelola dalam melakukan evaluasi secara objektif. Sehingga banyak ulasan yang bersifat subjektif dan tidak terstruktur menyebabkan analisis sentimen secara manual menjadi tidak efektif dan berpotensi menghasilkan kesimpulan yang kurang akurat atau bias [2]. Itulah mengapa, perlu adanya pendekatan berbasis komputasi yang dapat memproses data ulasan dalam jumlah besar dengan cara otomatis sehingga dapat menghasilkan informasi yang lebih sistematis serta memudahkan pemahaman masyarakat. Dalam konteks ini, *data mining* berperan penting sebagai proses ekstraksi informasi berharga, pola, dan pengetahuan tersembunyi dari kumpulan data yang besar dan kompleks [3], sehingga mampu membantu mengidentifikasi pola sentimen dari data ulasan pengguna secara lebih sistematis.

Beberapa penelitian yang sudah dilakukan sebelumnya mengindikasikan, berbagai teknologi telah diterapkan untuk mengatasi permasalahan analisis sentimen pada data ulasan pengguna. Pada penelitian sebelumnya mengindikasikan, dalam perbandingan algoritma klasifikasi teks, *Support Vector Machine* (SVM) memperlihatkan

performa yang lebih stabil dan mampu menangani kompleksitas data teks secara baik dengan *accuracy* mencapai 90%, *precision* mencapai 89%, *recall* mencapai 88% dan *F1-Score* mencapai 88% dibanding metode *Naïve Bayes* yang mencapai hasil *accuracy* mencapai 88%, *precision* mencapai 88%, *recall* mencapai 85% dan *F1-Score* mencapai 86% [4]. Pada penelitian lain juga diidentifikasi metode *K-Nearest Neighbor* (KNN) dapat diaplikasikan dalam rangka memetakan tingkat kepuasan wisatawan berdasarkan ulasan *Google Maps* dengan tingkat *accuracy* yang cukup baik yaitu sebesar 78% [5]. Selain itu, pada penelitian lain juga mengindikasikan, perkembangan model berbasis *Transformer* seperti *IndoBERT* juga telah dimanfaatkan untuk menganalisis opini publik yang muncul pada media sosial dengan temuan analisis yang lebih mendalam dengan hasil *accuracy* mencapai 98% namun memiliki persoalan krusial yang perlu ditangani berkenaan dengan ketidakseimbangan kelas (*class imbalance*) karena memperlihatkan performa dengan akurasi terbaik pada kelas negatif dengan *recall* mencapai 99% namun memperlihatkan kelemahan signifikan saat mengidentifikasi sentimen positif dengan hanya mencapai hasil *recall* 29% [6]. Sementara itu, pada penelitian lainnya, metode *Support Vector Machine* (SVM) juga banyak diaplikasikan pada penelitian analisis sentimen dikarenakan memiliki proses komputasi yang relatif cepat serta performa yang cukup baik dalam klasifikasi teks melalui perolehan *accuracy* mencapai 85%, *precision* mencapai 100%, *recall* mencapai 70% dan *F1-Score* mencapai 82% [7]. Penggunaan teknik *web scraping* otomatis juga menjadi bagian integral untuk memastikan data yang diambil dari *Google Maps* bersifat *up-to-date* dan representatif [8]. Meskipun penelitian sebelumnya telah membandingkan metode *Naïve Bayes* dan *SVM* dalam konteks analisis sentimen [4], penelitian tersebut dilakukan pada domain ulasan aplikasi *LinkedIn* yang bersumber dari *Google Play Store*. Hal ini menunjukkan adanya perbedaan karakteristik data dibandingkan dengan ulasan pada *Google Maps* yang cenderung lebih beragam, tidak terstruktur, serta berkaitan langsung dengan pengalaman pengguna terhadap fasilitas fisik seperti pusat perbelanjaan. Selain itu, penelitian lain yang menggunakan objek pusat perbelanjaan di Kota Semarang masih menggunakan metode yang berbeda, seperti *Random Forest* dengan teknik penyeimbangan data (*SMOTE*) [2], sehingga belum memberikan gambaran komparatif antara metode klasifikasi seperti *Naïve Bayes* dan *SVM* pada domain tersebut. Diartikan, ada celah penelitian (*research gap*) yang kentara, yakni belum diadakannya penelitian terti yang secara spesifik membandingkan performa kedua metode tersebut pada data ulasan *Google Maps* untuk pusat perbelanjaan di Kota Semarang.

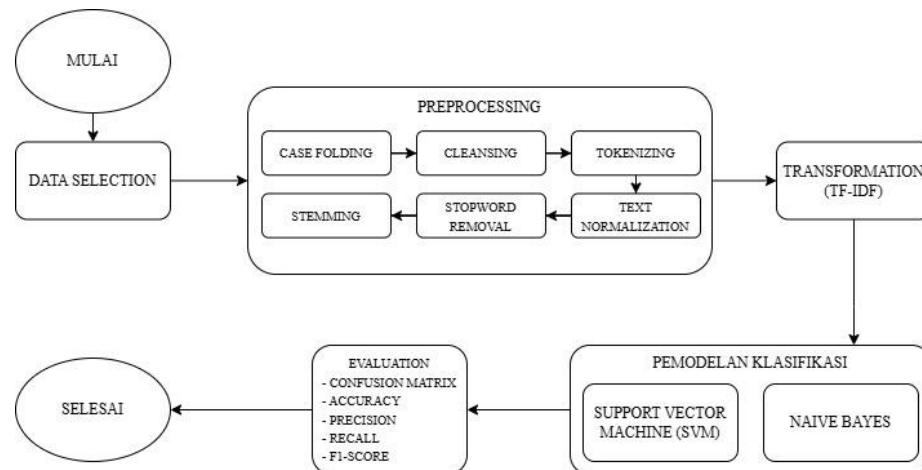
Pemilihan metode *Naïve Bayes* dan *SVM* pada penelitian ini didasarkan pada beberapa pertimbangan. Kedua metode termasuk algoritma klasifikasi yang paling umum diaplikasikan pada konteks analisis sentimen teks dikarenakan kesederhanaan, efisiensi komputasi, serta kemampuannya dalam menangani data berdimensi tinggi. Selain itu, meskipun metode lainnya, misal *K-Nearest Neighbor* dan *Logistic Regression* juga memperlihatkan performa yang cukup baik, penelitian ini secara khusus berfokus pada perbandingan dua metode klasik yang memiliki pendekatan berbeda, yaitu probabilistik (*Naïve Bayes*) dan berbasis margin (*SVM*), sehingga dapat memberikan pemahaman yang komprehensif berkenaan dengan karakter performa masing-masing metode dalam klasifikasi sentimen.

Berdasarkan permasalahan tersebut, penelitian ini menggunakan metode *Naïve Bayes* dan *SVM* melalui pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) guna melakukan analisis sentimen terkait ulasan pengguna pada *Google Maps* yang berkaitan dengan pusat perbelanjaan di Kota Semarang. Data ulasan diperoleh melalui teknik *web scraping* sehingga data yang digunakan bersifat aktual dan representatif. Melalui proses tersebut, ulasan pengguna akan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Penelitian ditujukan guna membandingkan kinerja metode *Naïve Bayes* dan *SVM* dalam mengklasifikasi sentimen terkait ulasan *Google Maps* pada pusat perbelanjaan di Semarang. Penelitian ini menyediakan gambaran mengenai persepsi pengunjung terhadap pusat perbelanjaan di Kota Semarang serta menyediakan informasi yang dapat dipergunakan menjadi acuan bahan evaluasi untuk pengelola guna memperbaiki mutu layanannya berdasarkan analisis data ulasan pengguna. Selain itu, penelitian ini berkontribusi melalui analisis komparatif performa dua metode klasifikasi populer, yaitu *Naïve Bayes* dan *Support Vector Machine*, pada data ulasan berbasis lokasi (*Google Maps*) yang memiliki karakteristik berbeda dibandingkan data media sosial atau aplikasi, serta menyajikan bukti empiris mengenai kemampuan tiap metode dalam konteks penanganan ketidakseimbangan data dan kompleksitas teks ulasan pengguna.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini memakai data primer berupa ulasan pengguna dan *rating* terbaru hingga tahun 2026 yang diperoleh melalui teknik *scraping* dari *Google Maps* pada lima pusat perbelanjaan di Kota Semarang dengan jumlah ulasan terbanyak. Data yang dipakai bersifat publik dan merepresentasikan pengalaman langsung pengguna. Metode penelitian mengacu pada pendekatan *Knowledge Discovery in Data* (KDD). Pada Gambar 1, merupakan alur penelitian yang digunakan pada penelitian ini. Proses diawali dari tahap *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Pada tahapan *preprocessing* diterapkan teknik *Natural Language Processing* (NLP) dalam rangka membersihkan dan menyiapkan data teks. Selanjutnya data ditransformasikan menggunakan TF-IDF, kemudian diklasifikasikan memakai metode *Naïve Bayes* dan *Support Vector Machine* (SVM), serta dievaluasi menggunakan *Confusion Matrix* dan metrik performa.



Gambar 1. Alur Penelitian

### 2.1.1 Data Selection

*Data selection* merupakan tahap pengumpulan sekaligus pemilihan data yang nantinya dipergunakan dalam penelitian. Proses pengumpulan data dilakukan memakai teknik *web scraping*, yakni cara pengambilan data secara otomatis dari sumber di internet menggunakan aplikasi atau kode pemrograman. Melalui teknik ini, data dapat dikumpulkan dalam jumlah besar dari berbagai sumber di dunia maya [9].

### 2.1.2 Preprocessing

Tahap *preprocessing* merupakan proses pembersihan dan penyiapan data teks agar dapat diproses oleh algoritma *machine learning*. Data ulasan yang diperoleh dari pengguna biasanya masih tidak terstruktur sehingga perlu dilakukan beberapa tahapan pengolahan teks di antaranya:

a. *Case Folding*

Proses mengonversi seluruh karakter huruf ke dalam huruf kecil [10].

b. *Cleansing*

Dalam pengolahan data teks, *cleansing* dilakukan untuk mereduksi entitas yang tidak diperlukan guna meningkatkan kualitas input data. Proses reduksi ini menargetkan komponen *noise* yang meliputi karakter HTML, simbol visual (emotikon), metadata media sosial (seperti *hashtag* dan *mention*), sintaks *retweet*, hingga protokol tautan digital (URL/alamat *website*) [11].

c. *Tokenizing*

*Tokenizing* merupakan tahapan pemecahan teks yang berawal dari kalimat menjadi kata per kata [12].

d. *Text Normalization*

*Text Normalization* yaitu mengubah kata tidak baku menjadi bentuk standar atau disebut dengan istilah *word normalization* [13]. Proses normalisasi dilakukan menggunakan kamus kata tidak baku (*slang dictionary*) yang disusun secara manual berdasarkan kata-kata yang sering muncul dalam ulasan pengguna Google Maps.

e. *Stopword Removal*

*Stopword Removal* merupakan proses menghilangkan kata-kata yang dianggap tidak memiliki makna [14].

f. *Stemming*

*Stemming* merupakan proses normalisasi kata-kata ke dalam bentuk dasar dengan memakai pustaka sastrawi guna menjamin konsistensi dalam teks [15].

### 2.1.3 Transformation TF-IDF

*Term Frequency-Inverse Document Frequency* atau yang biasa dikenal dengan TF-IDF adalah metode pembobotan *term* melalui penghitungan nilai *Term Frequency* (TF) untuk menghitung jumlah kemunculan suatu kata dalam sebuah dokumen. Makin seringnya sebuah kata muncul pada sebuah dokumen, maka skor TF juga makin besar. Sedangkan *Inverse Document Frequency* (IDF) digunakan dalam penghitungan tingkat kepentingan suatu kata terhadap keseluruhan dokumen dalam *dataset*. Kata yang muncul pada sebagian besar dokumen nantinya mempunyai nilai IDF yang lebih kecil dibandingkan kata yang frekuensi kemunculannya rendah [16]. Di tahapan ini hanya dipratikkan pembagian data ke dalam 1 skenario yakni 80% *data training* dan 20% *data testing*. Berikutnya, proses pembobotan kata nantinya memakai metode TF-IDF yang berfungsi untuk melakukan konversi data teks ke dalam vektor dengan bobot. Melalui proses pembobotan ini, setiap dokumen akan direpresentasikan dalam bentuk vektor numerik yang menggambarkan tingkat kepentingan kata dalam dokumen tersebut. Hasil transformasi TF-IDF kemudian dipergunakan sebagai fitur *input* pada proses klasifikasi sentimen dengan memakai metode *Naive Bayes* dan *Support Vector Machine* (SVM). Dengan demikian, model dapat mengenali pola kata yang berpengaruh terhadap penentuan kategori sentimen pada ulasan pengguna *Google Maps*. Berikut rumus dari *Term Frequency* (TF) adalah.

$$TF = \frac{\text{Total jumlah Dokumen dalam Korpus } D}{\text{Total jumlah Kata dalam Dokumen } D} \quad (1)$$

Berikut rumus dari *Inverse Document Frequency* (IDF) adalah.

$$IDF = \log\left(\frac{\text{Total jumlah Dokumen dalam Korpus}}{\text{jumlah Dokumen Mengandung Term}}\right) + 1 \quad (2)$$

Maka, dapat dihitung dari perhitungan TF-IDF, berikut rumusnya terlampir.

$$TFIDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

Pada rumus tersebut, TF (*Term Frequency*) dilambangkan sebagai  $TF(t, d)$ , yang memperlihatkan frekuensi kemunculan suatu *term* “*t*” dalam dokumen “*d*”, dihitung berdasarkan perbandingan jumlah kemunculan term terhadap total kata dalam dokumen yang bersangkutan. Selanjutnya, IDF (*Inverse Document Frequency*) dilambangkan sebagai  $IDF(t, D)$ , di mana “*D*” merupakan keseluruhan dokumen dalam korpus, dan nilai ini diperoleh dari perbandingan antara jumlah total dokumen dengan jumlah dokumen yang mengandung term “*t*”, kemudian ditambahkan konstanta 1 untuk menghindari nilai nol. Sementara itu, TF-IDF dilambangkan sebagai  $TFIDF(t, d, D)$ , yang merupakan hasil perkalian antara nilai TF dan IDF, sehingga menghasilkan bobot suatu term yang mencerminkan taraf kepentingan pada sebuah dokumen relatif terhadap seluruh dokumen dalam korpus.

#### 2.1.4 Pemodelan Klasifikasi

Pada penelitian ini, metode yang digunakan adalah *Naïve Bayes* dan *SVM* untuk melakukan analisis sentimen terhadap ulasan pengguna pada *Google Maps* yang berkaitan dengan pusat perbelanjaan di Kota Semarang. Dipilihnya kedua algoritma tersebut dikarenakan menampilkan performa yang baik dalam klasifikasi teks serta kerap dipakai dalam kajian analisis sentimen.

*Support Vector Machine* (SVM) merupakan salah satu metode klasifikasi dengan menggunakan pendekatan *machine learning* (*supervised learning*) yang memprediksi kelas didasarkan pada pola dari hasil proses *training* [17]. SVM bekerja dengan mencari *hyperplane* optimal yang mampu memisahkan data antar kelas dengan margin terbesar. Metode ini mempunyai keefektifan yang tinggi biasa dipakai pada data berdimensi tinggi, seperti data teks hasil pembobotan TF-IDF, karena mampu menangani jumlah fitur yang besar serta menghasilkan model yang stabil dalam proses klasifikasi [18].

*Naïve Bayes* adalah metode klasifikasi berbasis probabilitas yang menggunakan prinsip Teorema Bayes dalam menetapkan peluang suatu data tergolong ke dalam sebuah kelas [19]. Salah satu varian yang umum digunakan adalah *Multinomial Naïve Bayes*, yang banyak diterapkan dalam klasifikasi teks dan *Natural Language Processing* (NLP), seperti analisis sentimen, deteksi spam, dan sistem rekomendasi. Kinerja algoritma ini berlangsung melalui penghitungan probabilitas kemunculan suatu kata dalam dokumen terhadap kelas tertentu menggunakan distribusi *multinomial*, sehingga dapat menentukan kelas dengan probabilitas tertinggi sebagai hasil prediksi [20].

#### 2.1.5 Evaluation

Tahap evaluasi dilakukan guna mengetahui performa model klasifikasi yang telah dibangun. Evaluasi model dilakukan dengan memakai *Confusion Matrix* yang berfungsi untuk mengukur tingkat akurasi model dalam mengklasifikasikan sentimen ulasan pengguna. *Confusion Matrix* adalah tabel yang menunjukkan perolehan prediksi model dibandingkan dengan data sebenarnya. Pada penelitian ini digunakan beberapa metrik evaluasi yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Selection

Pada penelitian ini, proses pengumpulan data dilakukan menggunakan teknik *web scraping* dengan memanfaatkan layanan *Apify* yang dijalankan melalui *platform Google Colaboratory*. Teknik *web scraping* dipakai untuk mengambil data ulasan pengguna dari *Google Maps* secara otomatis sehingga proses pengumpulan data dapat dilakukan secara lebih efisien dan sistematis. Objek penelitian yang dipakai dalam proses pengambilan data adalah lima pusat perbelanjaan yang berada di Kota Semarang, yaitu Paragon Mall, Mall Ciputra, Java Mall, DP Mall, dan Queen City Mall. Data yang diperoleh dari proses *scraping* berupa teks ulasan pengguna serta nilai *rating* yang diberikan oleh pengunjung terhadap masing-masing pusat perbelanjaan. Pada proses *scraping*, jumlah ulasan yang diambil dibatasi maksimal sebanyak 1.624 ulasan untuk setiap objek penelitian. Namun jumlah data yang diperoleh dapat berbeda pada setiap lokasi, tergantung pada jumlah ulasan yang tersedia dan dapat diakses oleh sistem *scraping*. Berdasarkan hasil proses *scraping*, diperoleh total *dataset* sebanyak 7.638 ulasan pengguna dari lima pusat perbelanjaan yang diteliti. Pembatasan jumlah data ini dilakukan untuk menyesuaikan kebutuhan *dataset* dalam penelitian serta menjaga efisiensi pada proses pengolahan data selanjutnya. Dengan adanya batasan tersebut, sistem hanya mengambil ulasan pengguna hingga jumlah maksimal yang telah ditentukan untuk setiap lokasi yang diteliti. Setelah proses pengumpulan data selesai dilakukan, tahap berikutnya adalah pemilihan atribut data yang relevan untuk digunakan dalam penelitian. Pada tahap ini dilakukan proses seleksi terhadap beberapa atribut utama yang digunakan dalam analisis, yaitu nama

tempat, nama pengguna, teks ulasan, dan nilai rating. Selain itu, dilakukan pula proses pembersihan data awal (*initial data cleaning*) dengan menghapus data ulasan yang kosong, tidak lengkap, atau tidak sesuai dengan kebutuhan penelitian.

Selanjutnya, *dataset* yang telah melalui proses seleksi dan pembersihan awal akan dilakukan proses pelabelan sentimen berdasarkan nilai *rating* yang diberikan oleh pengguna. Proses pelabelan ini ditujukan guna membagi ulasan pengguna menjadi tiga kategori sentimen, diantaranya positif, netral, dan negatif. Pelabelan secara otomatis memakai fungsi pemrograman *Python* berdasarkan rentang nilai *rating* yang diberikan oleh pengguna. Hasil pelabelan sentimen pada Tabel 1 berikut:

**Tabel 1.** Tabel *Labeling* Sentimen

Rating	Sentimen
4–5	Positif
3	Netral
1-2	Negatif

Berdasarkan proses pengumpulan data yang telah dilakukan, diperoleh *dataset* ulasan pengguna dari lima pusat perbelanjaan di Kota Semarang yang kemudian melalui tahap pembersihan data untuk memastikan kualitas *dataset* yang akan dipakai. Setelah dilakukan proses *cleaning* dalam rangka mengeliminasi data yang tidak valid atau kosong, jumlah *dataset* yang dipakai pada penelitian ini sebanyak 3.286 data ulasan. *Dataset* tersebut selanjutnya digunakan pada tahap analisis sentimen untuk mengetahui persepsi pengguna terhadap pusat perbelanjaan yang menjadi objek penelitian. *Dataset* yang digunakan pada penelitian ini dapat disimak pada Gambar 2 berikut.

Rating	Nama_Mall	Nama_User	Review	Label
5	Mall Ciputra	Sukma Nurna	Mall tengah kota semarang yang wajib dikunjungi	Positif
5	Mall Ciputra	Andreas Fibrianto	Tmpt jalan² mencuci mata sambil cari makan	Positif
4	Mall Ciputra	Lily Ryan	It's a pretty big mall with plenty of shops an...	Positif
4	Mall Ciputra	elfina vina	Easy to find parking, not busies as usual, but...	Positif
5	Mall Ciputra	88 cs	Niceee	Positif

**Gambar 2.** *Sample Dataset*

### 3.2 Preprocessing Data

Proses ini ditujukan guna menyederhanakan struktur data sehingga hanya atribut yang relevan saja yang digunakan pada tahap analisis selanjutnya. Dengan mengurangi atribut yang tidak diperlukan, proses pengolahan data menjadi lebih efisien serta dapat meminimalkan gangguan (*noise*) yang berpotensi mempengaruhi hasil pemodelan. Terdapat beberapa tahapan diantaranya sebagai berikut:

#### 3.2.1 Case Folding

Dilakukan konversi seluruh karakter dalam dokumen ke dalam format *lowercase*.

Review	case_folding
Mall tengah kota semarang yang wajib dikunjungi	mall tengah kota semarang yang wajib dikunjungi
Tmpt jalan² mencuci mata sambil cari makan	tmpt jalan² mencuci mata sambil cari makan
It's a pretty big mall with plenty of shops an...	it's a pretty big mall with plenty of shops an...
Easy to find parking, not busies as usual, but...	easy to find parking, not busies as usual, but...
Niceee	niceee

**Gambar 3.** Hasil *Case Folding*

#### 3.2.2 Cleansing

Dilakukan penghapusan atribut yang tidak berpengaruh terhadap klasifikasi. Seperti karakter simbol, angka, *emoticon*, tanda baca, lambang bendera, dan sebagainya akan dihilangkan atau dihapus serta juga spasi yang berlebihan, *URL*, *hashtag* dan lainnya.

case_folding	cleansing
mall tengah kota semarang yang wajib dikunjungi	mall tengah kota semarang yang wajib dikunjungi
tmpt jalan² mencuci mata sambil cari makan	tmpt jalan mencuci mata sambil cari makan
it's a pretty big mall with plenty of shops an...	it s a pretty big mall with plenty of shops an...
easy to find parking, not busies as usual, but...	easy to find parking not busies as usual but m...
niceee	niceee

**Gambar 4.** Hasil *Cleansing*

### 3.2.3 Tokenizing

*Tokenizing* merupakan prosedur dekomposisi kalimat menjadi unit-unit kata berdasarkan pembatas karakter tertentu. Tahapan ini memiliki peran krusial sebagai fondasi implementasi algoritma TF-IDF, mengingat proses T memerlukan pengelompokan dan kalkulasi frekuensi kemunculan kata secara presisi. Karakter yang umumnya digunakan sebagai separator dalam proses ini meliputi spasi, titik (.), koma (,), titik koma (;), serta titik dua (:)

cleansing	tokenizing
mall tengah kota semarang yang wajib dikunjungi	[mall, tengah, kota, semarang, yang, wajib, di...]
tmpt jalan mencuci mata sambil cari makan	[tmpt, jalan, mencuci, mata, sambil, cari, makan]
it s a pretty big mall with plenty of shops an...	[it, s, a, pretty, big, mall, with, plenty, of...]
easy to find parking not busies as usual but m...	[easy, to, find, parking, not, busies, as, usu...]
niceee	[niceee]

Gambar 5. Hasil *Tokenizing*

### 3.2.4 Text Normalization

Proses normalisasi dilakukan dengan mengoreksi kata yang mengandung unsur singkatan, tidak baku, dan *typo* menjadi bentuk baku menggunakan kamus kata tidak baku (*slang dictionary*) yang disusun secara manual. Penyusunan kamus dilakukan dengan terlebih dahulu mengidentifikasi kata-kata yang kerap muncul dalam *dataset* melalui perhitungan frekuensi kata, kemudian menganalisisnya untuk menentukan kata tidak baku, singkatan, dan *typo* yang selanjutnya dipetakan ke dalam bentuk baku. Berdasarkan hasil analisis tersebut, ditemukan beberapa kata tidak baku yang sering muncul seperti ‘yg’, ‘ga’, dan ‘aja’, yang kemudian dikonversi menjadi ‘yang’, ‘tidak’, dan ‘saja’. Selain itu, kata seperti ‘tmpt’ juga dinormalisasi menjadi ‘tempat’. Proses normalisasi dilakukan dengan memeriksa setiap token hasil *tokenizing* dan mengganti kata yang terdapat dalam kamus dengan bentuk baku yang sesuai.

tokenizing	normalized
[mall, tengah, kota, semarang, yang, wajib, di...]	[mall, tengah, kota, semarang, yang, wajib, di...]
[tmpt, jalan, mencuci, mata, sambil, cari, makan]	[tempat, jalan, mencuci, mata, sambil, cari, m...]
[it, s, a, pretty, big, mall, with, plenty, of...]	[it, s, a, pretty, big, mall, with, plenty, of...]
[easy, to, find, parking, not, busies, as, usu...]	[easy, to, find, parking, not, busies, as, usu...]
[niceee]	[niceee]

Gambar 6. Hasil *Text Normalization*

### 3.2.5 Stopword Removal

Tahapan ini ialah penghapusan kata-kata yang tidak memiliki makna penting atau tidak berkontribusi signifikan terhadap proses analisis. Kata-kata yang tergolong kelompok ini umumnya adalah kata hubung, kata depan, atau kata umum yang kerap muncul dalam kalimat, seperti “dan”, “di”, “yang”, “ke”, dan sejenisnya. Keberadaan kata-kata tersebut cenderung tidak membawa informasi khusus yang dapat membantu dalam proses klasifikasi sentimen.

normalized	stopword_removed
[mall, tengah, kota, semarang, yang, wajib, di...]	[mall, kota, semarang, wajib, dikunjungi]
[tempat, jalan, mencuci, mata, sambil, cari, m...]	[jalan, mencuci, mata, cari, makan]
[it, s, a, pretty, big, mall, with, plenty, of...]	[it, s, a, pretty, big, mall, with, plenty, of...]
[easy, to, find, parking, not, busies, as, usu...]	[easy, to, find, parking, not, busies, as, usu...]
[niceee]	[niceee]

Gambar 7. Hasil *Stopword Removal*

### 3.2.6 Stemming

Proses ini menghapus imbuhan awal atau akhiran lalu menjadikannya ke kata dasar yang ditujukan guna menyederhanakan variasi kata yang bermakna serupa. Misalnya, kata "mencuci" dapat diubah menjadi "cuci". *Stemming* sangat berguna untuk mengurangi kompleksitas data teks, sehingga model atau algoritma dapat lebih mudah memahami konteksnya.

	stopword_removed	stemming
0	[mall, kota, semarang, wajib, dikunjungi]	[mall, kota, semarang, wajib, dikunjungi]
1	[jalan, mencuci, mata, cari, makan]	[jalan, cuci, mata, cari, makan]
2	[it, s, a, pretty, big, mall, with, plenty, of...]	[it, s, a, pretty, big, mall, with, plenty, of...]
3	[easy, to, find, parking, not, busies, as, usu...]	[easy, to, find, parking, not, busies, as, usu...]
4	[niceee]	[niceee]

Gambar 8. Hasil *Stemming*



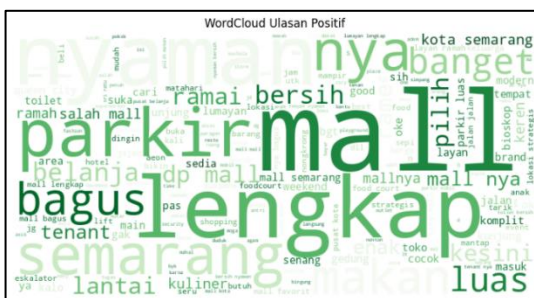
Pada Gambar 11, Secara keseluruhan menggambarkan adanya keluhan atau ketidakpuasan pengunjung terhadap beberapa aspek fasilitas dan kenyamanan saat berkunjung ke pusat perbelanjaan.



Gambar 12. Visualisasi *WordCloud* Sentimen Netral

Selanjutnya, pada Gambar 12. terlihat ulasan pada kategori netral umumnya berisi informasi atau deskripsi mengenai kondisi pusat perbelanjaan, fasilitas yang tersedia, serta aktivitas yang dilakukan pengunjung tanpa menunjukkan kecenderungan penilaian yang terlalu positif maupun negatif.

Mengacu visualisasi tersebut, dapat diambil kesimpulan bahwa ulasan dengan sentimen netral cenderung bersifat informatif dan deskriptif, sehingga memberikan gambaran umum mengenai kondisi pusat perbelanjaan tanpa menunjukkan penilaian yang ekstrem. Informasi ini tetap memiliki nilai penting karena dapat digunakan sebagai bahan pertimbangan tambahan dalam memahami persepsi pengunjung secara lebih menyeluruh.



Gambar 13. Visualisasi *WordCloud* Sentimen Positif

Sementara itu, pada Gambar 13, terlihat pengunjung memberikan ulasan positif terkait kenyamanan tempat, kelengkapan fasilitas, serta suasana pusat perbelanjaan yang dianggap menarik untuk dikunjungi.

### 3.4.2 Pembagian *Data Training* dan *Testing*

Selanjutnya, proses pembagian *dataset* menjadi *data training* dan *data testing* memakai fungsi *train\_test\_split*. Pembagian data ini dilakukan supaya model dapat dilatih menggunakan sebagian data dan diuji menggunakan data lainnya yang belum pernah dipelajari oleh model. Pada penelitian ini menggunakan rasio pembagian data sebesar 80% *data training* dan 20% *data testing* dengan parameter *random\_state* = 42 serta *stratify* untuk menjaga keseimbangan distribusi kelas sentimen. Berdasarkan hasil pembagian data diperoleh seperti Tabel 2 berikut:

Tabel 2. Jumlah Data

Skenario Rasio Perbandingan	Data Training	Data Testing
80:20	2628	658

Untuk memberikan gambaran yang lebih rinci terkait distribusi data pada tahap pengujian, dilakukan analisis jumlah data pada masing-masing kelas sentimen. Distribusi data testing berdasarkan kelas sentimen ditunjukkan melalui Tabel 3.

Tabel 3. Distribusi Data Testing

Sentimen	Jumlah Data
Negatif	77
Netral	23
Positif	558
Total	658

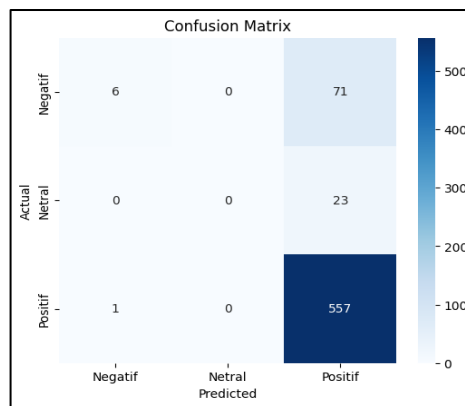
Berdasarkan Tabel 3, terlihat bahwa distribusi data tidak seimbang, dimana jumlah data sentimen positif jauh lebih dominan dibandingkan dengan sentimen negatif dan netral. Kondisi ini berpotensi mempengaruhi performa model dalam melakukan klasifikasi, khususnya pada kelas minoritas.

### 3.4.3 Proses Training Model SVM dan Naïve Bayes

Setelah *dataset* dibagi, proses selanjutnya adalah pelatihan model menggunakan algoritma SVM dengan metode *LinearSVC*. Algoritma ini digunakan karena dapat mengelola data teks dengan jumlah fitur yang besar serta memiliki performa yang baik dalam proses klasifikasi. Di tahapan ini model dilatih memakai *data training* yang telah diproses sebelumnya. Parameter *class\_weight = balanced* digunakan guna mengatasi kemungkinan ketidakseimbangan jumlah data pada masing-masing kelas sentimen. Setelah proses training selesai dilakukan, model digunakan untuk melakukan prediksi terhadap *data testing*.

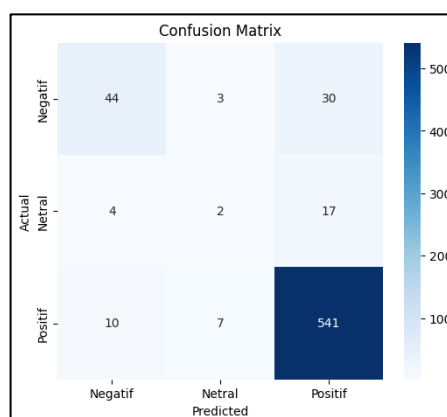
### 3.4.4 Confusion Matrix

Untuk mengetahui hasil klasifikasi yang dihasilkan oleh model, digunakan *Confusion Matrix*. *Confusion Matrix* dipakai guna melihat perbandingan antara label sebenarnya dengan perolehan prediksi model pada *data testing*. Melalui *Confusion Matrix* dapat diidentifikasi total data yang sukses diprediksi secara benar maupun data yang mengalami kesalahan klasifikasi pada setiap kategori sentimen.



Gambar 14. Confusion Matrix Naïve Bayes

Pada *Confusion Matrix* yang ditunjukkan pada Gambar 14 terlihat hasil klasifikasi model *Naïve Bayes* terhadap data uji yang dipakai dalam penelitian ini. Berdasarkan hasil tersebut, terdapat 6 data yang sebenarnya tergolong dalam kelas negatif dan sukses diprediksi dengan baik sebagai negatif. Namun, model juga melakukan kesalahan klasifikasi dengan memprediksi 71 data yang sebenarnya negatif sebagai positif. Selanjutnya, pada kelas sentimen netral terdapat 23 data yang sebenarnya netral, tetapi seluruh data tersebut diprediksi sebagai kelas positif oleh model. Hal ini mengindikasikan, model *Naïve Bayes* masih mengalami kesulitan dalam membedakan karakteristik ulasan yang memiliki sentimen netral dengan sentimen lainnya. Pada kelas sentimen positif, dari total 558 data, model berhasil mengklasifikasikan 557 data secara benar sebagai positif, sedangkan 1 data diprediksi sebagai negatif. Temuan tersebut mengindikasikan, model mempunyai kemampuan yang cukup baik dalam mengenali pola ulasan yang mengandung sentimen positif. Secara keseluruhan, dari total 658 data uji, hasil *Confusion Matrix* ini menyajikan gambaran mengenai performa model *Naïve Bayes* dalam melakukan klasifikasi sentimen. Meskipun model *Naïve Bayes* memperlihatkan tingkat akurasi yang cukup tinggi, terlihat bahwa model *Naïve Bayes* cenderung memprediksi sebagian besar data ke dalam kelas positif. Hal ini menunjukkan adanya ketidakseimbangan distribusi data yang mempengaruhi kemampuan model *Naïve Bayes* dalam mengenali sentimen negatif dan netral secara lebih akurat. Kondisi ini sejalan dengan distribusi data pada Tabel 3, dimana jumlah data sentimen positif jauh lebih besar dibandingkan kelas lainnya. Ketidakseimbangan ini menyebabkan model tendensi mengklasifikasikan data ke dalam kelas mayoritas, sehingga performa pada kelas negatif dan netral menjadi kurang optimal.



Gambar 15. Confusion Matrix SVM

Pada *Confusion Matrix* yang terlihat pada Gambar 15 ditampilkan hasil klasifikasi memakai metode SVM terhadap data uji. Berdasarkan hasil tersebut, sebagian besar data ulasan berhasil diklasifikasikan dengan benar sesuai dengan label aslinya. Model SVM mampu mengklasifikasikan sebagian besar data pada kelas sentimen positif dengan cukup baik, yang mengindikasikan, algoritma ini efektif dalam mengenali pola kata yang muncul pada ulasan dengan sentimen positif. Selain itu, pada kelas sentimen negatif, model juga sukses memprediksi beberapa data dengan benar, meskipun masih ditemukan sejumlah kekeliruan prediksi yang diklasifikasikan sebagai sentimen positif. Pada kelas sentimen netral, model SVM memperlihatkan performa yang lebih baik dibandingkan metode *Naïve Bayes*, meskipun masih terdapat sejumlah kekeliruan klasifikasi. Temuan tersebut mengindikasikan, metode SVM mempunyai kemampuan yang lebih baik dalam memisahkan data pada ruang fitur hasil pembobotan TF-IDF. Secara keseluruhan, *Confusion Matrix* ini memberikan Gambaran, metode SVM memperlihatkan performa klasifikasi yang lebih konsisten dan mampu menangani variasi data ulasan dengan lebih baik dibandingkan metode pada penelitian ini. Meskipun telah digunakan parameter class\_weight='balanced' untuk mengatasi ketidakseimbangan kelas, hasil klasifikasi masih menunjukkan adanya kecenderungan model terhadap kelas positif. Hal ini mengindikasikan, tingkat ketidakseimbangan data yang cukup tinggi masih mempengaruhi performa model, meskipun telah dilakukan penyesuaian bobot kelas.

### 3.5 Evaluation

Tahap evaluasi diterapkan dalam rangka mengetahui tingkat performa model klasifikasi yang sudah dirancang menggunakan matriks evaluasi berupa *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Berdasarkan Gambar 16, perolehan pengujian memakai metode SVM memperoleh nilai evaluasi model dengan tingkat akurasi yang lebih tinggi dan performa yang lebih stabil dalam mengklasifikasikan data ulasan pengguna dibandingkan dengan metode *Naïve Bayes*.

```

=====
MODEL: SUPPORT VECTOR MACHINE (SVM)
=====
Accuracy : 0.89209726443769
Precision : 0.8748405556541871
Recall : 0.89209726443769
F1 Score : 0.8809405172587271

Classification Report:
      precision    recall  f1-score   support

 Negatif      0.76      0.57      0.65        77
  Netral      0.17      0.09      0.11        23
  Positif      0.92      0.97      0.94       558

 accuracy          0.89          658
 macro avg      0.62      0.54      0.57          658
 weighted avg    0.87      0.89      0.88          658
    
```

Gambar 16. Nilai performa SVM

Sementara itu, hasil evaluasi *Naïve Bayes* pada Gambar 17 menunjukkan nilai *accuracy* sebesar 0.8556 atau sekitar 85.56%. Nilai *precision* 0.8258, *recall* 0.8556, serta *F1-score* 0.7981 mengindikasikan, model masih cukup baik dalam melakukan klasifikasi sentimen, terutama pada kelas positif. Namun demikian, berdasarkan hasil *classification report* dan *confusion matrix* terlihat bahwa performa model masih rendah pada kelas sentimen negatif dan netral.

```

=====
MODEL: NAIVE BAYES
=====
Accuracy : 0.8556231003039514
Precision : 0.8258792878853669
Recall : 0.8556231003039514
F1 Score : 0.798106149169979

Classification Report:
      precision    recall  f1-score   support

 Negatif      0.86      0.08      0.14        77
  Netral      0.00      0.00      0.00        23
  Positif      0.86      1.00      0.92       558

 accuracy          0.86          658
 macro avg      0.57      0.36      0.35          658
 weighted avg    0.83      0.86      0.80          658
    
```

Gambar 17. Nilai performa Naïve Bayes

Perbedaan performa antara kedua metode ini mengindikasikan bahwa algoritma SVM lebih efektif dalam konteks penanganan data teks dengan jumlah fitur yang besar hasil dari proses pembobotan TF-IDF. Selain itu, SVM

juga mempunyai kemampuan yang lebih baik dalam memisahkan data pada ruang fitur berdimensi tinggi dibandingkan metode *Naïve Bayes* yang berbasis probabilistik. Selain itu, ketidakseimbangan distribusi data pada masing-masing kelas sentimen juga menjadi faktor yang mempengaruhi hasil evaluasi model. Model cenderung memiliki performa yang lebih baik pada kelas dengan jumlah data yang lebih banyak, sehingga diperlukan penanganan lebih lanjut seperti teknik oversampling atau undersampling untuk meningkatkan performa klasifikasi pada kelas minoritas.

#### 4. KESIMPULAN

Mengacu temuan penelitian yang dilakukan, analisis sentimen terhadap ulasan pengguna *Google Maps* pada lima pusat perbelanjaan di Kota Semarang dapat dilakukan secara efektif menggunakan metode *machine learning*. Hasil pengujian mengindikasikan, metode *Support Vector Machine* (SVM) memperlihatkan performa yang lebih baik dibanding metode *Naïve Bayes* dalam melakukan klasifikasi sentimen. Hal ini ditunjukkan dari nilai evaluasi model yang lebih tinggi serta kemampuan SVM dalam mengklasifikasikan data secara lebih akurat. Sementara itu, metode *Naïve Bayes* cenderung menghasilkan prediksi yang bias terhadap kelas sentimen positif akibat distribusi data yang tidak seimbang, sehingga performanya pada kelas negatif dan netral kurang optimal. Temuan ini mengindikasikan, karakteristik data ulasan yang tidak seimbang sangat mempengaruhi kinerja model klasifikasi. Selain itu, hasil analisis *WordCloud* mengindikasikan, ulasan pengguna didominasi oleh pembahasan terkait fasilitas, kenyamanan, akses, dan area parkir, yang menjadi faktor utama dalam penilaian pengunjung terhadap pusat perbelanjaan. Akan tetapi, penelitian ini juga masih terbatas pada jumlah *dataset* yang relatif terbatas dan distribusi data yang tidak seimbang antar kelas sentimen. Itulah mengapa, penelitian selanjutnya direkomendasikan agar memakai *dataset* yang lebih besar serta mengimplementasikan teknik penanganan data tidak seimbang misalnya *oversampling* atau *undersampling*, serta mengembangkan metode berbasis *deep learning* untuk membuat akurasi klasifikasi sentimen lebih baik.

#### REFERENCES

- [1] R. T. Handayanto, Herlawati, P. D. Atika, F. N. Khasanah, A. Y. P. Yusuf, and D. Y. Saputra, "Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine," *J. Komtika (Komputasi dan Inform.,* vol. 5, no. 2, pp. 153–163, 2021, doi: <https://doi.org/10.31603/komtika.v5i2.6280> Received:
- [2] A. Maizaliyanti, K. Umam, W. D. Yuniarti, and M. R. Handayani, "Implementasi Algoritma Random Forest dalam Klasifikasi Ulasan Pengunjung Mall Semarang untuk Pengambilan Keputusan Layanan," *Edumatic J. Pendidik. Inform.,* vol. 9, no. 2, pp. 452–461, 2025, doi: 10.29408/edumatic.v9i2.30379.
- [3] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. Jambi: PT. Sonpedia Publising Indonesia, 2024.
- [4] Gishella Septania Al-Husna, Dian Asmarajati, Iman Ahmad Ihsannuddin, and Rina Mahmudati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi LinkedIn," *STORAGE J. Ilm. Tek. dan Ilmu Komput.,* vol. 3, no. 2, pp. 139–144, 2024, doi: 10.55123/storage.v3i2.3602.
- [5] S. Widodo and B. Hartono, "Analisis Sentimen Pengguna Google Terhadap Destinasi Wisata Di Kota Semarang Menggunakan Metode K-Nearest Neighbor," *Progresif J. Ilm. Komput.,* vol. 19, no. 2, p. 545, 2023, doi: 10.35889/progresif.v19i2.1364.
- [6] D. K. Sumartha, "Implementasi IndoBert Untuk Analisis Sentimen Opini Publik Terhadap Kebijakan Kenaikan UKT Di Era Pemerintahan," *J. Inform. Dan Tek. Elektro Terap.,* vol. 13, no. 3, p. 9, 2025, doi: [dx.doi.org/10.23960](https://doi.org/10.23960).
- [7] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.,* vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
- [8] C. H. P. Panjaitan and C. Supriadi, "Analisis Sentimen Pengunjung Wisata Heritage Kota Semarang Menggunakan Naive Bayes Pada Ulasan Google Maps," *J. Elektron. dan Teknol. Inf.,* vol. 4, no. 2, pp. 2721–9380, 2023, doi: <https://doi.org/10.5201/jet.v4i2.412>.
- [9] D. Rudini, D. G. Purnama, and A. A. Khan, "Penggunaan Teknik Web Scraping Dalam Aplikasi Pengambilan Data Dari Google Maps Untuk Menunjang Digital Marketing," *Lentera Multidiscip. Stud.,* vol. 2, no. 1, pp. 10–19, 2023, doi: <https://doi.org/10.57096/lentera.v2i1.61>.
- [10] A. E. Perkasa and A. N. Putri, "Penerapan Naïve Bayes Untuk Analisis Sentimen Pada Ulasan Aplikasi Mobile Legends," vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6507.
- [11] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [12] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.,* vol. 2, no. 5, pp. 178–185, 2022, doi: <https://doi.org/10.30865/klik.v2i5.362>.
- [13] D. Angraini and E. I. Wahyuni, "Analisis Sentimen Pengguna Terhadap Aplikasi Glints Di Google Play Store Menggunakan Metode Naive Bayes," *J. MEDIA Akad.,* vol. 3, no. 5, pp. 3031–5220, 2025, doi: <https://doi.org/10.62281/v3i5.1903>.
- [14] D. A. Tarigan, Z. Situmorang, and R. Rosnelly, "Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine (SVM), Naive Bayes Classifier Dan Random Forest," *J. Teknol. Inf. dan Ilmu Komput.,* vol. 11, no. 3, pp. 661–670, 2025, doi: <https://doi.org/10.25126/jtiik.2025129608>.
- [15] E. M. Windy Livia Azzahra, "Analisis Sentimen Terhadap RSUD Salatiga Menggunakan SVM Dan TF-IDF," vol. 6, no. 1, pp. 478–489, 2025, doi: <https://doi.org.10.35870/jimik.v6i1.1208>.
- [16] N. Wiliani, N. L. Chusna, and P. B. Ramadhan, *Analisis Sentimen terhadap Pro Kontra Aksi Unjuk Rasa Mahasiswa dengan Naive Bayes dan Information Gain*. Pekalongan: Penerbit NEM, 2023.
- [17] I. S. Aisah, B. Irawan, and T. Suprapti, "Algoritma Support Vector Machine (SVM) Untuk Analisis Sentimen Ulasan Aplikasi



Al Qur'an Digital," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 6, pp. 3759–3765, 2023, doi: <https://doi.org/10.36040/jati.v7i6.8263>.

[18] G. Maulani *et al.*, *Machine Learning*. Sumedang: CV. Mega Press Nusantara, 2025.

[19] D. Wijaya, R. A. Saputra, and F. Irwiensyah, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes," vol. 4, no. 4, 2024, doi: 10.30865/klik.v4i4.1738.

[20] H. Setyawan, M. Yusril, P. Herwanto, and T. Wiharko, *Machine Learning: Memahami Multinomial, Distribusi Probabilitas & Multinomial Naive Bayes*. Kupang: Tangguh Denara Jaya Publisher, 2025.