



## *Sentiment Analysis of Tourist Attractions Based on Reviews on Google Maps Using the Support Vector Machine Algorithm*

### **Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine**

**Joang Ipmawati<sup>1\*</sup>, Saifulloh<sup>2</sup>, Kusnawi<sup>3</sup>**

<sup>1</sup>Jurusan Teknik Komputer, Universitas Nahdlatul Ulama Yogyakarta, Indonesia

<sup>2</sup>Jurusan Teknik Informatika, Universitas PGRI Madiun, Indonesia

<sup>3</sup>Jurusan Informatika, Universitas Amikom Yogyakarta, Indonesia

E-Mail: <sup>1</sup>joang@unu-jogja.ac.id, <sup>2</sup>saifulloh@unipma.ac.id, <sup>3</sup>khusnawi@amikom.ac.id

*Received Nov 02nd 2023; Revised Dec 15th 2023; Accepted Jan 14th 2024*  
*Corresponding Author: Joang Ipmawati*

#### **Abstract**

*The Fourth Industrial Revolution era is characterized by the abundant availability of data, creating opportunities for information retrieval processes. One such data retrieval process involves searching for information about tourist destinations in Yogyakarta (DIY). This information retrieval can be carried out through Google Maps, which provides details such as location, distance, and even visitor reviews in the comments section, stemming from reviews of these tourist destinations. Within the collected information data, various issues arise that require identification, leading to research ideas for analyzing sentiments related to tourist destinations by utilizing user reviews on Google Maps. The research methodology employed in this study utilizes the Support Vector Machine (SVM) to categorize reviews into positive, negative, or neutral sentiment categories. User reviews from the Google Maps platform are processed and trained using SVM to identify sentiment patterns. The experimental results demonstrate the effectiveness of the SVM method in handling large volumes of review data for sentiment analysis, providing a deeper understanding of public perceptions regarding tourist destinations. This research can contribute to the development of marketing strategies and real-time user feedback-based management of tourist destinations. The study's findings regarding the performance of the SVM method in sentiment analysis classification using Support Vector Machine (SVM) indicate an average accuracy rate of 83.8% based on visitor reviews on the Google Maps website.*

*Keywords: DIY, Google Maps, Sentiment Analysis, SVM, Tourist Attractions*

#### **Abstrak**

Era Revolusi Industri 4.0 ditandai oleh ketersediaan data yang melimpah, menciptakan peluang dalam proses pengambilan informasi. Salah satu proses pengambilan data tersebut mencakup pencarian informasi tentang tempat wisata di Yogyakarta (DIY). Proses pengambilan informasi ini dapat dilakukan melalui Google Maps, yang menyediakan detail seperti lokasi, jarak, bahkan ulasan pengunjung dalam bagian komentar, yang berasal dari ulasan tentang destinasi wisata tersebut. Dalam data informasi yang dikumpulkan, muncul berbagai masalah yang memerlukan identifikasi, mengarah pada gagasan penelitian untuk menganalisis sentimen terkait destinasi wisata dengan memanfaatkan ulasan pengguna di Google Maps. Metodologi penelitian yang digunakan dalam studi ini menggunakan Support Vector Machine (SVM) untuk mengategorikan ulasan ke dalam kategori sentimen positif, negatif, atau netral. Ulasan pengguna dari platform Google Maps diolah dan dilatih menggunakan SVM untuk mengidentifikasi pola sentimen. Hasil eksperimen menunjukkan efektivitas metode SVM dalam mengelola volume besar data ulasan untuk analisis sentimen, memberikan pemahaman yang lebih dalam tentang persepsi masyarakat terhadap destinasi wisata. Penelitian ini dapat berkontribusi pada pengembangan strategi pemasaran dan manajemen berdasarkan umpan balik pengguna secara real-time. Temuan penelitian mengenai kinerja metode SVM dalam klasifikasi analisis sentimen menggunakan Support Vector Machine (SVM) menunjukkan tingkat akurasi rata-rata sebesar 83,8% berdasarkan ulasan pengunjung di situs Google Maps.

Kata Kunci: Analisis Sentiment, DIY, Google Maps, SVM, Tempat Wisata

## 1. PENDAHULUAN

Revolusi industri 4.0 ditandai dengan ketersediaan data yang berlimpah memungkinkan peluang melakukan proses mencari informasi dari data [1]. Data ini dihasilkan dengan pemanfaatan teknologi yang tersebar disekitar kita sehingga membentuk database raksasa. Dengan mencari informasi atau pola penting yang ada dalam database raksasa ini atau proses datamining [2], bisa dimanfaatkan oleh pengambil keputusan [3] untuk memperbaiki apa yang menjadi tanggungjawabnya. Tak terkecuali dengan sektor wisata, masyarakat bisa memutuskan untuk berkunjung atau tidak dapat melihat ulasan dari tempat itu pada situs seperti google maps.

Google Maps dapat mencari dan menampilkan informasi mengenai suatu tempat. Salah satu informasi yang dapat ditampilkan adalah ulasan pengunjung yang bisa menjadi rekomendasi calon pengunjung lainnya. Menurut [4], 78% pengunjung lebih mempercayai ulasan online daripada rekomendasi seseorang. Sehingga, dapat dikatakan bahwa ulasan pengunjung pada berbagai platform merupakan hal yang cukup penting untuk dipertimbangkan bagi pihak pengelola tempat wisata. Karenanya diperlukan analisa terkait ulasan ini menggunakan analisis sentimen.

Dalam penelitian ini, analisis sentimen diterapkan untuk mengetahui jenis ulasan pengunjung apakah positif atau negatif dan aspek yang dibahas pada ulasan tersebut seperti kebersihan, daya tarik, fasilitas, dan pelayanan. Terdapat beberapa penelitian yang relevan untuk menunjang penelitian ini diantaranya yang dilakukan mubarak, dkk (2023) dengan judul Analisis User Sentiment Aplikasi Google Maps, Maps ME dan Waze menggunakan Metode Support Vector Machine menghasilkan data analisis ulasan pada ketiga platform digital Dimana Maps.ME memperoleh nilai tertinggi berdasarkan komentar positif [5]. Adapun penelitian lainnya untuk penggunaan Metode Support Vector Machine dengan studi kasus untuk mengetahui Perguruan Tinggi termewah menurut Goggle Maps Erfina dan Wardani (2022) menghasilkan Tingkat akurasi analisis dari kelima Perguruan Tinggi sebagai objek penelitian dan penilaian berdasarkan hasil ulasan Google Maps diperoleh hasil analisis menggunakan metode SVM adalah Universitas Ciputra 85% [6]. Sedangkan, Terdapat penelitian serupa dalam pembahasan pemilihan tempat wisata yakni menurut syahlan, dkk (2023) penggunaan SVM sebagai pengumpulan ulasan pengguna berdasarkan tiga kategori sentiment (positif), Negatif dan Netral untuk menganalisis tanggapan Masyarakat mengenai wisata air mancur Sri Baduga Purwakarta [7].

Pada penelitian ini algoritma yang dipilih adalah SVM karena cocok untuk diterapkan dengan pembobotan TF-IDF yang memberi bobot terhadap setiap kata pada suatu kalimat yang menjadi feature pada SVM. Selain itu, menurut [8] SVM memiliki akurasi yang relatif tinggi. Berdasarkan ulasan diatas tujuan penelitian ini untuk mengetahui performa dan akurasi dari penggunaan algoritma SVM terhadap data Google Maps Review. Sehingga penelitian ini diharapkan dapat menjadi suatu model dalam analisis sentimen tempat wisata bagi pengelola tempat wisata dan dikembangkan dalam penelitian lain.

## 2. BAHAN DAN METODE

### 2.1 Sentiment Analisis

Pada era saat ini data mining bermain peran yang sangat peting dalam perkembangan industry terutama pada area text mining [9]. Data mining adalah sebuah proses untuk mengekstrak sebuah informasi dari data data untuk mengetahui pola pada sebuah kumpulan data yang besar. Dengan ditemukannya informasi dan pola dalam kumpulan data tersebut maka sebuah *trend*, *pattern*, *relationship* dan *route* akan terbantu dan kemudian diolah untuk memenuhi kebutuhan perusahaan.

Seiring perkembangan zaman dengan makin maraknya peneliti dibidang artificial intelegence dan data science teknologi text mining menjadi perhatian tersendiri dimata para peneliti. Text mining memfokuskan kaidah text analisis dan computer science untuk menghandle masalah yang kompleks [10]. Text mining dapat menemukan dan menerima sebuah informasi dalam sebuah kumpulan data atau text corpus. Text mining mengkombinasikan *artificial intelligence*, information retrieval dan data mining dalam memahami *analytical processing systems*.

Perkembangan e-commerce menuntut semua perusahaan harus mengembangkan usahanya. Dengan menggunakan sentiment analis perusahaan akan mendapatkan kemampuan untuk melihat target pasar mereka dengan ini sentiment analis akan memperkuat product dan service perusahaan terkait. Sentiment analisis merupakan sebuah bidang yang berkaitan dengan language processing dan dengan adanya sentiment analis ini pihak perusahaan maupun industri dapat merancang dan mengidentifikasi kepuasan pelanggan. Sebuah aspek dalam sentiment analisis merujuk pada identifikasi sebuah ekspresi untuk mengekstrak fine-grained information [11].

Penelitian [12] menerapkan gabungan teknik *Lexicon* dan Machine Learning dalam penyelesaian masalah analisis sentimen. Teknik *Lexicon* diterapkan pada bagian pelabelan dokumen dan Machine Learning dalam proses klasifikasi sentimen. Dengan menerapkan kedua teknik tersebut akan memudahkan proses analisis sentimen terhadap suatu kasus.

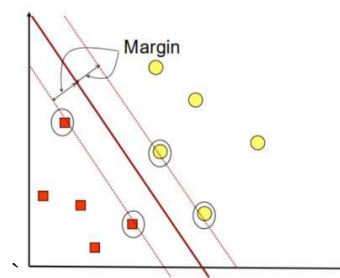
Penelitian [13] menggunakan Word2Vec dalam proses ekstraksi fitur atau representasi vektor dari tiap kata. Hasilnya menunjukkan bahwa ekstraksi fitur menggunakan Word2Vec memiliki akurasi terendah dibandingkan dengan model lainnya, yang mana Word2Vec menghasilkan akurasi 70% sedangkan model

lainnya lebih dari 80%. Hal tersebut dikarenakan Word2Vec memerlukan lebih banyak data agar mencapai hasil yang lebih optimal.

Proses *stopword removal* dan *stemming* merupakan salah satu proses dalam *text preprocessing* sebelum ekstraksi fitur dalam penyelesaian masalah analisis sentimen. Penelitian [14] membandingkan pendekatan yang berbeda dalam proses tersebut, seperti stop – stem, no stop – stem, stop – no stem, dan no stop – no stem. Hasilnya menunjukkan bahwa proses *stopword removal* dan *stemming* tidak berpengaruh banyak terhadap akurasi yang dihasilkan setelah proses klasifikasi dilakukan. Masalah analisis sentimen dapat diselesaikan menggunakan beberapa algoritma. Penelitian [15] menggunakan berbagai algoritma dan masing-masing algoritma tersebut menghasilkan akurasi yang relatif tinggi. SVM menghasilkan akurasi tertinggi yaitu sebesar 97.72% dibandingkan dengan algoritma lainnya.

## 2.2 Support Vector Machine (SVM)

SVM merupakan algoritma machine learning dengan tipe supervised berbasis vektor. SVM mengklasifikasikan data dengan membaginya menjadi dua kelas melalui hyperplane. Hyperplane berada di antara dua kelas dengan jarak  $d$  pada titik terdekat setiap kelas. Jarak  $d$  ini disebut dengan margin dan titik yang berada pada margin ini disebut sebagai support vector. Tujuan dari SVM adalah menentukan hyperplane terbaik yang dapat memberikan jarak terjauh dari suatu titik [16]. Representasi SVM dinyatakan gambar 1.



Gambar 1. Representasi SVM

Jika data dapat dipisahkan dengan sempurna oleh sebuah hyperplane linear, SVM disebut sebagai SVM linier. Namun, jika data tidak dapat dipisahkan dengan sempurna secara linear, SVM menggunakan teknik transformasi kernel untuk mengubah data ke dalam dimensi fitur yang lebih tinggi, dimana hyperplane linier dapat dibuat.

Dengan menggunakan berbagai jenis fungsi kernel, seperti kernel linier, kernel polinomial atau kernel Gauss, SVM dapat mengatasi data yang memiliki karakteristik non-linier dengan efisien. Fungsi kernel adalah fungsi yang mengubah data ke dimensi yang lebih tinggi dengan tujuan meningkatkan struktur data sehingga mempermudah proses pemisahan [17]. Dalam hal ini, rumus umum untuk SVM linear dapat dituliskan pada persamaan 1.

$$F(x) = \text{sign}(w \cdot x + b) \tag{1}$$

Di mana,  $f(x)$  adalah fungsi prediksi,  $w$  adalah vektor normal hyperplane,  $x$  adalah vektor fitur input, dan  $b$  adalah bias atau intercept.

## 2.3 Ekstraksi Fitur

Proses ini akan mengubah setiap kata berbentuk token yang sudah melalui tahap preprocessing menjadi vektor yang akan merepresentasikan kata yang ada. Term Frequency (TF) merupakan frekuensi kata yang muncul pada sebuah dokumen. Inverse Document Frequency (IDF) mengukur seberapa pentingnya kata pada sebuah dokumen [18]. Adapun persamaan dalam penghitungan TF-IDF ditunjukkan pada persamaan 2.

Bobot kata  $i$  pada dokumen  $j$

$$W_{ij} = tf_{ij} \times \log\left(\frac{N}{df_i}\right) \tag{2}$$

Keterangan:

- $tf_{ij}$  : jumlah kata  $i$  muncul dalam dokumen  $j$
- $df_i$  : jumlah dokumen yang mengandung  $i$
- $N$  : total seluruh dokumen

## 2.4 Confusion Matrix

*Confusion Matrix* merupakan salah satu metode dalam evaluasi performa suatu model dalam bentuk tabel yang menampilkan hasil prediksi terhadap dua kelas dapat dilihat pada tabel 1 [19].

**Tabel 1.** Confusion Matrix

Kelas Prediksi	Kelas Asli		
	Kelas-1	True Positive (TP)	False Negative (FN)
	Kelas-2	False Positive (FP)	True Negative (TN)

Adapun fungsi dari confusion matrix ini adalah mengukur tingkat akurasi dari suatu model dengan persamaan, ditunjukkan pada persamaan 3.

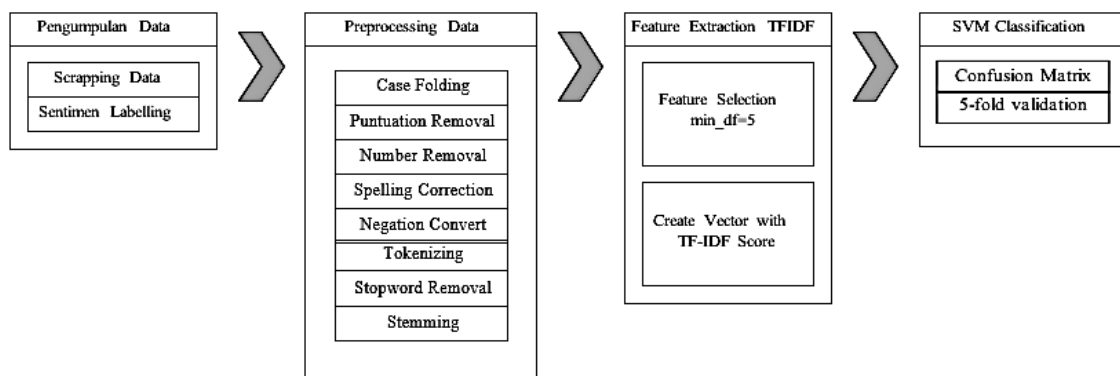
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

## 2.5 K-Fold Cross Validation

Evaluasi yang lebih mendalam dapat dilakukan dengan *k-fold cross-validation*. Metode ini akan membagi dataset secara acak sebanyak “k” yang ditentukan secara sama besar [20]. Metode ini menguji stabilitas model yang digunakan terhadap suatu data dan menghasilkan rata-rata akurasi. Pada penelitian ini, nilai “k” yang digunakan adalah 10.

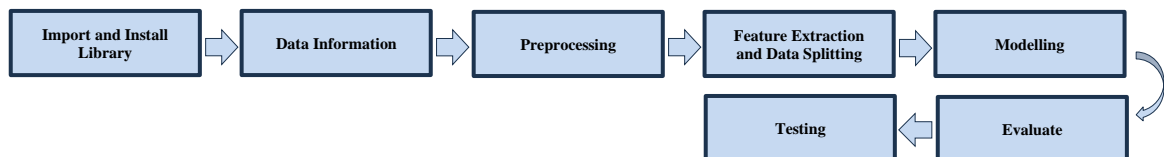
## 3. IMPLEMENTASI DAN PEMBAHASAN

Penelitian ini terdiri atas empat tahapan, dimulai dari tahap pengumpulan data, text preprocessing, ekstraksi fitur, dan klasifikasi menggunakan SVM [21]. Adapun alur penelitian ditunjukkan pada Gambar 2.



**Gambar 2.** Diagram Alur Penelitian

Berdasarkan alur penelitian pada gambar 2, selanjutnya dalam implementasiannya Dimana penelitian ini menggunakan program dalam Google Colab yang telah dibuat seperti pada gambar 3.



**Gambar 3.** Tahapan Program

### 3.1 Import and Install Library

Dalam tahapan ini dilakukan *importing library* dengan keterangan *library* seperti: Pandas, String, Sklearn, Matplotlib, cross\_val\_score, nltk, re, sastrawi. Penggunaan *library* memungkinkan proses pengodingan dan penelitian akan menjadi lebih mudah.

### 3.2 Data Information (Read Dataset)

Dalam tahapan ini dataset yang telah terbaca oleh program akan di *import* dan kemudian ditampilkan dalam tampilan seperti pada Tabel 2.

**Tabel 2.** Dataset

	Text	Value
0	Banyak resto seafood/boga Bahari, tempat bagus...	POSITIF
1	Pantai kotor	NEGATIF
2	Enak buat makan seafood	POSITIF
3	Tempatnya bagus, hanya saja untuk saat ini ma...	POSITIF
4	Aku benci masa 1 orang bayarnya 90.000 behhh...	NEGATIF

### 3.2.1 General Information

Dalam tahapan ini program akan menampilkan informasi dataset yang akan diolah. Program akan memberikan informasi tentang jumlah kolom dan baris dalam dataset tersebut. Dalam tahapan ini juga diberikan detail info dari dataset dengan diberikan keterangan index, column, Non-Null Count, Dtype dapat dilihat pada gambar 4 dan Dataset juga akan ditampilkan dengan menggunakan tabel yang berisi jumlah baris, jumlah baris unik, kata teratas dalam dataset, dan frekuensi pada dataset (tabel 3).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2434 entries, 0 to 2433
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0   text    2434 non-null    object
1   value   2434 non-null    object
dtypes: object(2)
memory usage: 38.2+ KB
```

Gambar 4. Data Information

Tabel 3. Data Describe

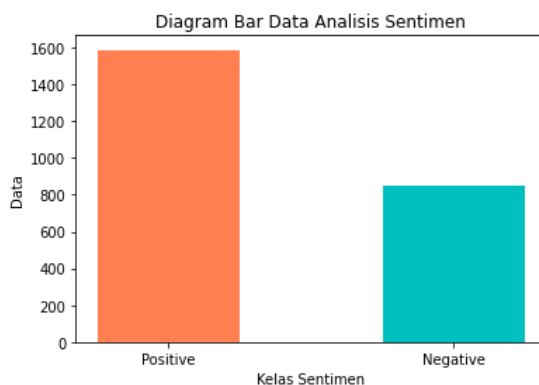
	Text	Value
Count	2434	2434
Unique	2443	2
Top	Bagus	POSITIF
Freq	12	1588

### 3.2.2 Check Missing Values

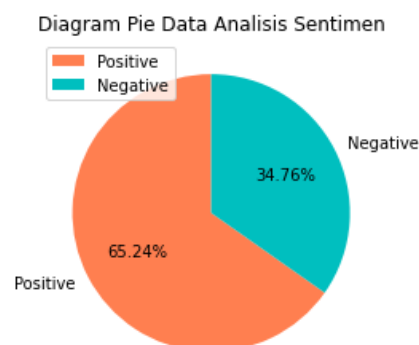
Dalam tahapan ini program akan menampilkan data yang bersifat null dalam dataset.

### 3.2.3 Visualize Data

Dataset yang telah dimasukkan kedalam program akan diolah dan dijadikan diagram batang untuk mengukur tingkat polaritas dataset (gambar 5) dan juga Program juga akan menampilkan diagram pie dengan menampilkan perbandingan dalam bentuk persen sebagai kondisi polaritas dataset (gambar 6).



Gambar 5. Visualisasi Data Diagram Batang



Gambar 6. Diagram Pie

Penelitian ini menggunakan data ulasan pada berbagai tempat wisata di DIY seperti Candi Prambanan, Gembira Loka, Hartono Mall, Jogja Bay, Pantai Depok, Pantai Parangtritis, dan Tebing Breksi dari Google Maps Review. Data dari tempat- tempat tersebut di-scraping menggunakan Apify kemudian digabungkan.

Keseluruhan data berjumlah 2434 data yang kemudian diberi label sentimen sebagai ulasan positif dan negatif berdasarkan bintang masing-masing ulasan yang kemudian dicek ulang secara manual. Data yang sudah

diberi label menunjukkan ulasan positif berjumlah 1588 data dan ulasan negatif berjumlah 846. Jumlah data tersebut menandakan bahwa data yang digunakan imbalance atau tidak seimbang.

**Tabel 4.** Contoh Dataset

Text	Value
Banyak resto seafood/boga bahari, tempat bagus, dan cocok untuk menikmati senja bareng keluarga atau kekasih	Positif
Pantai kotor	Negatif
Enak buat makan seafood	Positif
Tempat nya bagus, hanya saja untuk saat ini masih banyak wahana yang masih tutup. Karna efek pandemi.	Positif
Aku benci masa 1 orang bayarnya mahal bankrut dah!!	Negatif

### 3.3 Preprocessing Data

Tahapan pada Text Preprocessing yang akan dilakukan pada penelitian ini meliputi casefolding (mengubah huruf kapital menjadi huruf kecil), punctuation removal (menghilangkan tanda baca dan karakter selain alfabet), number removal (menghilangkan angka), tokenizing (memisahkan setiap kata pada kalimat menjadi token), stopword removal (menghilangkan kata yang tidak dibutuhkan), dan stemming (mengubah kata berimbuhan menjadi kata dasar).

**Tabel 5.** Hasil Text Preprocessing

Text	Value
Teks Awal	Aku benci masa 1 orang bayarnya mahal bankrut dah!!
Case Folding	aku benci masa 1 orang bayarnya mahal bankrut dah!!
Punctuation Removal	aku benci masa 1 orang bayarnya mahal bankrut dah
Number Removal	aku benci masa orang bayarnya mahal bankrut dah
Tokenizing	[aku, benci, masa, orang, bayarnya, mahal, bankrut, dah]
Stopword Removal	[benci, orang, bayarnya, mahal, bankrut]
Stemming	[benci, orang, bayar, mahal, bankrut]

#### 3.3.1 Create Preprocessing Function

Pada tahapan ini dataset akan melalui beberapa process dengan *detail casefolding, tokenizing, stopword, dan stemming*.

#### 3.3.2 Start Preprocessing

Case folding adalah proses penghilangan karakter selain *alfabet*. Dengan menggunakan proses ini dataset yang memiliki karakter unik didalamnya akan dihilangkan atau dihapus. Data pasca *case folding* dapat dilihat pada tabel 6.

**Tabel 6.** Data Pasca Case Folding

	Text	Value
0	Banyak resto seafood/boga Bahari, tempat bagus d...	POSITIF
1	Pantai kotor	NEGATIF
2	Enak buat makan seafood	POSITIF
3	Tempat nya bagus hanya saja untuk saat ini mas...	POSITIF
4	Aku benci masa orang bayarnya behhhhh bankr...	NEGATIF

Hasil *case folding*:

1. Tokenizing adalah sebuah proses penandaan setiap kata menggunakan token. Setiap kata yang telah ditandai dengan token akan dipisah menggunakan “,” (koma).
2. Stopword removal merujuk referensi library untuk menghilangkan kata yang tidak memiliki makna.
3. Proses stemming adalah proses terakhir dalam melakukan preprocessing penelitian ini. Merubah semua kata kata berimbuhan pada setiap baris dataset menjadi kata dasar. Proses ini mengubah sebuah baris dalam dataset menjadi point penting percakapan.
4. Dataset yang telah melalui semua proses diatas akan dikumpulkan dan dijadikan sebuah dataset baru.

### 3.4 Feature Extraction and Data Splitting

*Feature extraction* dilakukan dengan mengubah setiap text menjadi sebuah vector dan angka yang mudah dikenali dengan mesin.

### 3.4.1 Weighting (Vectorize)

Feature extraction ini dilakukan menggunakan pembobotan yang dilakukan oleh TF-IDF. Hasil data setelah dilakukan tahapan weighting pada tabel 7.

**Tabel 7.** Hasil Pembobotan TF-IDF

Vectorize	Nilai Pembobotan
(0, 2123)	0.30436768740017517
(0, 2196)	0.2220377133298921
(0, 430)	0.3193650200786121
(0, 4369)	0.34738587228546236
(0, 3290)	0.23206673845789558
(0, 926)	0.2251867973034359
(0, 319)	0.17182628325932367
(0, 338)	0.4343346547292049
(0, 4238)	0.4572207016418224
(0, 4019)	0.32187405255317275
(1, 2423)	0.8561271948482095
(1, 3542)	0.5167651557925863
(2, 4236)	0.6200548364229839
(2, 2772)	0.4605598241403215
(2, 1312)	0.6351508861808373
(3, 3528)	0.35405211155738725
(3, 1278)	0.5541885792540818
...	...
(2423, 2423)	0.18176853802525525
(2423, 3542)	0.1097169292554311

### 3.4.2 Data Split

Data splitting dilakukan untuk memisahkan data menjadi dua bagian. Pada kolom x terdapat data training dan data test dan pada kolom y terdapat data training dan data test. Dalam melakukan test data yang digunakan dalam tabel bisa ditentukan melalui test\_size yaitu menggunakan sample data untuk pengujian keseluruhan dataset.

### 3.5 Modelling

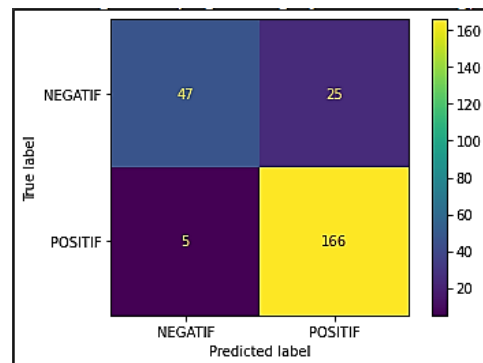
Text classification akan diproses dalam tahap ini. Dengan menggunakan metode Support Vector Machine dataset yang telah melalui semua proses sebelumnya akan dilatih dan dilakukan pengujian dengan kernel 'Linear' pada Probabilitas = True. Program juga akan memprediksi semua data dalam pengujian dengan dibandingkan pada dataset yang telah di fit-kan pada data split.

### 3.6 Evaluate

Ditahap ini semua proses pada program akan ditampilkan dengan menarik semua proses pada SVM untuk ditampilkan akurasi, presisi, recall, dan f1 score. Hasil Evaluasi confusion matrix ditunjukkan dalam tabel 8. Pada gambar 7 merupakan hasil dai confusion matrix diperolej tampilan True Negative sebanyak 47, True Positive sebanyak 166, False Positive sebanyak 25 dan False Negative sebanyak 5.

**Tabel 8.** Confusion Matrix SVM

Parameter	Nilai
Precision	90%
Recall	65%
Accuracy	87%
f1 score	75%



**Gambar 7.** Confusion Matrix

### 3.6.1 WordCloud

Wordcloud atau bisa disebut sebagai gugus kata akan menampilkan kata kata yang paling sering muncul dalam pengujian dataset baik dengan pengujian data positive maupun negative. Dengan ini bisa disimpulkan

bahwa kata “pantai”, ”Bagus”, “makan” menjadi kata kata yang sering muncul dalam pengujian data positive (gambar 8) sedangkan dalam pengujian data negative (gambar 9) menampilkan “parkir”,”pantai”, ”mahal”,”harga”, “masuk”.



Gambar 8. Wordcloud Positif



Gambar 9. Wordcloud Negatif

### 3.6.2 Cross Validation

Dalam melakukan sebuah penelitian maka akan ada sebuah validasi dalam penelitian tersebut. Dengan menggunakan *cross validation* yang disetting pada volume sebanyak 10 atau melakukan pengujian model sebanyak 10 kali dataset yang diacak. Penelitian ini menunjukkan bahwa nilai rata – rata *cross validation* berada pada 84% dapat dilihat pada tabel 9.

Tabel 9. Cross Validation

K uji model	Nilai
uji 1	88%
uji 2	86%
uji 3	81%
uji 4	86%
uji 5	81%
uji 6	84%
uji 7	81%
uji 8	83%
uji 9	82%
uji 10	84%
Rata-rata Cross Validation	84%

### 3.7 Testing

Dalam tahap ini dilakukan percobaan dengan memasukkan kata atau kalimat dengan tujuan menguji model penelitian ini. Dalam bagian testing ini semua proses mulai dari *preprocessing* hingga pengujian sample akan dilakukan dengan cara mengimport proses tersebut. Dengan menggunakan sample kata “tampanya bersih dan nyaman” kemudian dimasukkan dalam proses testing akan menampilkan hasil ‘Positive’ dengan Probabilitas 86% yang berarti program memiliki keyakinan sebesar 86% bahwa “bersih dan nyaman” adalah kalimat *positive*.



#### 4. KESIMPULAN

Dalam penelitian ini, metode SVM digunakan untuk menganalisis sentimen ulasan dari pengunjung tempat wisata di Yogyakarta yang diperoleh melalui Google Maps. Hasil analisis tersebut menunjukkan tingkat akurasi sebesar 87%, presisi sebesar 90%, recall sebesar 65%, dan nilai f1 sebesar 75% terhadap seluruh dataset dengan pengaturan random state 10. Secara rata-rata, akurasi keseluruhan mencapai 83,8%. Hasil pengujian yang dilakukan dari tahap preprocessing hingga uji sampel menggunakan kata kunci "Tempat Bersih dan Nyaman" mengindikasikan bahwa ulasan tersebut dikategorikan sebagai "Positif" dengan tingkat probabilitas sebesar 86%. Oleh karena itu, program ini memiliki keyakinan bahwa kalimat tersebut bersifat "Positif".

#### REFERENSI

- [1] K. Kusnawi, M. A. F. E. Putra, and J. Ipmawati, "Price Prediction Of Basic Material Using ARIMA Forecasting Method Through Open Data Sumedang District," *SISTEMASI*, vol. 12, no. 2, pp. 293–307, May 2023, Accessed: Dec. 21, 2023. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/2282>
- [2] E. Virantika, K. Kusnawi, and J. Ipmawati, "Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, pp. 1657–1666, Jul. 2022, doi: 10.30865/MIB.V6I3.4325.
- [3] "Prediction of Student Satisfaction with Academic Services Using The C4.5 Algorithm (Case Study: Yogyakarta Nahdlatul Ulama University) | Ipmawati | Prosiding Seminar Sains Nasional dan Teknologi." Accessed: Dec. 21, 2023. [Online]. Available: [https://publikasiilmiah.unwahas.ac.id/index.php/PROSIDING\\_SNST\\_FT/article/view/7214](https://publikasiilmiah.unwahas.ac.id/index.php/PROSIDING_SNST_FT/article/view/7214)
- [4] "Local Consumer Review Survey 2023: Customer Reviews and Behavior." Accessed: Dec. 21, 2023. [Online]. Available: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- [5] I. F. A. Mubarak, B. Huda, A. Hananto, T. Tukino, and H. Kabir, "Analisis User Sentiment Aplikasi Google Maps, Maps.Me Dan Waze Menggunakan Metode Support Vector Machine," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 1, pp. 69–74, 2023, doi: 10.36341/rabit.v8i1.3020.
- [6] N. R. W. Adhithia Erfina, "Analisis Sentimen Perguruan Tinggi Termewah Di Indonesia Menurut Ulasan Google Maps Menggunakan Algoritma Support Vector Machine (SVM)," *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 77–85, 2022.
- [7] M. S. Syahlan, D. Irmayanti, and S. Alam, "Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Support Vector Machine (Svm)," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 2, pp. 315–319, 2023, doi: 10.51876/simtek.v8i2.281.
- [8] A. Sasmito Aribowo, H. Basiron, N. Fazilla, A. Yusof, and S. Khomsah, "Cross- domain sentiment analysis model on Indonesian YouTube comment," *Int. J. Adv. Intell. Informatics*, vol. 7, no. 1, pp. 12–25, 2021, doi: 10.26555/ijain.v7i1.554.
- [9] F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab, and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," *IEEE Access*, vol. 9, pp. 43711–43735, 2021, doi: 10.1109/ACCESS.2021.3059343.
- [10] P. Wang et al., "Classification of Proactive Personality: Text Mining Based on Weibo Text and Short-Answer Questions Text," *IEEE Access*, vol. 8, pp. 97370–97382, 2020, doi: 10.1109/ACCESS.2020.2995905.
- [11] N. Zhao, H. Gao, X. Wen, and H. Li, "Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis," *IEEE Access*, vol. 9, pp. 15561–15569, 2021, doi: 10.1109/ACCESS.2021.3052937.
- [12] C. V D, "Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 5, pp. 4452–4459, 2019, doi: 10.11591/ijece.v9i5.pp4452-4459.
- [13] M. A. Fauzi, "Word2Vec model for sentiment analysis of product reviews in Indonesian," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, pp. 525–530, 2019, doi: 10.11591/ijece.v9i1.pp525-530.
- [14] A. W. Pradana and M. Hayaty, "The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 4, pp. 375–380, 2019, doi: 10.22219/kinetik.v4i4.912.
- [15] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of Indonesia's capital city," *Bull. Electr. Eng. Informatics*, vol. 9, no. 4, pp. 1620–1630, 2020, doi: 10.11591/eei.v9i4.2352.
- [16] Y. Al Amrani, M. Lazaar, and K. E. El Kadirp, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," *Procedia Comput. Sci.*, vol. 127, pp. 511–520, Jan. 2018, doi: 10.1016/J.PROCS.2018.01.150.

- 
- [17] G. Li and J. Li, "Research on Sentiment Classification for Tang Poetry based on TF-IDF and FP-Growth," Proc. 2018 IEEE 3rd Adv. Inf. Technol. Electron. Autom. Control Conf. IAEAC 2018, pp. 630–634, Dec. 2018, doi: 10.1109/IAEAC.2018.8577715.
- [18] F. Rozi, F. Sukmana, and M. N. Adani, "Pengelompokan Judul Buku dengan Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Term Frequency – Inverse Document Frequency (TF-IDF)," *JIMP J. Inform. Merdeka Pasuruan*, vol. 6, no. 3, pp. 1–5, 2021.
- [19] P. R. Sihombing and A. M. Arsani, "Comparison of Machine Learning Methods in Classifying Poverty in Indonesia in 2018," *J. Tek. Inform.*, vol. 2, no. 1, pp. 51–56, 2021, doi: 10.20884/1.jutif.2021.2.1.52.
- [20] A. Peryanto, A. Yudhana, and R. Umar, "Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 45–51, 2020, doi: 10.30871/jaic.v4i1.2017.
- [21] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, p. 86, 2020, doi: 10.33365/jtk.v14i2.792.