



Analisis Sentimen Ulasan Wisata Alun-Alun Brebes pada Google Maps Menggunakan Support Vector Machine

Azkiyatul Maulida*, Bambang Irawan, Nur Ariesanto Ramdhan

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Muhadi Setiabudi, Brebes, Indonesia

Email: ^{1*}azkiyatulmaulida2@gmail.com, ²bambangumus@gmail.com, ³ariesantoramdhan@gmail.com

Email Penulis Korespondensi: azkiyatulmaulida2@gmail.com

Abstrak—Perkembangan teknologi informasi mendorong pemanfaatan platform digital sebagai media berbagi opini, termasuk ulasan tempat wisata pada *Google Maps*. Alun-Alun Brebes sebagai salah satu ruang publik yang ramai dikunjungi menghasilkan ribuan ulasan dengan karakteristik teks yang beragam, sehingga sulit dianalisis secara manual. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengunjung Alun-Alun Brebes menggunakan pendekatan *text mining* dengan algoritma *Support Vector Machine*. Data yang digunakan terdiri dari 1.000 ulasan *Google Maps*, dengan 327 data berlabel sentimen positif dan negatif secara manual serta 673 data tanpa label. Tahapan penelitian meliputi pengumpulan data, *preprocessing* teks, ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*, pelatihan dan pengujian model *Support Vector Machine*, serta pelabelan otomatis data tanpa label. Evaluasi performa model dilakukan menggunakan *confusion matrix* dengan metrik akurasi, *precision*, *recall*, dan *F1-score* berdasarkan data berlabel manual. Hasil pengujian menunjukkan bahwa model *Support Vector Machine* dengan *kernel* linear mencapai tingkat akurasi sebesar 100% dengan nilai *F1-score* sebesar 1.00, yang menunjukkan performa klasifikasi sentimen yang sangat baik. Analisis visualisasi *word cloud* menunjukkan bahwa sentimen positif didominasi oleh aspek kenyamanan dan fasilitas, sementara sentimen negatif berkaitan dengan kebersihan, kepadatan, dan pengelolaan lingkungan. Hasil penelitian ini memberikan informasi berbasis data mengenai aspek-aspek utama yang perlu dipertahankan dan ditingkatkan dalam pengelolaan Alun-Alun Brebes sebagai ruang publik.

Kata Kunci: Analisis sentimen; *Support Vector Machine*; *TF-IDF*; *Text Mining*; Ulasan Wisata

Abstract—The rapid development of information technology has encouraged the use of digital platforms as media for sharing opinions, including tourism reviews on *Google Maps*. Alun-Alun Brebes, as one of the most frequently visited public spaces, has generated thousands of reviews with diverse textual characteristics, making manual analysis inefficient and impractical. This study aims to analyze visitor sentiment toward Alun-Alun Brebes by applying a *text mining* approach using the *Support Vector Machine* algorithm. The *dataset* consists of 1,000 *Google Maps* reviews, including 327 reviews manually labeled as positive and negative sentiments and 673 unlabeled reviews. The research stages include data collection, text *preprocessing*, *feature extraction* using the *Term Frequency–Inverse Document Frequency (TF-IDF)* method, *Support Vector Machine* model training and testing, and automatic *labeling* of unlabeled data. Model performance was evaluated using a *confusion matrix* with *accuracy*, *precision*, *recall*, and *F1-score* metrics based on the manually labeled data. The results show that the *Support Vector Machine* model with a *linear kernel* achieved an *accuracy* of 100% with an *F1-score* of 1.00, indicating excellent sentiment *classification* performance. Furthermore, *word cloud visualization* reveals that positive sentiment is dominated by aspects related to comfort and facilities, while negative sentiment is associated with cleanliness, crowd density, and environmental management. These findings provide data-driven insights into key aspects that should be maintained and improved in managing Alun-Alun Brebes as a public space.

Keywords: Sentiment Analysis; Support Vector Machine; TF-IDF; Text Mining; Tourism Reviews

1. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah cara masyarakat memperoleh, menyampaikan, dan mengevaluasi informasi, termasuk dalam konteks pariwisata dan pemanfaatan ruang publik. Platform digital kini tidak hanya berfungsi sebagai penyedia informasi lokasi, tetapi juga sebagai media utama bagi pengguna untuk menyampaikan pengalaman, opini, dan penilaian mereka secara terbuka. Ulasan daring telah banyak dimanfaatkan sebagai sumber data penelitian pariwisata karena mampu merepresentasikan persepsi dan pengalaman pengunjung secara langsung dan berskala besar (Mariani et al., 2020). Salah satu platform yang paling banyak digunakan dalam konteks ini adalah *Google Maps*, yang menyediakan fitur ulasan berbasis teks dan penilaian numerik terhadap berbagai lokasi, seperti destinasi wisata, fasilitas umum, dan ruang publik. Ulasan pengguna pada *Google Maps* merepresentasikan pengalaman nyata pengunjung dan mengandung informasi penting mengenai persepsi masyarakat terhadap kualitas layanan, fasilitas, serta pengelolaan suatu lokasi. Oleh karena itu, data ulasan digital semakin banyak dimanfaatkan sebagai sumber informasi dalam penelitian analisis sentimen untuk mendukung evaluasi berbasis data (Abdullah et al., 2021; Kumar et al., 2020; Li et al., 2021).

Alun-Alun Brebes merupakan salah satu ruang publik utama di Kabupaten Brebes yang memiliki peran strategis sebagai pusat aktivitas sosial, rekreasi, dan ekonomi masyarakat. Tingginya intensitas kunjungan masyarakat terhadap kawasan ini mendorong munculnya ribuan ulasan pengguna pada *Google Maps*. Ulasan tersebut berisi berbagai bentuk opini, mulai dari apresiasi terhadap kenyamanan dan fasilitas hingga kritik terkait kebersihan, kepadatan pengunjung, parkir, serta pengelolaan lingkungan. Informasi yang terkandung dalam ulasan-ulasan tersebut sangat berharga bagi pengelola dan pemerintah daerah karena mencerminkan persepsi dan pengalaman pengunjung secara langsung. Namun, jumlah data yang besar dan karakteristik teks yang tidak terstruktur menyebabkan proses analisis secara manual menjadi tidak efisien, memerlukan waktu yang lama, serta berpotensi menghasilkan interpretasi yang subjektif.

Keterbatasan analisis manual terhadap data ulasan digital menimbulkan kebutuhan akan pendekatan otomatis yang mampu mengolah data teks dalam jumlah besar secara cepat, konsisten, dan objektif. Salah satu pendekatan yang

banyak digunakan untuk tujuan tersebut adalah analisis sentimen, yang merupakan bagian dari *text mining*. Analisis sentimen bertujuan untuk mengidentifikasi polaritas opini pengguna, seperti sentimen positif dan negatif, berdasarkan isi teks yang dianalisis. Pendekatan ini telah diterapkan secara luas dalam berbagai bidang, termasuk evaluasi layanan publik, analisis ulasan aplikasi digital, serta kajian pariwisata berbasis data ulasan daring (Cambria et al., 2020; Yadav & Vishwakarma, 2020). Dalam beberapa tahun terakhir, berbagai penelitian telah mengkaji penerapan analisis sentimen pada ulasan wisata dan layanan publik menggunakan algoritma pembelajaran mesin. (Ipmawati et al., 2024) menerapkan analisis sentimen ulasan wisata pada *Google Maps* menggunakan algoritma *Support Vector Machine* dan menunjukkan bahwa metode tersebut mampu menghasilkan performa klasifikasi yang baik. Penelitian lain oleh (Damayanti et al., 2024) serta (Ichwani et al., 2024) juga membuktikan bahwa algoritma *Support Vector Machine* memiliki kinerja yang stabil dalam mengklasifikasikan sentimen ulasan digital (Syahlan et al., 2023). Selain itu, (Mubarok et al., 2023) dan (Nursalim & Novita, 2023) menunjukkan bahwa *Support Vector Machine* efektif digunakan untuk mengolah data teks berdimensi tinggi dengan variasi bahasa yang umum ditemukan pada ulasan pengguna.

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada klasifikasi polaritas sentimen positif dan negatif sebagai tujuan utama analisis. Pendekatan tersebut menghasilkan informasi yang bersifat umum dan deskriptif, tetapi belum sepenuhnya mampu menjawab kebutuhan praktis dalam pengelolaan ruang publik. Dalam konteks pengambilan keputusan, pengelola tidak hanya membutuhkan informasi mengenai dominasi sentimen positif atau negatif, tetapi juga perlu mengetahui aspek-aspek spesifik yang menjadi penyebab munculnya sentimen tersebut. Tanpa adanya identifikasi aspek, hasil analisis sentimen menjadi kurang operasional dan sulit dimanfaatkan sebagai dasar perbaikan kebijakan atau pengelolaan fasilitas (Zhang & Luo, 2022). Seiring dengan perkembangan penelitian analisis sentimen, beberapa studi mulai mengintegrasikan analisis polaritas sentimen dengan teknik visualisasi dan eksplorasi kata dominan, seperti *word cloud*. Pendekatan ini memungkinkan peneliti untuk mengidentifikasi tema dan aspek utama yang sering muncul dalam ulasan pengguna, sehingga memberikan konteks yang lebih kaya terhadap hasil klasifikasi sentimen (Al-Adaileh et al., 2024). Namun, penerapan pendekatan tersebut pada ruang publik lokal, khususnya di wilayah non-metropolitan seperti Kabupaten Brebes, masih relatif terbatas.

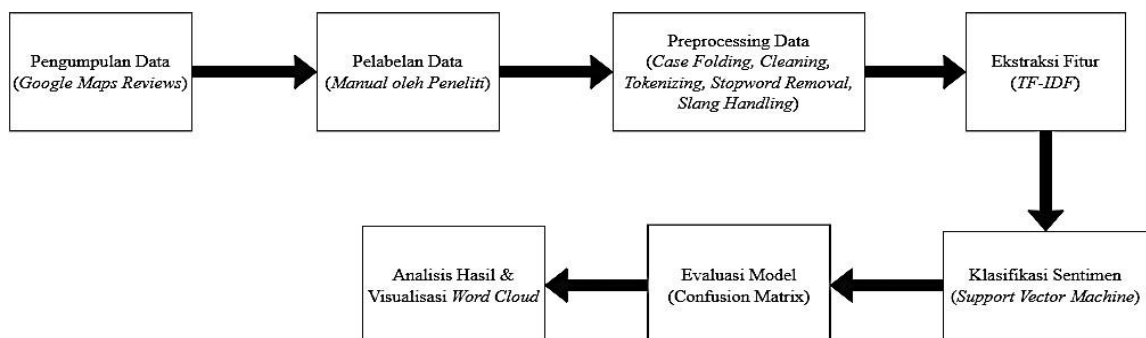
Berdasarkan kondisi tersebut, terdapat urgensi untuk melakukan penelitian yang tidak hanya mengklasifikasikan sentimen ulasan pengguna, tetapi juga mampu mengidentifikasi aspek-aspek utama yang menjadi perhatian pengunjung ruang publik. Oleh karena itu, penelitian ini menawarkan solusi berupa penerapan analisis sentimen berbasis *text mining* menggunakan algoritma *Support Vector Machine* terhadap ulasan pengguna *Google Maps* mengenai Alun-Alun Brebes. Algoritma *Support Vector Machine* dipilih karena kemampuannya dalam menangani data teks berdimensi tinggi serta menghasilkan performa klasifikasi yang stabil, sebagaimana dibuktikan dalam berbagai penelitian terkini (Ichwani et al., 2024; Joachims, 2020). Tujuan penelitian ini adalah untuk menganalisis sentimen ulasan pengguna *Google Maps* terhadap Alun-Alun Brebes menggunakan algoritma *Support Vector Machine* serta mengidentifikasi pola opini dan aspek utama yang menjadi perhatian pengunjung melalui analisis kata dominan dan visualisasi *word cloud*. Kontribusi utama penelitian ini terletak pada pengembangan analisis sentimen yang bersifat lebih operasional dan aplikatif pada konteks ruang publik lokal, sehingga hasil penelitian dapat dimanfaatkan sebagai dasar evaluasi dan pengambilan keputusan berbasis data oleh pengelola dan pemerintah daerah.

2. METODOLOGI PENELITIAN

Analisis sentimen digunakan untuk mengklasifikasikan opini pengguna ke dalam dua kelas sentimen, yaitu sentimen positif dan sentimen negatif. Pendekatan ini umum digunakan dalam penelitian analisis ulasan digital karena mampu mengolah data teks dalam jumlah besar secara sistematis dan objektif (Yadav & Vishwakarma, 2020).

2.1 Tahapan Penelitian

Tahapan penelitian dilakukan secara berurutan mulai dari pengumpulan data, pelabelan data, *preprocessing* teks, ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency (TF-IDF)*, proses klasifikasi menggunakan *Support Vector Machine (SVM)*, hingga evaluasi performa model dan analisis hasil. Gambar 1 menunjukkan bagan alir tahapan penelitian yang dilakukan, mulai dari proses pengumpulan data hingga analisis hasil sentimen.



Gambar 1. Bagan Alir Tahapan Penelitian



2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna *Google Maps* terhadap Alun-Alun Brebes. Pengumpulan data dilakukan menggunakan teknik *web scraping* dengan bantuan ekstensi peramban Google Chrome. Data yang diperoleh kemudian disimpan dalam format Microsoft Excel (.xlsx). *Dataset* yang digunakan berjumlah 1.000 ulasan yang mencerminkan pengalaman dan persepsi pengunjung terhadap berbagai aspek pengelolaan Alun-Alun Brebes sebagai ruang publik. Dari total 1.000 data ulasan yang diperoleh, sebanyak 327 data diberi label sentimen secara manual, sedangkan 673 data lainnya tidak memiliki label. Pelabelan sentimen dilakukan secara manual oleh peneliti yang memahami konteks pariwisata dan karakteristik ulasan wisata, dengan mengacu pada panduan pelabelan sentimen yang disusun berdasarkan makna dan konteks kalimat dalam ulasan. Proses pelabelan dilakukan dengan menentukan polaritas sentimen positif atau negatif berdasarkan opini dominan yang disampaikan oleh pengguna dalam setiap ulasan. Penggunaan panduan pelabelan ini bertujuan untuk mengurangi subjektivitas dalam proses penentuan label sentimen sehingga hasil pelabelan dapat digunakan secara konsisten sebagai data latih dan data uji dalam pembentukan serta evaluasi model klasifikasi sentimen. Proses pelabelan data merupakan bagian dari tahapan pengumpulan data sebagaimana ditunjukkan pada Gambar 1.

2.3 Preprocessing Data Teks

Tahap *preprocessing* data teks bertujuan untuk membersihkan dan menormalkan data ulasan sebelum diproses lebih lanjut pada tahap ekstraksi fitur dan klasifikasi sentimen. Mengingat data ulasan *Google Maps* bersifat tidak terstruktur dan mengandung banyak *noise*, tahapan *preprocessing* dilakukan secara bertahap agar representasi data teks menjadi lebih konsisten dan informatif.

Tahapan *preprocessing* yang diterapkan meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *slang handling*. *Case folding* dilakukan dengan mengubah seluruh teks ulasan menjadi huruf kecil untuk menghindari perbedaan makna akibat variasi penggunaan huruf kapital. Tahap *cleaning* diterapkan untuk menghapus tanda baca, angka, simbol khusus, serta karakter non-alfabet yang tidak memiliki kontribusi terhadap analisis sentimen. Selanjutnya, tahap *tokenizing* dilakukan untuk memecah teks ulasan menjadi unit kata. Tahap *stopword removal* digunakan untuk menghilangkan kata-kata umum dalam bahasa Indonesia yang tidak memiliki makna sentimen yang signifikan, seperti kata penghubung dan kata depan. Selain itu, tahap *slang handling* diterapkan untuk menormalisasi penggunaan bahasa informal dan singkatan yang umum ditemukan dalam ulasan *Google Maps* dengan mengubahnya ke dalam bentuk kata baku. Seluruh tahapan *preprocessing* ini bertujuan untuk meningkatkan kualitas representasi data teks dan mendukung kinerja model klasifikasi sentimen yang digunakan (Haddi et al., 2021; Safawi et al., 2024).

2.4 Ekstraksi Fitur

Setelah melalui tahap *preprocessing*, data teks direpresentasikan ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*. Metode *TF-IDF* digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan tingkat kepentingannya terhadap keseluruhan dokumen dalam *dataset*. Bobot *TF-IDF* dihitung menggunakan persamaan sebagai berikut:

$$W_{i,j} = tf_{i,j} \times \log \left(\frac{D}{df_j} \right) \quad (1)$$

Dengan keterangan $W_{i,j}$ adalah bobot *TF-IDF* term ke- j pada dokumen ke- i , $tf_{i,j}$ adalah frekuensi term, D adalah jumlah dokumen, df_j adalah jumlah dokumen yang mengandung term j . Pendekatan ini memungkinkan kata-kata yang bersifat informatif dan relevan terhadap sentimen memiliki bobot yang lebih tinggi, sementara kata-kata yang sering muncul tetapi kurang bermakna akan memiliki bobot yang lebih rendah. Metode *TF-IDF* banyak digunakan dalam penelitian analisis sentimen karena efektif dalam merepresentasikan data teks berdimensi tinggi dan meningkatkan performa algoritma klasifikasi berbasis pembelajaran mesin (Kumar et al., 2020; Safawi et al., 2024). Rumus *TF-IDF* tersebut digunakan untuk mengukur tingkat kepentingan suatu term dalam dokumen tertentu dengan mempertimbangkan frekuensi kemunculannya pada keseluruhan dokumen, sehingga fitur teks yang dihasilkan lebih representatif untuk proses klasifikasi.

2.5 Klasifikasi Sentimen

Proses klasifikasi sentimen dilakukan menggunakan algoritma *Support Vector Machine (SVM)*. Algoritma ini dipilih karena memiliki kemampuan yang baik dalam menangani data berdimensi tinggi serta menghasilkan performa klasifikasi yang stabil pada data teks. Berbagai studi menunjukkan bahwa algoritma pembelajaran mesin tradisional seperti *Support Vector Machine* masih memberikan performa yang kompetitif dalam analisis sentimen teks, khususnya ketika dikombinasikan dengan representasi fitur *TF-IDF* dan digunakan pada dataset berukuran menengah (Vargas-Calderón & Camacho, 2021). Pada penelitian ini, model *SVM* digunakan untuk mengklasifikasikan ulasan ke dalam dua kelas sentimen, yaitu sentimen positif dan sentimen negatif, dengan memanfaatkan fitur *TF-IDF* sebagai masukan. Model dilatih menggunakan data berlabel manual yang kemudian digunakan untuk mempelajari pola sentimen dalam data ulasan. Pemilihan *SVM* sebagai algoritma klasifikasi didasarkan pada hasil penelitian terdahulu yang menunjukkan keunggulan *SVM* dalam tugas klasifikasi sentimen berbasis teks (Ichwani et al., 2024; Joachims, 2020).

2.6 Evaluasi Model

Evaluasi performa model dilakukan untuk mengukur tingkat keberhasilan model SVM dalam mengklasifikasikan sentimen ulasan secara akurat. Evaluasi dilakukan menggunakan *confusion matrix* yang menghasilkan metrik evaluasi berupa nilai akurasi, *precision*, *recall*, dan *F1-score*. Data berlabel manual dibagi menjadi data latih dan data uji dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Penggunaan metrik evaluasi tersebut bertujuan untuk memberikan gambaran yang komprehensif mengenai kinerja model klasifikasi yang digunakan dalam penelitian ini (Minaee et al., 2021). Hasil perhitungan *confusion matrix* pada penelitian ini disajikan pada Tabel 1.

Tabel 1. *Confusion Matrix*

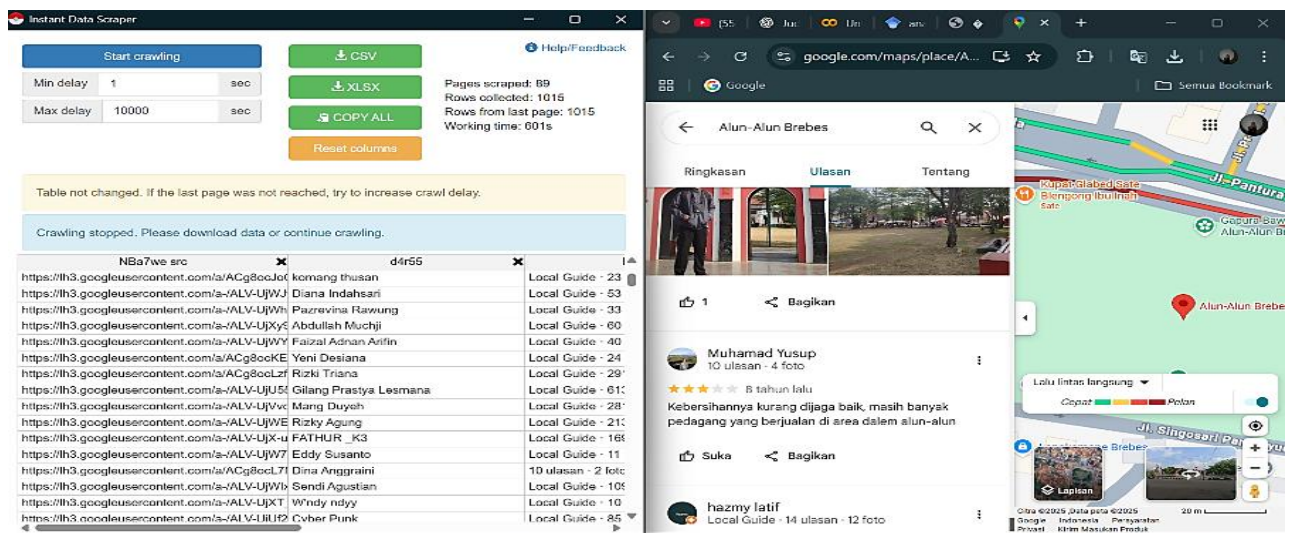
Prediksi / Aktual	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Berdasarkan *confusion matrix*, nilai evaluasi ditentukan oleh True Positive (TP), yaitu data positif yang diklasifikasikan dengan benar, True Negative (TN) sebagai data negatif yang diklasifikasikan dengan benar, False Positive (FP) sebagai data negatif yang salah diklasifikasikan sebagai positif, serta False Negative (FN) sebagai data positif yang salah diklasifikasikan sebagai negatif.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Scraping Data

Pada penelitian ini, proses pengumpulan data dilakukan dengan teknik *web scraping* terhadap ulasan *Google Maps* Alun-Alun Brebes. Scraping data dilakukan untuk memperoleh ulasan pengunjung secara langsung yang mencerminkan opini dan pengalaman masyarakat terhadap fasilitas, kenyamanan, dan kondisi kawasan Alun-Alun Brebes. Proses scraping dilakukan menggunakan bantuan ekstensi peramban Google Chrome dan hasil ulasan disimpan dalam format Microsoft Excel (.xlsx). Data yang dikumpulkan terdiri dari atribut No, Nama Pengguna, *Review*, dan Label. Dari hasil scraping diperoleh sebanyak 1000 data ulasan. Selanjutnya, sebagian data diberi label sentimen secara manual untuk keperluan pelatihan dan evaluasi model klasifikasi. Tampilan proses pengambilan data ulasan *Google Maps* ditunjukkan pada Gambar 2, yang memperlihatkan proses scraping menggunakan melalui Instant Data Scraper di Google Chrome.



Gambar 2. Hasil Scraping Data

3.2 Preprocessing Teks

Tahap *preprocessing* data teks menghasilkan representasi teks yang lebih bersih dan terstruktur. Proses ini mencakup *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *slang handling*. Tahapan ini berhasil menghilangkan kata tidak bermakna, simbol, serta menormalisasi penggunaan bahasa informal yang umum ditemukan dalam ulasan *Google Maps*. Hasil *preprocessing* ini berkontribusi pada peningkatan kualitas fitur teks yang digunakan pada tahap ekstraksi fitur dan klasifikasi sentimen.

3.1.1 Case Folding

Case folding bertujuan untuk mengubah seluruh teks ulasan menjadi huruf kecil (lowercase) guna menghindari perbedaan makna akibat variasi penggunaan huruf kapital. Contoh perubahan hasil *case folding* ditunjukkan pada Tabel 2.

**Tabel 2.** Proses *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
Alun Alun Sangat Nyaman Untuk Wisata	alun alun sangat nyaman untuk wisata
Tempatnya Bagus dan Bersih	tempatnya bagus dan bersih

3.1.2 Cleaning

Tahap *cleaning* bertujuan untuk membersihkan teks ulasan dari karakter yang tidak relevan seperti tanda baca, angka, simbol khusus, serta spasi berlebih yang dapat menimbulkan *noise* dalam proses analisis. Proses *cleaning* dilakukan setelah tahap *case folding* dan sebelum proses *tokenizing* sehingga teks yang dihasilkan menjadi lebih rapi dan mudah diproses. Dengan menghilangkan karakter yang tidak memiliki kontribusi terhadap analisis sentimen, tahap ini membantu meningkatkan kualitas data teks yang digunakan pada tahap selanjutnya. Contoh hasil proses *cleaning* ditunjukkan pada Tabel 3.

Tabel 3. Proses *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
alun-alun sangat nyaman!!!	alun alun sangat nyaman
tempatnya bagus & bersih □	tempatnya bagus bersih
parkirnya sempit!!! □	parkirnya sempit

3.1.3 Tokenizing

Tahap *tokenizing* bertujuan untuk memecah kalimat ulasan menjadi potongan kata atau token. Proses ini dilakukan setelah teks melalui tahap *case folding* sehingga menghasilkan token yang lebih konsisten. Contoh hasil proses *tokenizing* terhadap ulasan pengunjung ditunjukkan pada Tabel 4, yang memperlihatkan perubahan teks ulasan dari bentuk kalimat menjadi sekumpulan token kata.

Tabel 4. Proses *Tokenizing*

Kalimat	Hasil <i>Tokenizing</i>
alun alun sangat nyaman	['alun','alun','sangat','nyaman']
tempatnya bagus dan bersih	['tempatnya','bagus','dan','bersih']

3.1.4 Stop Word Removal

Tahap *stop word removal* dilakukan untuk menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap sentimen, seperti dan, yang, di, ke, dan sebagainya. Penghapusan *stopword* dilakukan agar fitur yang dihasilkan lebih representatif. Contoh hasil penerapan proses *stop word removal* pada ulasan pengunjung ditampilkan pada Tabel 5, yang menunjukkan perbedaan teks sebelum dan sesudah penghapusan *stopword*.

Tabel 5. Proses *Stop Word Removal*

Sebelum <i>Stop Word</i>	Sesudah <i>Stop Word</i>
tempatnya bagus dan bersih	['tempatnya','bagus','bersih']
alun alun sangat nyaman	['alun','alun','nyaman']

3.1.5 Slang Handling

Tahap *slang handling* bertujuan untuk menormalisasi penggunaan bahasa informal atau kata tidak baku yang sering digunakan oleh pengguna dalam ulasan *Google Maps*. Bahasa informal seperti singkatan, kata gaul, dan variasi ejaan dapat menyebabkan perbedaan representasi kata yang memiliki makna serupa. Oleh karena itu, kata-kata tidak baku diubah ke dalam bentuk kata baku agar makna teks menjadi lebih seragam. Tahap ini penting untuk meningkatkan konsistensi data teks dan kualitas fitur yang dihasilkan. Contoh hasil proses *slang handling* ditunjukkan pada Tabel 6.

Tabel 6. Proses *Slang Handling*

Sebelum <i>Slang Handling</i>	Sesudah <i>Slang Handling</i>
tempatnya mantap bgt	tempatnya bagus banget
parkirnya gk nyaman	parkirnya tidak nyaman
suasananya oke banget	suasananya baik banget

3.3 Ekstraksi Fitur *TF-IDF*

Setelah proses ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* dilakukan, analisis dilanjutkan dengan menghitung rata-rata bobot *TF-IDF* untuk setiap kelas sentimen, yaitu sentimen positif dan sentimen negatif. Penyajian nilai rata-rata bobot *TF-IDF* ini bertujuan untuk mengidentifikasi kata-kata yang paling dominan dan representatif dalam masing-masing kelas sentimen. Perlu ditegaskan bahwa nilai *TF-IDF* yang disajikan pada tabel ini bukan merupakan vektor *TF-IDF* per dokumen individu, melainkan rata-rata bobot *TF-IDF*

yang diperoleh dari agregasi seluruh dokumen dalam setiap kelas sentimen. Pendekatan ini digunakan untuk menghindari kesalahpahaman pembaca serta memudahkan analisis perbedaan karakteristik kata antara sentimen positif dan negatif. Hasil perhitungan rata-rata bobot *TF-IDF* per kelas sentimen ditunjukkan pada Tabel 7.

Tabel 7. Rata-rata Bobot *TF-IDF* per Kelas Sentimen

Kata	Positif	Negatif
nyaman	0.36	0.02
bersih	0.31	0.01
kotor	0.03	0.28
ramai	0.22	0.19

3.4 Klasifikasi Sentimen Menggunakan *Support Vector Machine* (SVM)

Klasifikasi sentimen pada penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dengan *kernel* linear. Algoritma SVM dipilih karena kemampuannya dalam menangani data berdimensi tinggi, khususnya data teks yang telah direpresentasikan dalam bentuk vektor menggunakan metode *Term Frequency-Inverse Document Frequency* (*TF-IDF*). Sebanyak 327 data ulasan yang telah diberi label secara manual digunakan sebagai data latih dan data uji. Pembagian data dilakukan dengan perbandingan 80% data latih dan 20% data uji. Proses pembagian data, pembentukan fitur menggunakan *TF-IDF*, serta pelatihan dan pengujian model SVM ditunjukkan pada Gambar 3. Pada tahap ini, model dilatih menggunakan data latih dan selanjutnya diuji menggunakan data uji untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen ulasan ke dalam kelas positif dan negatif.

```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)

tfidf = TfidfVectorizer(max_features=5000, ngram_range=(1,2))
X_train_tf = tfidf.fit_transform(X_train)
X_test_tf = tfidf.transform(X_test)

svm = LinearSVC()
svm.fit(X_train_tf, y_train)

y_pred = svm.predict(X_test_tf)

print("Akurasi:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

Gambar 3. Proses pelatihan dan pengujian model *Support Vector Machine*

Setelah proses pelatihan dan pengujian selesai, model SVM yang telah terbentuk kemudian digunakan untuk melakukan prediksi sentimen pada data ulasan yang belum memiliki label. Sebanyak 673 data ulasan tanpa label diproses menggunakan model SVM hasil pelatihan untuk menghasilkan label sentimen secara otomatis. Proses prediksi sentimen pada data tanpa label ini ditunjukkan pada Gambar 4.

```
X_unlab_tf = tfidf.transform(df_unlabeled['clean_review'])
df_unlabeled['label_pred'] = svm.predict(X_unlab_tf)

# Masukkan ke data utama
df.loc[df['label'].isna(), 'label'] = df_unlabeled['label_pred']
```

Gambar 4. Proses prediksi sentimen pada data tanpa label menggunakan model SVM

Melalui tahapan tersebut, seluruh *dataset* yang berjumlah 1000 data ulasan berhasil memiliki label sentimen, yang terdiri dari 327 data berlabel manual dan 673 data hasil prediksi model SVM.

3.5 *Confusion Matrix*

Berdasarkan hasil evaluasi menggunakan *confusion matrix* pada data berlabel manual yang ditunjukkan pada Gambar 4, diperoleh kinerja model klasifikasi sentimen menggunakan *Support Vector Machine* (SVM) yang sangat baik. Dari total 327 data ulasan berlabel manual, yang terdiri dari 54 ulasan sentimen negatif dan 273 ulasan sentimen positif, seluruh data berhasil diklasifikasikan dengan benar oleh model. Hasil pengujian menunjukkan bahwa model mencapai nilai akurasi sebesar 100%, dengan nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 1.00 untuk kedua kelas sentimen. Nilai-nilai tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan sentimen positif dan negatif pada ulasan pengguna. Dengan demikian, hasil ini membuktikan bahwa model SVM yang digunakan efektif dalam melakukan klasifikasi sentimen ulasan berbasis data manual, sebagaimana ditunjukkan pada Gambar 5, serta menunjukkan konsistensi, stabilitas, reliabilitas, generalisasi yang kuat, dan potensi penerapan lebih lanjut pada *dataset* serupa di masa mendatang.



4. KESIMPULAN

Penelitian ini berhasil menjawab permasalahan analisis otomatis terhadap ulasan pengunjung Alun-Alun Brebes dengan menerapkan analisis sentimen berbasis *text mining* menggunakan algoritma *Support Vector Machine* (SVM) dan ekstraksi fitur *Term Frequency–Inverse Document Frequency* (TF-IDF). Hasil penelitian menunjukkan bahwa model SVM mampu mengklasifikasikan sentimen ulasan ke dalam kelas positif dan negatif secara sangat akurat, dengan tingkat akurasi sebesar 100%, serta nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 1.00 pada data berlabel manual. Temuan ini membuktikan bahwa kombinasi tahapan *preprocessing* data teks, ekstraksi fitur TF-IDF, dan algoritma SVM efektif dalam menangani karakteristik ulasan *Google Maps* yang mengandung variasi bahasa, teks singkat, serta keberadaan *noise*. Selain itu, hasil analisis *word cloud* memberikan gambaran yang lebih mendalam mengenai aspek-aspek utama yang menjadi perhatian pengunjung, di mana sentimen positif didominasi oleh kata-kata yang berkaitan dengan kenyamanan, fasilitas, dan suasana kawasan, sementara sentimen negatif lebih banyak berkaitan dengan isu kebersihan, pengelolaan parkir, kondisi lingkungan, dan kepadatan pengunjung. Informasi ini menunjukkan bahwa hasil analisis sentimen tidak hanya berkontribusi pada aspek teknis pengolahan data teks, tetapi juga pengambilan keputusan berbasis data dalam upaya peningkatan kualitas layanan dan pengelolaan ruang publik. Meskipun demikian, penelitian ini memiliki keterbatasan pada jumlah data yang dilabeli secara manual serta belum sepenuhnya mampu menangani penggunaan bahasa daerah atau dialek lokal seperti Brebesan yang muncul dalam ulasan. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas jumlah data berlabel, melibatkan lebih dari satu pelabel independen guna meningkatkan objektivitas, serta mengembangkan metode atau pendekatan yang lebih adaptif terhadap variasi bahasa lokal agar hasil analisis sentimen menjadi lebih komprehensif, akurat, dan memiliki tingkat generalisasi yang lebih baik.

REFERENCES

- Abdullah, N. A., Salleh, M. N. M., & Omar, K. (2021). Sentiment analysis of tourism reviews using machine learning techniques. *Journal of Tourism Futures*, 7(3), 287–299.
- Al-Adaileh, A., Al-Taani, A., & Alsmadi, I. (2024). Sentiment analysis approach for understanding users' opinions using word cloud visualization. *Journal of Big Data Analytics*, 9(2), 45–58.
- Alharbi, A., & de Silva, L. C. (2022). Aspect-based sentiment analysis using text mining and visualization techniques. *Information Processing & Management*, 59(4), 102105–102105.
- Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2020). Sentiment analysis: Beyond polarity. *IEEE Intelligent Systems*, 35(2), 58–62.
- Damayanti, E., Prasetyo, E., & Nugroho, A. (2024). Sentiment analysis of mobile application reviews using Support Vector Machine. *Journal of Information Systems Engineering and Business Intelligence*, 10(1), 23–32.
- Haddi, E., Liu, X., & Shi, Y. (2021). The role of text preprocessing in sentiment analysis. *Procedia Computer Science*, 182, 191–198.
- Ichwani, A., Sari, D. R., & Hidayat, R. (2024). Sentiment analysis of marketplace application reviews using Support Vector Machine. *Journal of Applied Informatics and Computing*, 8(1), 15–24.
- Ipmawati, J., Saifulloh, S., & Kusnawi, K. (2024). Analisis sentimen tempat wisata berdasarkan ulasan pada Google Maps menggunakan algoritma Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1).
- Joachims, T. (2020). Text categorization with Support Vector Machines: Learning with many relevant features. *Machine Learning*, 46(1), 137–142.
- Kumar, A., Sebastian, T. M., & Ravi, V. (2020). Sentiment analysis on online reviews using machine learning algorithms. *Expert Systems with Applications*, 141, 112930–112930.
- Li, X., Zhang, L., & Wang, Y. (2021). Sentiment analysis of online tourism reviews using machine learning techniques. *Journal of Hospitality and Tourism Management*, 46, 50–59.
- Mariani, M. M., Borghi, M., & Gretzel, U. (2020). Online reviews as a data source for tourism research: A systematic literature review. *Tourism Management*, 81, 104128–104128.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning–based text classification: A comprehensive review. *ACM Computing Surveys*, 54(3), 1–40.
- Mubarok, I. F. A., Huda, B., Hananto, A., Tukino, T., & Kabir, H. (2023). Analisis user sentiment aplikasi Google Maps, Maps.Me, dan Waze menggunakan metode Support Vector Machine. *RABIT: Jurnal Teknologi Dan Sistem Informasi Univrab*, 8(1), 69–74.
- Nursalim, A., & Novita, R. (2023). Sentiment analysis of comments on Google Play Store, Twitter, and YouTube using Support Vector Machine. *Jurnal Teknik Informatika (JUTIF)*, 4(6), 1305–1312.
- Safawi, N. U. C. M., Hamzah, M. H., & Yusof, N. (2024). Performance evaluation of TF-IDF for text classification. *International Journal of Advanced Computer Science and Applications*, 15(2), 112–119.
- Syahlan, M. S., Irmayanti, D., & Alam, S. (2023). Analisis sentimen tempat wisata berdasarkan komentar pengunjung menggunakan Support Vector Machine. *Simtek: Jurnal Sistem Informasi Dan Teknik Komputer*, 8(2), 315–319.



- Vargas-Calderón, V., & Camacho, D. (2021). A survey on sentiment analysis: From traditional to deep learning methods. *Artificial Intelligence Review*, 54, 6207–6254.
- Yadav, S. K., & Vishwakarma, A. (2020). Sentiment analysis using machine learning techniques: A review. *International Journal of Computer Applications*, 176(29), 1–7.
- Zhang, Y., & Luo, J. (2022). Aspect-based sentiment analysis on online reviews: A survey. *Knowledge-Based Systems*, 235, 107651–107651.