Jurnal Teknologi Terpadu

https://journal.nurulfikri.ac.id/index.php/jtt
ISSN : 2477-0043 ISSN ONLINE : 2460-7908

PERBANDINGAN ALGORITMA SVM DAN *DECISION TREE* UNTUK ANALISIS SENTIMEN ULASAN GOOGLE MAPS WISATA GUCI

Gilang Fajar Al-Fatih[1], Reza Dwi Cahya Kurniawan[2]

[1,2] Informatika, Universitas Muhammadiyah Tegal
Tegal, Jawa Tengah, Indonesia 52464
gilangfajar505@umtegal.ac.id, rezadck@umtegal.ac.id

Abstract

Online reviews on digital platforms significantly influence tourist decision-making. This study performs sentiment analysis on visitor reviews of the Guci Hot Spring tourist destination using machine learning algorithms. A total of 1,127 reviews were collected from Google Maps and preprocessed through cleaning, case folding, normalization, tokenization, stop-word removal, and duplicate elimination. Sentiment labelling was performed using a lexicon-based approach, classifying reviews into positive, negative, and neutral categories. The dataset was split into 80% for training and 20% for testing. Two algorithms were compared: Support Vector Machine (SVM) and Decision Tree (DT). Performance evaluation employed accuracy, precision, recall, and F1-score metrics. Results show that SVM *outperforms Decision Tree, achieving 69.9% accuracy and 0.680 macro-average F1-score, compared to Decision Tree's 64.2% accuracy and 0.622 macro-average F1-score. These findings demonstrate that SVM is more effective in handling high-dimensional text data, particularly in distinguishing ambiguous neutral-class reviews. This research provides valuable insights for tourism management to improve service quality based on visitor sentiment patterns.*

Keywords: *Decision Tree, Google Maps, Guci Hot Spring, Sentiment Analysis, SVM*

Abstrak

Ulasan daring pada *platform* digital semakin memengaruhi pengambilan keputusan wisatawan. Penelitian ini bertujuan melakukan analisis sentimen terhadap ulasan pengunjung destinasi wisata Pemandian Air Panas Guci menggunakan algoritma *machine learning*. Sebanyak 1.127 ulasan dikumpulkan dari Google Maps dan diproses melalui tahapan *preprocessing* meliputi *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, *stopword removal*, serta penghapusan data duplikat dan kosong. Pelabelan sentimen dilakukan menggunakan metode *lexicon-based* dengan tiga kategori: positif, negatif, dan netral. *Dataset* dibagi menjadi 80% data latih dan 20% data uji, kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM) dan *Decision Tree* (DT). Evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa SVM memberikan performa lebih baik dengan akurasi 69,9% dan *macro-average F1-score* sebesar 0,680, dibandingkan *Decision Tree* dengan akurasi 64,2% dan *macro-average F1-score* sebesar 0,622. Temuan ini membuktikan bahwa SVM lebih efektif menangani data teks berdimensi tinggi, khususnya dalam membedakan ulasan netral yang bersifat ambigu. Penelitian ini memberikan wawasan berharga bagi pengelola wisata untuk meningkatkan kualitas layanan berdasarkan pola sentimen pengunjung.

Kata kunci: Analisis Sentimen, *Decision Tree*, Google Maps, SVM, Wisata Guci

1. PENDAHULUAN

Analisis sentimen merupakan salah satu cabang *text mining* yang banyak digunakan untuk memahami opini publik dari data berbasis teks. Metode ini telah banyak diterapkan pada berbagai domain, mulai dari ulasan aplikasi layanan digital [1], *e-commerce*, hingga pariwisata [2]. Kehadiran ulasan *online* seperti di Google Maps menjadi salah satu sumber informasi penting yang memengaruhi keputusan calon pengunjung dalam memilih destinasi wisata.

Pemandian Air Panas Guci di Kabupaten Tegal merupakan salah satu destinasi wisata populer yang banyak mendapatkan ulasan dari pengunjung. Ulasan tersebut mengandung informasi berharga terkait kepuasan, kritik, maupun saran terhadap fasilitas dan layanan wisata. Namun, banyaknya jumlah ulasan membuat sulit untuk menganalisis secara manual, sehingga dibutuhkan pendekatan komputasi seperti analisis sentimen [3].

Diterima 29 Oktober 2025, Direvisi 2 November 2025, Diterima untuk publikasi 15 Desember 2025

Penelitian terdahulu telah banyak menerapkan *machine learning* dan *deep learning* untuk analisis sentimen ulasan digital maupun pariwisata. Pada penelitian [1] menggunakan metode Bi-LSTM dan Word2Vec pada ulasan aplikasi Identitas Kependudukan Digital (IKD) dan berhasil mencapai akurasi di atas 96%, meskipun penelitian tersebut masih terbatas pada domain layanan aplikasi pemerintah digital. Pada sektor pariwisata, pada penelitian [4] melakukan analisis sentimen terhadap ulasan destinasi wisata di Kota Tegal menggunakan algoritma *Naïve Bayes* dan *Decision Tree*, dengan akurasi tertinggi sebesar 77,50% pada *Naïve Bayes*. Sementara itu, pada penelitian [5] membandingkan algoritma SVM dan *Decision Tree* untuk sistem rekomendasi wisata Yogyakarta dan menunjukkan bahwa SVM memiliki performa lebih unggul dengan akurasi mencapai 98,97%.

Selanjutnya, beberapa penelitian pada sektor pariwisata juga telah menerapkan analisis sentimen berbasis ulasan Google Maps. Penelitian oleh Arianto dkk. menganalisis sentimen destinasi wisata Borobudur dan Prambanan menggunakan pendekatan *Aspect-Based Sentiment Analysis* pada ulasan berbahasa campuran (*code-mixed*), dan hasilnya mampu mengidentifikasi aspek yang paling sering memperoleh respons negatif dari wisatawan [6].

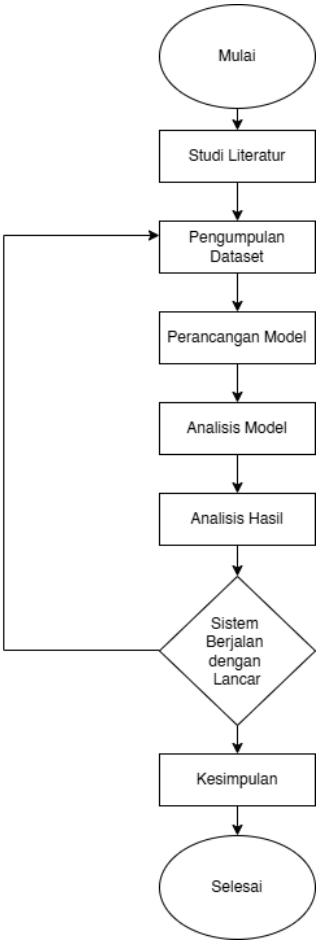
Selain itu, Ipmawati dkk. menerapkan algoritma *Support Vector Machine* (SVM) untuk analisis sentimen tempat wisata yang diulas pada Google Maps dan menunjukkan bahwa SVM memberikan performa klasifikasi yang cukup baik dalam membedakan sentimen positif dan negatif [7].

Di sisi lain, Rahman dkk. menggunakan metode *Decision Tree* untuk klasifikasi sentimen objek wisata dan menemukan bahwa akurasi model sangat bergantung pada kualitas *preprocessing* serta representasi fitur teks [8]. Ketiga studi ini memperlihatkan bahwa Google Maps merupakan sumber data yang relevan dalam mengetahui persepsi wisatawan secara langsung, serta SVM dan *Decision Tree* merupakan algoritma yang umum digunakan dalam pemodelan sentimen di bidang pariwisata.

Berdasarkan celah penelitian yang ada, mayoritas studi hanya mengklasifikasikan polaritas sentimen (positif, negatif, atau netral) tanpa mengaitkan opini dengan aspek layanan secara spesifik. Oleh karena itu, penelitian ini berkontribusi dengan menerapkan analisis sentimen berbasis aspek *Aspect-Based Sentiment Classification* (ABSC) pada ulasan Google Maps Wisata Pemandian Air Panas Guci serta membandingkan performa klasifikasi antara algoritma SVM dan *Decision Tree*. Hasil penelitian diharapkan mampu memberikan wawasan yang lebih mendalam terkait persepsi wisatawan, sebagai dasar evaluasi dalam peningkatan kualitas pelayanan wisata, serta memperkaya kajian akademik tentang penerapan *machine learning* dalam sektor pariwisata lokal.

2. METODE PENELITIAN

Penelitian ini melibatkan berbagai langkah yang dimulai dari pengumpulan data ulasan Google Maps pada destinasi Wisata Pemandian Air Panas Guci, dilanjutkan dengan proses pra-pemrosesan data meliputi pembersihan teks, translasi bahasa, *case folding*, normalisasi, tokenisasi, serta penghapusan *stopword*. Selanjutnya, label sentimen diklasifikasikan ke dalam tiga kategori positif, negatif, dan netral menggunakan pendekatan yang berbasis kosa kata dan ekstraksi aspek dengan metode *Aspect-Based Sentiment Classification* (ABSC). Data yang telah diproses kemudian dibagi menjadi data latih dan data uji, sebelum diklasifikasikan menggunakan algoritma SVM dan *Decision Tree*. Hasil dari tes dievaluasi dengan metrik akurasi, *confusion matrix*, serta *classification report*, kemudian divisualisasikan dalam bentuk *WordCloud*, frekuensi kata, dan analisis *N-Gram* untuk memberikan gambaran mendalam terkait opini wisatawan.



Gambar 1. Diagram Alir Proses Penelitian

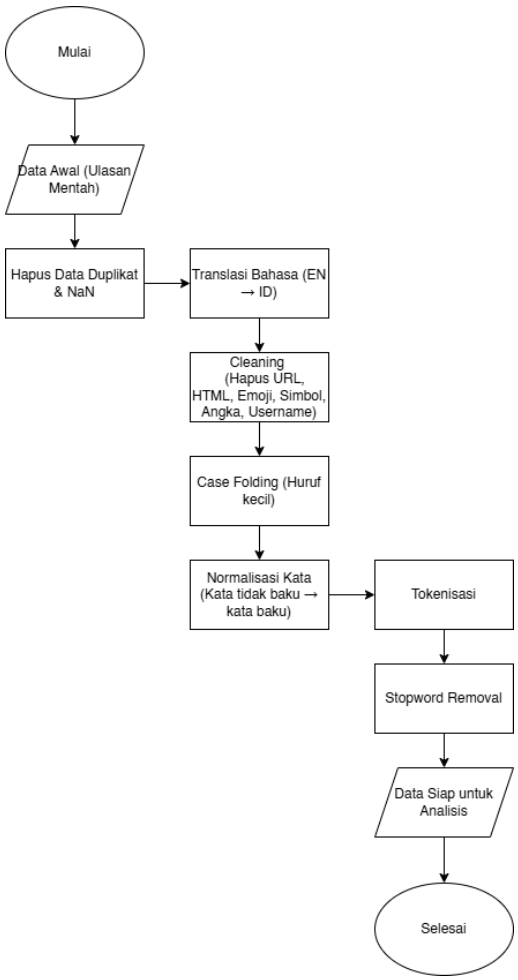
Proses penelitian digambarkan secara bertahap di Gambar 1. Pada diagram, titik awal adalah Mulai (*oval*). Pertama, penelitian literatur yang berkelanjutan dilakukan untuk mempelajari penelitian sebelumnya dan teknik yang relevan. Selanjutnya dilakukan Pengumpulan *Dataset* (persegi panjang), proses pengambilan ulasan dan penilaian dari Google Maps. Tahap ini dilanjutkan ke tahap Perancangan Model (persegi panjang), yang mencakup

konfigurasi eksperimen, pemilihan metode (SVM dan *Decision Tree*), dan pemilihan fitur. Setelah model dirancang, dilakukan Analisis Model (persegi panjang), yang melibatkan pelatihan dan pengujian model pada data latih dan uji. Tahap Analisis Hasil (persegi panjang) kemudian memasukkan analisis aspek, visualisasi, dan metrik performa (akurasi, ketepatan, *recall*, skor F1).

Pada titik berikutnya, terdapat sebuah keputusan (belah ketupat) bertuliskan Sistem Berjalan dengan Lancar yang mengevaluasi apakah model atau sistem telah berjalan sesuai harapan. Jika jawaban adalah ya, alur dilanjutkan ke Kesimpulan (persegi panjang) dan berakhir pada Selesai (*oval*). Jika jawaban adalah tidak, diagram menunjukkan alur balik (panah kembali) menuju tahap Pengumpulan *Dataset* untuk melakukan perbaikan siklus atau pemutakhiran data. Tindakan seperti penambahan data, perbaikan *preprocessing*, atau pengaturan model dilakukan sebelum mengulangi proses perancangan dan analisis. Alur balik ini menekankan sifat iteratif penelitian, yang berarti perbaikan berulang hingga model menjadi cukup efektif.

2.1. Pemrosesan Data

Penelitian ini menerapkan tahapan pemrosesan data secara menyeluruh, yang ditunjukkan pada Gambar 2 sebagai representasi Diagram Alir Pemrosesan Data.



Gambar 2. Diagram Alir Pemrosesan Data

Proses pemrosesan data pada penelitian ini dimulai dari penghapusan data duplikat dan data kosong, kemudian ulasan berbahasa Inggris diterjemahkan ke dalam Bahasa Indonesia. Selanjutnya dilakukan pembersihan teks dengan menghapus URL, *tag* HTML, emoji, simbol, angka, dan *username*, disertai *case folding* untuk menyeragamkan huruf menjadi kecil. Tahap berikutnya mencakup normalisasi kata, tokenisasi, serta *stopword removal* sehingga diperoleh data yang bersih dan siap untuk dianalisis lebih lanjut.

2.1.1. Pengumpulan Data

Tahap pengumpulan data pada penelitian ini dilakukan dengan mengambil ulasan dari platform Google Maps yang terkait dengan destinasi Wisata Pemandian Air Panas Guci di Kabupaten Tegal. Data yang diperoleh berupa teks ulasan yang ditulis langsung oleh pengunjung serta informasi tambahan berupa *rating* bintang yang diberikan pada lokasi tersebut. Proses pengambilan data dilakukan secara otomatis menggunakan pustaka Python Selenium [9], yang berfungsi untuk melakukan *web scraping* sehingga seluruh ulasan dapat diunduh dan disimpan dalam bentuk *dataset* [10]. Hasil dari tahap ini berupa kumpulan data mentah yang kemudian akan diproses lebih lanjut pada tahap pra-pemrosesan sebelum digunakan untuk analisis sentimen.

2.1.2. Proses Hapus Data Duplikat

Tahap awal yang dilakukan setelah data terkumpul adalah penghapusan data duplikat. Proses ini bertujuan untuk memastikan bahwa tidak ada ulasan yang muncul lebih dari satu kali di dalam *dataset*. Keberadaan data duplikat dapat menimbulkan bias dalam analisis karena ulasan yang sama akan dihitung berulang kali sehingga memengaruhi distribusi kelas sentimen maupun performa model klasifikasi. Dengan menghapus data yang berulang, kualitas *dataset* menjadi lebih representatif dan hasil analisis dapat mencerminkan kondisi sebenarnya [11].

2.1.3. Translasi Bahasa (EN – IND)

Dalam tahap ini dilakukan proses translasi bahasa dari ulasan yang diterjemahkan dari bahasa Inggris ke bahasa Indonesia. Hal ini penting karena sebagian pengunjung wisata menggunakan Bahasa Inggris dalam menuliskan ulasan di Google Maps, sementara analisis sentimen dalam penelitian ini difokuskan pada Bahasa Indonesia. Proses translasi dilakukan secara otomatis dengan bantuan pustaka Python yang memanfaatkan layanan *translation* API, sehingga seluruh teks ulasan dapat diseragamkan dalam satu bahasa. Dengan adanya translasi, data yang digunakan menjadi lebih konsisten dan meminimalisasi kesalahan interpretasi makna kata akibat perbedaan bahasa [12].

2.1.4. Pra-Pemrosesan Data

Pra-pemrosesan teks atau *text preprocessing* merupakan tahapan awal yang berfungsi untuk mengolah teks mentah menjadi bentuk yang lebih terstruktur sehingga siap

digunakan dalam analisis. Tahap ini memiliki peran penting dalam bidang *Natural Language Processing* (NLP) karena menentukan kualitas hasil pengolahan data teks [13]. Tujuan utama pra-pemrosesan teks adalah untuk membersihkan, menyeragamkan, dan menyederhanakan data teks mentah agar dapat diproses lebih efektif oleh algoritma *Natural Language Processing* (NLP) maupun *machine learning*. Melalui tahapan seperti *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, dan *stopword removal*, teks diubah menjadi format yang lebih terstruktur sehingga dapat mengurangi *noise*, meningkatkan konsistensi data, serta memperbaiki kualitas representasi teks untuk analisis lebih lanjut [14].

a. *Cleaning*

Pada tahap ini, ulasan dibersihkan dari elemen-elemen yang tidak relevan, seperti *Uniform Resource Locator* (URL), HTML *tag*, emoji, simbol, angka, maupun *username*. Tujuannya adalah agar data yang digunakan hanya berfokus pada teks ulasan yang mengandung opini [15].

b. *Case Folding*

Case folding adalah tahap pra-pemrosesan teks yang berfungsi untuk menyeragamkan penulisan huruf pada ulasan dengan cara mengubah seluruh huruf diubah menjadi huruf kecil (*lowercase*). Proses ini dilakukan agar kata dengan bentuk huruf berbeda tetapi memiliki makna sama, seperti “Bagus”, “BAGUS”, dan “bagus”, diperlakukan sebagai satu kata yang identik. Dengan demikian, jumlah variasi kata yang tidak perlu dapat dikurangi, sehingga analisis menjadi lebih efisien dan hasil klasifikasi lebih akurat [16].

c. Normalisasi Kata

Normalisasi kata adalah proses mengubah kata yang tidak baku, singkatan, maupun kata gaul menjadi bentuk kata baku sesuai dengan kaidah bahasa. Tahap ini penting karena ulasan pengguna sering kali menggunakan bahasa informal, singkatan, atau ejaan yang tidak sesuai standar. Contohnya, kata “gk”, “nggak”, atau “ga” *dinormalisasi* menjadi “tidak”, dan kata “bgt” menjadi “banget”. Dengan adanya normalisasi, variasi penulisan kata yang merujuk pada makna sama dapat diminimalisasi, sehingga data menjadi lebih konsisten dan memudahkan algoritma dalam mengenali pola teks. Normalisasi juga membantu meningkatkan akurasi analisis sentimen karena kata yang digunakan lebih terstruktur dan seragam [17].

d. Tokenisasi

Tokenisasi adalah proses memecah teks ulasan menjadi unit-unit terkecil berupa kata, frasa, atau simbol yang disebut token. Tahap ini bertujuan untuk memisahkan kalimat panjang menjadi potongan kata yang lebih sederhana, sehingga memudahkan dalam analisis dan perhitungan frekuensi kata. Misalnya, kalimat “Pemandangannya indah sekali” akan diubah menjadi token: [“pemandangannya”, “indah”, “sekali”]. Dengan tokenisasi, setiap kata dapat dianalisis secara individual untuk melihat kontribusinya terhadap sentimen yang dikandung dalam ulasan. Tahap ini menjadi sangat penting karena sebagian besar algoritma *Natural Language Processing* (NLP) bekerja dengan representasi kata sebagai unit analisis dasar [18].

e. *Stopword Removal*

Stopword removal adalah proses menghapus kata-kata umum yang sering muncul dalam teks namun tidak memiliki makna penting dalam analisis, seperti “yang”, “dan”, “di”, “ke”, atau “dengan”. Kata-kata tersebut biasanya tidak memberikan kontribusi signifikan terhadap penentuan sentimen karena hanya berfungsi sebagai kata penghubung atau penunjang kalimat. Dengan menghapus *stopword*, jumlah kata dalam *dataset* dapat diperkecil, sehingga mempercepat proses komputasi dan membuat analisis lebih fokus pada kata-kata bermakna yang benar-benar memengaruhi sentimen. Tahap ini juga membantu mengurangi *noise* pada data dan meningkatkan kualitas representasi teks untuk digunakan pada algoritma klasifikasi [19].

2.1.5. Pelabelan Sentimen

Tahap pelabelan data bertujuan untuk menentukan kelas sentimen dari setiap ulasan yang telah melalui proses pra-pemrosesan. Pada penelitian ini digunakan metode *lexicon-based*, yaitu pendekatan yang memanfaatkan kamus kata (*lexicon*) berisi daftar kata dengan bobot sentimen positif maupun negatif. Kata-kata dalam ulasan dicocokkan dengan daftar *lexicon* tersebut untuk menghitung skor sentimen secara keseluruhan.

Hasil dari proses pelabelan ini membagi ulasan ke dalam tiga kelas sentimen (3-*Class*), yaitu positif, negatif, dan netral. Ulasan dikategorikan positif apabila skor kata yang mengandung sentimen positif lebih dominan, dikategorikan negatif apabila kata-kata bernilai negatif lebih mendominasi, dan dikategorikan netral apabila tidak ada dominasi yang jelas di antara keduanya. Dengan pendekatan ini, pelabelan dapat dilakukan secara otomatis tanpa harus melakukan anotasi manual, sehingga lebih efisien untuk *dataset* dengan jumlah besar [20].

2.1.6. *Aspect-Based Sentiment Classification* (ABSC)

Aspect-Based Sentiment Classification (ABSC) digunakan dalam penelitian ini untuk menganalisis sentimen tidak hanya berdasarkan polaritas (positif, negatif, netral), tetapi juga mengaitkan opini dengan aspek tertentu dari Wisata Pemandian Air Panas Guci. Aspek yang ditinjau meliputi fasilitas, kenyamanan/pengalaman wisatawan, serta harga dan aksesibilitas. Dengan metode ini, kata kunci dalam ulasan seperti “toilet”, “parkir”, atau “nyaman” dipetakan ke dalam aspek yang sesuai sehingga dapat diketahui bagian mana yang paling diapresiasi atau dikritik. Pendekatan ini

memberikan wawasan lebih mendalam karena tidak hanya menunjukkan kecenderungan sentimen secara umum, tetapi juga mengidentifikasi aspek spesifik yang perlu ditingkatkan atau dipertahankan oleh pengelola wisata [21].

2.1.7. *Splitting* Data (80% Latih, 20% Uji)

Tahap *splitting* data dilakukan untuk membagi *dataset* yang telah melalui proses pra-pemrosesan dan pelabelan menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*). Pada penelitian ini digunakan perbandingan 80% data latih dan 20% data uji. Data latih digunakan untuk membangun dan melatih model klasifikasi sehingga algoritma dapat mempelajari pola dari ulasan yang ada. Sementara itu, data uji digunakan untuk mengukur kemampuan model dalam melakukan prediksi pada data yang belum pernah dilihat sebelumnya. Teknik ini bertujuan untuk menghindari *overfitting* serta memastikan model yang dihasilkan memiliki kemampuan generalisasi yang baik pada data baru [22].

2.1.8. Klasifikasi

a. *Support Vector Machine* (SVM)

Support Vector Machine (SVM) merupakan salah satu algoritma *supervised machine learning* yang dapat digunakan baik untuk klasifikasi maupun regresi [23]. Prinsip utama SVM adalah mencari *hyperplane* terbaik yang memisahkan data antar kelas. Pada kasus data *non-linier*, SVM melakukan transformasi data ke ruang berdimensi lebih tinggi melalui prosedur yang dikenal sebagai kernel *trick*. Karakteristik kernel memungkinkan datanya divisualisasikan dari sudut pandang sederhana hingga ke representasi yang lebih kompleks [24]. Beberapa jenis kernel yang umum digunakan dalam SVM adalah *kernel linear*, kernel *polynomial*, kernel *Gaussian Radial Basis Function (RBF)*, dan kernel *sigmoid*. Persamaan untuk masing-masing kernel adalah sebagai berikut:

1) Kernel *Linear*

$$K(x_i, x) = x_i x$$

2) Kernel RBF

$$K(x_i, x) = exp\left(\frac{-||_{x_i} - x||^2}{2\sigma^2}\right)$$

3) Kernel *Polynominal*

$$K(x_i, x) = (x_i x)^d$$

4) Kernel *Sigmoid*

$$K(x_i, x) = \tanh(\sigma(x_i x) + c)$$

Keterangan:

x_i *dan* x adalah data pada data pelatihan, parameter σ adalah *sigma*, kompleksitas C adalah derajat, d adalah *degree*, dan $-||x_i - x||^2$ merupakan kuadrat dari jarak total *vector* x_i dan x.

b. *Decision Tree*

Algoritma pohon keputusan, juga dikenal sebagai "Pohon Keputusan", didasarkan pada nilai informasi yang diperoleh. Metode untuk memilih atribut sesuai dengan pencapaian tujuan kelas yang bergantung pada penyelesaian objek. Algoritma ini menghasilkan akar dari pohon keputusan, yaitu dari atribut yang dipilih. Aspek yang memiliki nilai informasi yang paling besar adalah atribut akar dari pohon keputusan [25]. Persamaan untuk menghitung *entropy* dan gain data adalah sebagai berikut:

1) Mengumpulkan data pelatihan, yang dapat dikategorikan ke dalam kelompok tertentu, dan

2) Gunakan persamaan berikut untuk menghitung *entropy* setiap fitur:

$$Entropy\ (S) = \sum_{i=1}^{n} -p_i log_2 p_i$$

Di mana:

$S =$ Total Kasus

$n =$ Jumlah partisi S

$p_i =$ Jumlah kasus total dibagi dengan jumlah kelas adalah probabilitas

3) Menggunakan persamaan untuk menghitung informasi gain dari setiap atribut:

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|si|}{|s|} * Entropy\ (Si)$$

Di mana:

$S =$ Total Kasus

$A =$ Atribut

$n =$ Jumlah partisi S

$|Si| =$ Jumlah partisi ke $-$ i

$|S| =$ Jumlah kasus dalam S

4) Selanjutnya, hitung rasio keuntungan dengan menggunakan istilah *Split Information*, rumusnya:

$$split\ information = -\sum_{t=1}^{c} \frac{s1}{s} log_2 \frac{s1}{s}$$

5) Kemudian hitunglah *gain ratio*

$$gainratio(S.A) = \frac{Gain\ (S.A)}{SplitInformation(S.A)}$$

6) Ulangi langkah kedua sampai semua rekaman terkumpul.

2.1.9. Evaluasi Model (Akurasi, *Confusion Matrix*, *Classification Report*)

Evaluasi model dalam analisis sentimen biasanya menggunakan metrik seperti akurasi, *confusion matrix*, dan *classification report* yang sangat penting untuk mengukur performa model klasifikasi.

a) Akurasi mengukur persentase prediksi yang benar dari keseluruhan data. Nilai ini memberikan gambaran umum seberapa baik model melakukan klasifikasi [26].

b) *Confusion Matrix* adalah tabel yang menunjukkan jumlah prediksi benar dan salah per kelas, memberikan rincian kesalahan model terhadap kategori tertentu (positif, negatif, netral). Ini membantu mengidentifikasi kelas yang sulit diprediksi [27].

c) *Classification Report* menyediakan metrik lengkap seperti *precision* (ketepatan), *recall* (kelengkapan), dan *F1-score* (harmonik antara *precision* dan *recall*) untuk setiap kelas, sehingga evaluasi menjadi lebih komprehensif [28].

2.1.10. Visualisasi

Visualisasi dalam analisis sentimen ulasan wisata sering menggunakan tiga metode utama yaitu *WordCloud*, Frekuensi Kata, dan *N-Gram* untuk memberikan gambaran visual dan statistik mengenai kata atau frasa yang paling sering muncul.

a) *WordCloud* adalah visualisasi yang menampilkan kata-kata yang sering muncul dalam teks dengan ukuran proporsional terhadap frekuensinya. Misalnya, kata “indah”, “wisata”, “pantai”, dan “alam” sering muncul dalam ulasan wisata dan divisualisasikan untuk menunjukkan aspek yang dominan dalam ulasan wisatawan. *WordCloud* membantu mengidentifikasi tema utama yang disorot oleh pengunjung [29].

b) Frekuensi Kata mengukur banyaknya kemunculan setiap kata dalam ulasan. Ini memberikan data statistik yang lebih terperinci dan digunakan untuk memperkuat wawasan dari *WordCloud*. Frekuensi kata membantu menentukan kata-kata penting yang berkontribusi terhadap sentimen secara keseluruhan [30].

c) *N-Gram* adalah teknik yang mengelompokkan kata dalam urutan berpasangan (*bi-gram*), tiga kata *(tri-gram*), atau lebih, untuk menangkap konteks frasa daripada hanya kata individual. Dalam analisis sentimen wisata, *N-Gram* bisa menunjukkan pola seperti “harga tiket mahal” atau “pelayanan sangat baik” yang memberikan wawasan lebih spesifik tentang opini pengunjung [30].

Gabungan ketiga teknik ini membantu memberikan pemahaman yang lebih baik dan menyeluruh tentang ulasan pengunjung, mendukung analisis sentimen yang lebih akurat dan informatif dalam bidang pariwisata.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data ulasan diperoleh dari Google Maps pada lokasi Pemandian Air Panas Guci, Kabupaten Tegal, dengan menggunakan pustaka Python Selenium yang secara otomatis mengekstraksi teks ulasan (*review text*) dan *rating* bintang (*stars*). Seluruh data disimpan dalam format CSV untuk diolah lebih lanjut menggunakan pustaka *pandas*. *Dataset* yang diperoleh berisi kombinasi ulasan positif, negatif, dan netral yang ditulis langsung oleh pengunjung beserta *rating* bintang yang mereka berikan. Ringkasan hasil *scraping* ulasan ditampilkan pada Tabel 1.

Tabel 1. Contoh dan *Rating* pada *Dataset*

ulasan	*Rating*
Kalau datang pagi-pagi, udaranya dingin banget...	4
Tempat berlibur komplit dengan wahana waterboo...	5
Guci dari dulu mata air panasnya tdk pernah su...	4
Tempatnya luas. Ada beberapa lokasi untuk pema...	5
Lokasi sangat bersahabat untuk ajak keluarga d...	5

3.2. Proses Hapus Data Duplikat

Tahap ini bertujuan untuk menghapus data yang duplikat (*duplicate entries*) serta data dengan nilai kosong (*NaN*) agar kualitas *dataset* tetap terjaga. Proses pembersihan dilakukan menggunakan fungsi *drop_duplicates*() dan *dropna*() dari pustaka *pandas*. Dengan langkah ini, hanya data yang unik dan valid yang digunakan untuk tahap analisis berikutnya sehingga hasil tidak mengalami bias akibat adanya pengulangan ulasan dengan isi serupa. Perbandingan jumlah data sebelum dan sesudah proses pembersihan dapat dilihat pada Gambar 3 dan Gambar 4.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1130 entries, 0 to 1129
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   stars   1130 non-null   int64
 1   text    1130 non-null   object
dtypes: int64(1), object(1)
memory usage: 17.8+ KB
```

Gambar 3. Data awal

```
<class 'pandas.core.frame.DataFrame'>
Index: 1128 entries, 0 to 1129
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   stars   1128 non-null   int64
 1   text    1128 non-null   object
dtypes: int64(1), object(1)
memory usage: 26.4+ KB
```

Gambar 4. Setelah Hapus Data Duplikat

3.3. Translasi Bahasa (EN – IND)

Beberapa ulasan pengguna ditulis dalam Bahasa Inggris sehingga perlu dilakukan translasi agar seluruh data seragam dalam Bahasa Indonesia. Proses penerjemahan dilakukan secara otomatis menggunakan pustaka Googletrans (v4.0.0-rc1) pada Python, dengan hasil terjemahan disimpan dalam kolom *trans_indo*. Langkah ini memastikan konsistensi bahasa sehingga analisis sentimen dapat dilakukan secara akurat. Contoh hasil translasi ditampilkan pada Tabel 2.

Tabel 2. Contoh hasil translasi

stars	*text*	*trans_indo*
4	Kalau datang pagi-pagi, udaranya dingin banget...	Kalau datang pagi-pagi, udaranya dingin banget...
5	Tempat berlibur komplit dengan wahana waterboo...	Tempat Berlibur Komplit Delangan Wahana Waterb...
4	Guci dari dulu mata air panasnya tdk pernah su...	GUCI DARI DULU MATA AIR PANASYA TDK PERNAH SUR...
5	Tempatnya luas. Ada beberapa lokasi untuk pema...	Tempatnya Luas.ADA BEBERAPA LOKASI UNTUK PEMAN...
5	Lokasi sangat bersahabat untuk ajak keluarga d...	Lokasi sangat bersahabat untuk ajak keluarga d...

3.4. Pra-Pemrosesan Data

3.4.1. *Cleaning*

Tahap *cleaning* dilakukan untuk menghapus elemen-elemen yang tidak relevan seperti URL, tag HTML, emoji, simbol, angka, dan *username* dari teks ulasan. Tujuannya adalah menghilangkan *noise* agar data hanya berfokus pada isi opini pengguna yang relevan terhadap analisis sentimen. Hasil proses pembersihan teks ditampilkan pada Tabel 3.

Tabel 3. Contoh proses pembersihan

stars	*trans_indo*	*cleaning*
4	Kalau datang pagi-pagi, udaranya dingin banget...	Kalau datang pagipagi udaranya dingin bangetDi...
5	Tempat Berlibur Komplit Delangan Wahana Waterb...	Tempat Berlibur Komplit Delangan Wahana Waterb...
4	GUCI DARI DULU MATA AIR PANASYA TDK PERNAH SUR...	GUCI DARI DULU MATA AIR PANASYA TDK PERNAH SUR...
5	Tempatnya Luas.ADA BEBERAPA LOKASI UNTUK PEMAN...	Tempatnya LuasADA BEBERAPA LOKASI UNTUK PEMAND...
5	Lokasi sangat bersahabat untuk ajak keluarga d...	Lokasi sangat bersahabat untuk ajak keluarga d...

3.4.2. *Case Folding*

Tahap *case folding* bertujuan untuk menyeragamkan penulisan huruf dengan mengubah seluruh teks ulasan menjadi huruf kecil (*lowercase*). Langkah ini memastikan bahwa kata dengan bentuk huruf berbeda, seperti “Bagus”, “BAGUS”, dan “bagus”, dikenali sebagai satu kata yang sama oleh sistem. Hasil proses *case folding* ditampilkan pada Tabel 4.

Tabel 4. Contoh proses *case folding*

stars	*cleaning*	*case_folding*
4	Kalau datang pagipagi udaranya dingin bangetDi...	kalau datang pagipagi udaranya dingin bangetdi...
5	Tempat Berlibur Komplit Delangan Wahana Waterb...	tempat berlibur komplit delangan wahana waterb...
4	GUCI DARI DULU MATA AIR PANASYA TDK PERNAH SUR...	guci dari dulu mata air panasya tdk pernah sur...
5	Tempatnya LuasADA BEBERAPA LOKASI UNTUK PEMAND...	tempatnya luasada beberapa lokasi untuk pemand...
5	Lokasi sangat bersahabat untuk ajak keluarga d...	lokasi sangat bersahabat untuk ajak keluarga d...

3.4.3. Normalisasi Kata

Tahap normalisasi kata dilakukan untuk mengubah kata tidak baku, singkatan, atau bentuk slang menjadi kata baku sesuai kaidah Bahasa Indonesia. Proses ini menggunakan kamus kamuskatabaku.xlsx yang berisi pasangan kata tidak baku dan padanannya. Langkah ini penting untuk meningkatkan konsistensi data karena ulasan pengguna sering menggunakan bahasa informal, seperti “gk” atau “ga” yang dinormalisasi menjadi “tidak”, dan “bgt” menjadi “banget”. Hasil proses *normalisasi* ditampilkan pada Tabel 5.

Tabel 5. Contoh proses *normalisasi*

stars	*case_folding*	normalisasi
4	kalau datang pagipagi udaranya dingin bangetdi...	kalau datang pagipagi udaranya dingin bangetdi...
5	tempat berlibur komplit delangan wahana waterb...	tempat berlibur komplit delangan wahana waterb...
4	guci dari dulu mata air panasya tdk pernah sur...	guci dari dulu mata air panasya tidak pernah s...
5	tempatnya luasada beberapa lokasi untuk pemand...	tempatnya luasada beberapa lokasi untuk pemand...
5	lokasi sangat bersahabat untuk ajak keluarga d...	lokasi sangat bersahabat untuk ajak keluarga d...

3.4.4. Tokenisasi

Tahap tokenisasi bertujuan untuk memecah teks ulasan menjadi unit-unit terkecil berupa kata atau token. Proses ini dilakukan menggunakan fungsi *split*() agar setiap kata dalam kalimat dapat dianalisis secara terpisah. Dengan demikian, kalimat panjang diubah menjadi kumpulan kata sederhana yang siap digunakan pada tahap ekstraksi fitur atau perhitungan frekuensi kata. Hasil proses tokenisasi ditampilkan pada Tabel 6.

Tabel 6. Contoh proses tokenisasi

stars	normalisasi	*tokenize*
4	kalau datang pagipagi udaranya dingin bangetdi...	[kalau, datang, pagipagi, udaranya, dingin, ba...
5	tempat berlibur komplit delangan wahana waterb...	[tempat, berlibur, komplit, delangan, wahana, ...
4	guci dari dulu mata air panasya tidak pernah s...	[guci, dari, dulu, mata, air, panasya, tidak, ...
5	tempatnya luasada beberapa lokasi untuk pemand...	[tempatnya, luasada, beberapa, lokasi, untuk, ...
5	lokasi sangat bersahabat untuk ajak keluarga d...	[lokasi, sangat, bersahabat, untuk, ajak, kelu...

3.4.5. *Stopword Removal*

Tahap *stopword removal* dilakukan untuk mengurangi kata-kata yang sering muncul dalam teks namun tidak memiliki makna penting dalam analisis sentimen, seperti “yang”, “dan”, “di”, “ke”, dan “dengan”. Proses ini menggunakan pustaka *nltk.corpus.stopwords* untuk Bahasa Indonesia, ditambah dengan daftar *custom stopwords* agar hasil pembersihan lebih optimal. Dengan menghapus kata-kata yang tidak relevan ini, analisis menjadi lebih fokus pada kata-kata bermakna yang memengaruhi polaritas sentimen. Hasil proses *stopword removal* disajikan pada Tabel 7.

Tabel 7. Contoh proses *stopword removal*

stars	*tokenize*	*stopword_removal*
4	[kalau, datang, pagipagi, udaranya, dingin, ba...	pagipagi udaranya dingin bangetdi pilihan kola...
5	[tempat, berlibur, komplit, delangan, wahana, ...	berlibur komplit delangan wahana waterboom air...
4	[guci, dari, dulu, mata, air, panasya, tidak, ...	guci mata air panasya surut sebaikya kesini pa...
5	[tempatnya, luasada, beberapa, lokasi, untuk, ...	tempatnya luasada lokasi pemandianyaada kolam ...
5	[lokasi, sangat, bersahabat, untuk, ajak, kelu...	lokasi bersahabat ajak keluarga teman menikmat...

3.5. Pelabelan Sentimen

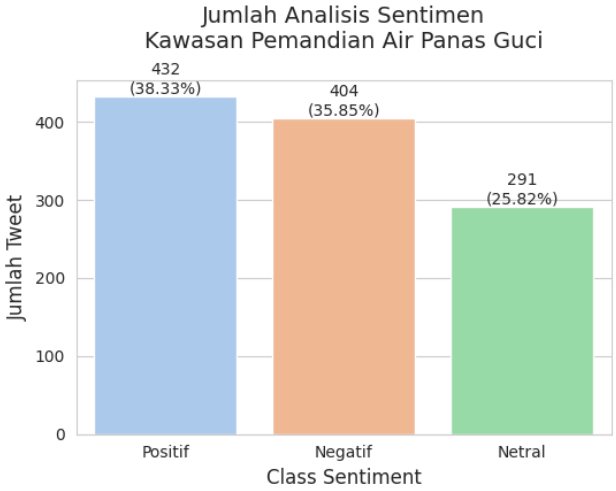
Tahap pelabelan sentimen dilakukan menggunakan metode *Lexicon-Based Sentiment Analysis* dengan memanfaatkan *dataset InSet* (Indonesian *Sentiment Lexicon*). Setiap kata dalam ulasan diberi skor berdasarkan polaritasnya, yaitu positif atau negatif, kemudian skor tersebut dijumlahkan untuk menentukan kategori sentimen. Klasifikasi dilakukan dengan ketentuan: positif jika skor total > 0, negatif jika skor total < 0, dan netral jika skor total = 0. Hasil pelabelan ini menghasilkan tiga kelas utama yang menggambarkan persepsi pengguna terhadap objek wisata dan divisualisasikan dalam bentuk diagram batang untuk menunjukkan distribusi data sentimen. Contoh hasil pelabelan ditampilkan pada Tabel 8.

Tabel 8. Contoh hasil pelabelan

no	*stopword_removal*	*score*	*sentiment*
1	pagipagi udaranya dingin bangetdi pilihan kola...	2	Positif
2	berlibur komplit delangan wahana waterboom air...	0	Netral
3	guci mata air panasya surut sebaikya kesini pa...	1	Positif
4	tempatnya luasada lokasi pemandianyaada kolam ...	-1	Negatif
5	lokasi bersahabat ajak keluarga teman menikmat...	0	Netral
6	airnya bersih viewnya bagus memuaskan harga ti...	5	Positif

Berdasarkan Gambar 5, sentimen positif menjadi yang paling dominan dengan 432 ulasan (38,33%), diikuti negatif sebanyak 404 ulasan (35,85%), dan netral sebanyak 291

ulasan (25,82%). Hasil ini menunjukkan mayoritas pengunjung memiliki pengalaman yang baik saat berwisata di Guci, meskipun ulasan negatif yang cukup tinggi masih memberi pengelola pelajaran untuk meningkatkan layanan dan fasilitas.



Gambar 5. Jumlah Analisis Sentimen

3.6. *Aspect-Based Sentiment Classification* (ABSC)

Aspect-Based Sentiment Classification (ABSC) digunakan untuk mengelompokkan sentimen berdasarkan aspek tertentu yang dibahas oleh pengguna dalam ulasan. Pada penelitian ini, terdapat tiga aspek utama, yaitu:

a. Fasilitas & Infrastruktur (mis. parkir, toilet, penginapan),

b. Pengalaman Wisatawan (mis. nyaman, indah, bersih), dan

c. Harga & Aksesibilitas (mis. tiket, murah, terjangkau).

Proses klasifikasi dilakukan dengan pendekatan *keyword matching* untuk mendeteksi kata kunci pada setiap ulasan dan mengaitkannya dengan aspek terkait. Hasil *ABSC* menggambarkan aspek yang paling sering menjadi perhatian pengunjung serta memberikan informasi lebih spesifik mengenai bagian wisata yang diapresiasi maupun dikritik, seperti ditunjukkan pada Tabel 9, 10, dan 11.

Tabel 9. Contoh *Aspect Extraction*

no	*stopword_removal*	*aspect_extraction*
1	pagipagi udaranya dingin bangetdi pilihan kola...	Fasilitas & Infrastruktur
2	berlibur komplit delangan wahana waterboom air...	Fasilitas & Infrastruktur
3	guci mata air panasya surut sebaikya kesini pa...	Fasilitas & Infrastruktur
4	tempatnya luasada lokasi pemandianyaada kolam ...	Fasilitas & Infrastruktur
5	lokasi bersahabat ajak keluarga teman menikmat...	Pengalaman Wisatawan
6	airnya bersih viewnya bagus memuaskan harga ti...	Fasilitas & Infrastruktur

Tabel 10. Kata Kunci untuk Ekstraksi Aspek

Aspek	Aspek
Fasilitas & Infrastruktur	[toilet, parkir, jalan, akses, wifi, kebersihan...
Pengalaman Wisatawan	[seru, indah, ramai, sunyi, tenang, nyaman, ram...
Harga & Aksesibilitas	[murah, mahal, gratis, terjangkau, diskon, biay...
[tempatnya, luasada, beberapa, lokasi, untuk, ...	tempatnya luasada lokasi pemandianyaada kolam ...
[lokasi, sangat, bersahabat, untuk, ajak, kelu...	lokasi bersahabat ajak keluarga teman menikmat...

Tabel 11. Contoh Hasil *Aspect Extraction*

No	Ulasan (Setelah *Preprocessing*)	Aspek yang Terdeteksi
1	[pagipagi udaranya dingin bangetdi pilihan kolam...	Fasilitas & Infrastruktur
2	[berlibur komplit delangan wahana waterboom air...	Fasilitas & Infrastruktur
3	[guci mata air panasya surut sebaikya kesini pa...	Fasilitas & Infrastruktur
4	[tempatnya luasada lokasi pemandianyaada kolam...	Pengalaman Wisatawan
5	[lokasi bersahabat ajak keluarga teman menikmat...	Pengalaman Wisatawan
6	[airnya bersih viewnya bagus memuaskan harga ti...	Pengalaman Wisatawan, Harga & Aksesibilitas
7	[toilet kotor parkir penuh fasilitas kurang	Fasilitas & Infrastruktur
8	[pemandangan indah udara sejuk nyaman sekali	Pengalaman Wisatawan
9	[tiket murah worth it untuk keluarga	Harga & Aksesibilitas
10	[jalan menuju lokasi rusak berkelok tanjakan	Harga & Aksesibilitas

3.6.1. Distribusi Sentimen per Aspek

Analisis distribusi sentimen per aspek dilakukan untuk mengidentifikasi aspek mana yang paling banyak mendapat perhatian pengunjung serta kecenderungan sentimen pada setiap aspek tersebut. Dari total 1.127 ulasan yang

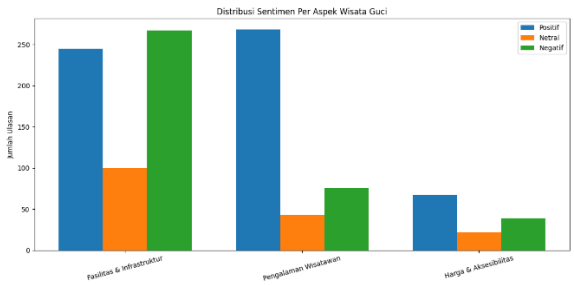
dianalisis, diperoleh distribusi pada Tabel 12 sebagai berikut.

Tabel 12. Distribusi Sentimen per Aspek

Aspek	Total Ulasan	Positif	%	Negatif	%	Netral	%
Fasilitas & Infrastruktur	612	245	40.0%	267	43.6%	100	16.4%
Pengalaman Wisatawan	387	268	69.2%	76	19.6%	43	11.1%
Harga & Aksesibilitas	128	67	52.3%	39	30.5%	22	17.2%
Total	**1.127**	**432**	**38.3%**	**404**	**35.8%**	**291**	**25.8%**

Catatan: Total ulasan per aspek dapat melebihi 1.127 karena satu ulasan dapat membahas lebih dari satu aspek

Visualisasi distribusi sentimen per aspek ditampilkan pada Gambar 6.



Gambar 6. Visualisasi distribusi sentimen per aspek

Dari data Gambar 6, dapat diidentifikasi beberapa temuan penting:

a. Aspek Fasilitas & Infrastruktur merupakan aspek yang paling sering dibahas dalam ulasan (612 ulasan atau 54.3% dari total *dataset*). Hal ini mengindikasikan bahwa pengunjung sangat memperhatikan kondisi fasilitas fisik destinasi wisata.

b. Sentimen negatif pada aspek Fasilitas & Infrastruktur mendominasi dengan 43.6%, menunjukkan adanya gap antara harapan pengunjung dengan kondisi fasilitas yang tersedia.

c. Aspek Pengalaman Wisatawan memiliki sentimen paling positif (69.2%), yang mengindikasikan bahwa daya tarik alam dan pengalaman keseluruhan di Wisata Guci mendapat apresiasi tinggi dari pengunjung.

d. Aspek Harga & Aksesibilitas menunjukkan sentimen yang relatif seimbang (52.3% positif vs 30.5% negatif), menandakan bahwa harga tiket dinilai wajar meskipun masih ada keluhan terkait akses jalan dan biaya tambahan.

e. Perbedaan distribusi sentimen antar aspek ini memberikan *insight* strategis bahwa meskipun destinasi memiliki daya tarik alam yang kuat, kualitas pengalaman pengunjung secara keseluruhan dapat ditingkatkan signifikan melalui perbaikan infrastruktur dan fasilitas pendukung.

3.6.2. Kata Kunci Dominan per Aspek

Analisis kata kunci dominan dilakukan dengan menghitung frekuensi kemunculan kata-kata spesifik dalam ulasan yang telah dikategorikan per aspek dan sentimen. Metode yang digunakan adalah TF (*Term Frequency*) untuk mengidentifikasi kata-kata yang paling representatif untuk setiap kombinasi aspek-sentimen. Dan ditampilkan pada Tabel 13.

Metodologi Ekstraksi Kata Kunci:

a. Filter ulasan berdasarkan aspek dan sentimen

b. Hitung frekuensi setiap kata setelah *stopword removal*

c. Identifikasi top 15 kata dengan frekuensi tertinggi

d. Analisis konteks penggunaan kata-kata tersebut

Tabel 13. Kata Kunci Dominan per Aspek

Rank	Sentimen Negatif	*Freq*	Sentimen Positif	*Freq*	Sentimen Netral	*Freq*
1	toilet	89	kolam	76	kolam	34
2	parkir	67	waterboom	58	ada	28
3	kotor	58	luas	52	fasilitas	26
4	penginapan	54	lengkap	48	toilet	22
5	bau	47	banyak	45	parkir	19
6	kurang	45	Bersih	42	tersedia	18

Rank	Sentimen Negatif	*Freq*	Sentimen Positif	*Freq*	Sentimen Netral	*Freq*
7	gazebo	42	seru	38	penginapan	16
8	terbatas	38	anak	35	area	15
9	tua	35	air	33	tempat	14
10	terawat	32	panas	31	gazebo	13

3.6.3. Implikasi Praktis Hasil ABSC

Hasil ABSC memberikan implikasi praktis yang penting bagi pengelola Wisata Pemandian Air Panas Guci dalam merumuskan kebijakan peningkatan layanan. Pertama, tingginya sentimen negatif pada aspek Fasilitas & Infrastruktur menegaskan perlunya prioritas dalam peningkatan dan pemeliharaan fasilitas wisata, seperti kebersihan toilet, kelayakan area parkir, kondisi jalan internal kawasan wisata, serta kualitas fasilitas penunjang seperti gazebo, warung, dan area tunggu. Peningkatan pada aspek ini dinilai akan memberikan dampak signifikan terhadap kenyamanan dan persepsi pengunjung.

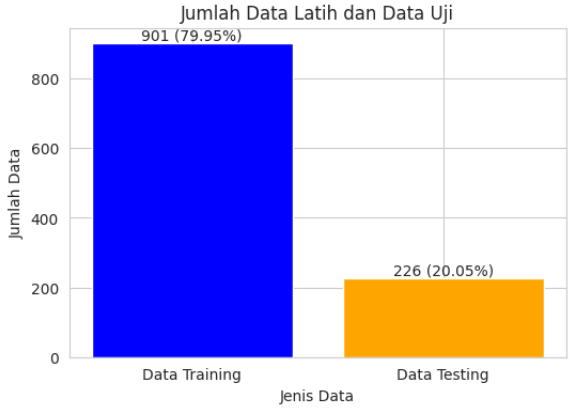
Kedua, dominasi sentimen positif pada aspek Pengalaman Wisatawan memberikan peluang bagi pengelola untuk memperkuat strategi promosi berbasis pengalaman, seperti memaksimalkan potensi keindahan alam, menghadirkan titik foto yang lebih menarik (*photo spot*), dan menyediakan aktivitas wisata bernilai tambah yang dapat memperkaya pengalaman pengunjung.

Ketiga, pada aspek Harga & Aksesibilitas, persepsi yang beragam menunjukkan perlunya transparansi informasi harga, evaluasi biaya tambahan, serta peningkatan kondisi akses jalan menuju lokasi wisata. Perbaikan pada aspek ini dinilai mampu meningkatkan kemudahan akses dan mengurangi keluhan wisatawan.

Secara keseluruhan, temuan ABSC ini memberikan gambaran strategis bahwa meskipun Wisata Guci memiliki daya tarik alam yang kuat, peningkatan kualitas fasilitas fisik dan aksesibilitas merupakan faktor kunci untuk meningkatkan kepuasan pengunjung secara berkelanjutan. Dengan memanfaatkan temuan ini sebagai acuan perbaikan, pengelola wisata dapat menerapkan pendekatan pengembangan layanan yang lebih terarah dan berbasis data.

3.7. *Splitting* Data (80% Latih, 20% Uji)

Dataset yang telah selesai melalui tahap pra-pemrosesan dan pelabelan selanjutnya dibagi menjadi dua *subset*, yaitu data latih sebesar 80% dan data uji sebesar 20% menggunakan fungsi *train_test_spli*t dari pustaka *scikit-learn*. Pembagian dilakukan secara *stratified*, sehingga proporsi tiap kelas sentimen (positif, negatif, netral) tetap seimbang pada kedua *subset*. Data latih digunakan untuk melatih model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi performa model secara objektif. Distribusi hasil dari pembagian data ditunjukkan pada Gambar 7.



Gambar 7. Jumlah Data Latih dan Data Uji

3.8. Klasifikasi

Tahap klasifikasi dilakukan untuk membangun model yang mampu memprediksi sentimen ulasan secara otomatis. Dua algoritma digunakan dalam penelitian ini, yaitu:

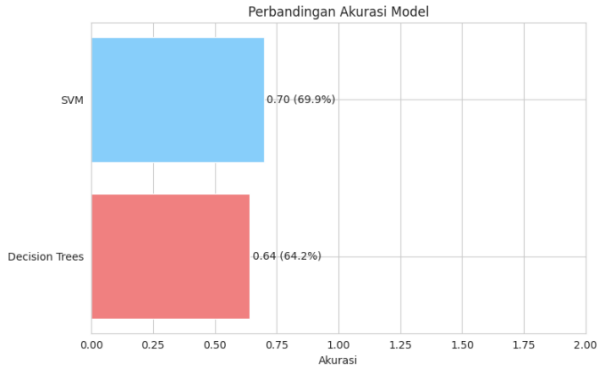
3.8.1. *Suport Vector Machine (*SVM*)*

SVM merupakan algoritma *supervised learning* yang bekerja dengan mencari *hyperplane* terbaik untuk memisahkan kelas sentimen. Pada penelitian ini digunakan *kernel linear*, karena data hasil *vektorisasi* dengan *CountVectorizer* cenderung memiliki pemisahan linier. SVM banyak digunakan pada analisis sentimen karena mampu menangani data berteks berdimensi tinggi secara efektif.

3.8.2. *Decision Tree*

Decision Tree mengklasifikasikan ulasan melalui struktur pohon keputusan, di mana setiap *node* ditentukan berdasarkan nilai *information gain* tertinggi. Pendekatan ini mudah diinterpretasikan karena dapat menunjukkan kata atau fitur apa yang paling memengaruhi penentuan sentimen.

Perbandingan hasil akurasi kedua algoritma dapat dilihat pada Gambar 8, di mana model dengan akurasi tertinggi dianggap paling optimal dalam menyelesaikan klasifikasi sentimen ulasan Wisata Guci.



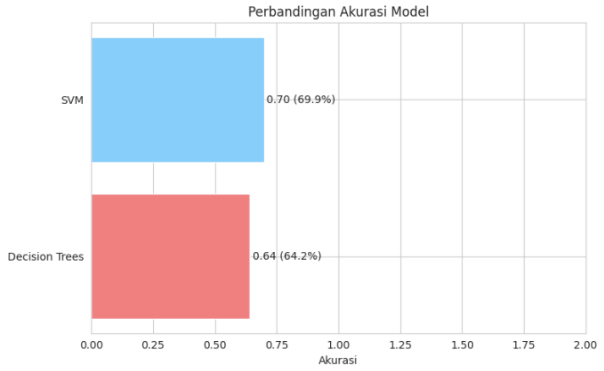
Gambar 8. Perbandingan Hasil Akurasi

3.9. Evaluasi Model (Akurasi, *Confusion Matrix*, *Classification Report*)

Evaluasi model dilakukan menggunakan data uji untuk menilai kemampuan algoritma dalam mengklasifikasikan sentimen secara tepat. Tiga metrik utama digunakan dalam penelitian ini, yaitu:

3.9.1. Akurasi

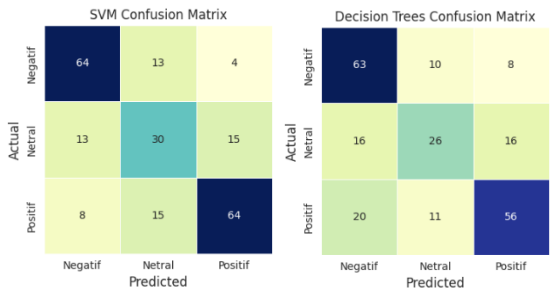
Mengukur persentase prediksi yang benar dari seluruh data uji. Nilai ini memberikan gambaran umum performa model dalam klasifikasi sentimen. Hasil akurasi untuk masing-masing model ditampilkan pada Gambar 9.



Gambar 9. Hasil Akurasi

3.9.2. *Confusion Matrix*

Menunjukkan jumlah prediksi benar dan salah pada tiap kelas sentimen (positif, negatif, netral), sehingga dapat terlihat kelas mana yang paling sulit diprediksi. Visualisasi *confusion matrix* disajikan pada Gambar 10.



Gambar 10. *Confusion Matrix*

3.9.3. *Classication Report*

Berisi metrik *precision*, *recall*, dan *F1-score* untuk setiap kelas, sehingga penilaian performa menjadi lebih komprehensif. Hasil lengkap dapat dilihat pada Gambar 11.

Classification Report for SVM:

	precision	recall	f1-score	support
Negatif	0.753	0.790	0.771	81.000
Netral	0.517	0.517	0.517	58.000
Positif	0.771	0.736	0.753	87.000
accuracy	0.699	0.699	0.699	0.699
macro avg	0.680	0.681	0.680	226.000
weighted avg	0.699	0.699	0.699	226.000

Classification Report for Decision Trees:

	precision	recall	f1-score	support
Negatif	0.636	0.778	0.700	81.000
Netral	0.553	0.448	0.495	58.000
Positif	0.700	0.644	0.671	87.000
accuracy	0.642	0.642	0.642	0.642
macro avg	0.630	0.623	0.622	226.000
weighted avg	0.640	0.642	0.636	226.000

Gambar 11. *Classification Report*

Berdasarkan hasil evaluasi pada Gambar 11, model SVM menunjukkan performa terbaik dibandingkan D*ecision Tree*, dengan akurasi sebesar 69,9% dan nilai *F1-Score* sebesar 69,9%. Hasil ini menunjukkan bahwa algoritma SVM lebih efektif dalam mengklasifikasikan sentimen ulasan Wisata Guci dibandingkan *Decision Tree* yang memiliki akurasi 64,2% dan F1-Score 63,6%. Dengan demikian, SVM dinilai lebih optimal dalam memahami pola teks dan menentukan polaritas opini dari ulasan pengguna.

3.10. Visualisasi

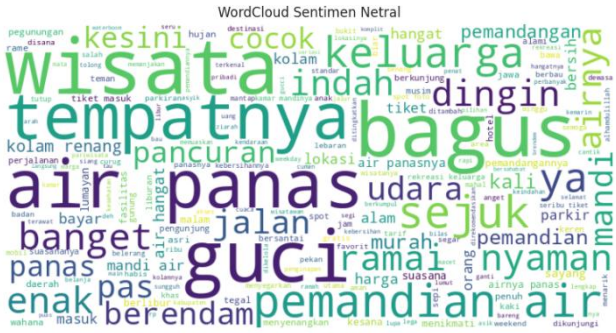
Tahap visualisasi dilakukan untuk memahami pola opini wisatawan secara lebih intuitif dan mendalam. Tiga jenis visualisasi digunakan dalam penelitian ini:

3.10.1. *WordCloud*

Visualisasi *WordCloud* pada Gambar 12, 13, dan 14 menampilkan kata-kata yang paling sering muncul dalam ulasan, di mana ukuran kata menunjukkan tingkat frekuensi kemunculannya. Melalui visualisasi ini, tema utama yang sering dibahas oleh pengunjung dapat diidentifikasi dengan mudah, baik pada sentimen positif, negatif, maupun netral, sehingga memberikan gambaran umum persepsi wisatawan terhadap Wisata Pemandian Air Panas Guci.

Gambar 12. *WordCloud* Sentimen Negatif

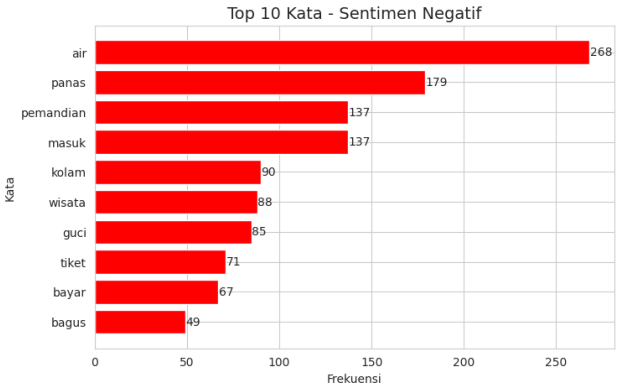
Gambar 13. *WordCloud* Sentimen Positif



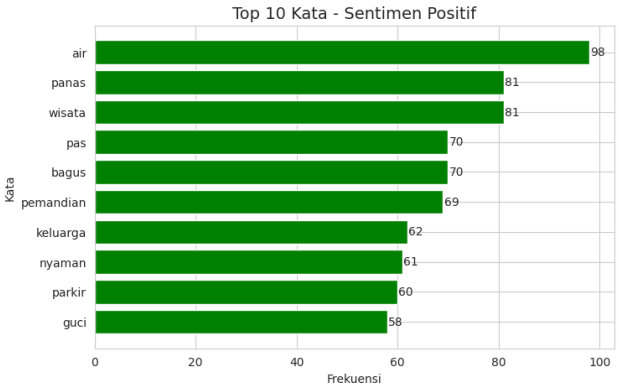
Gambar 14. *WordCloud* Sentimen Netral

3.10.2. Frekuensi Kata

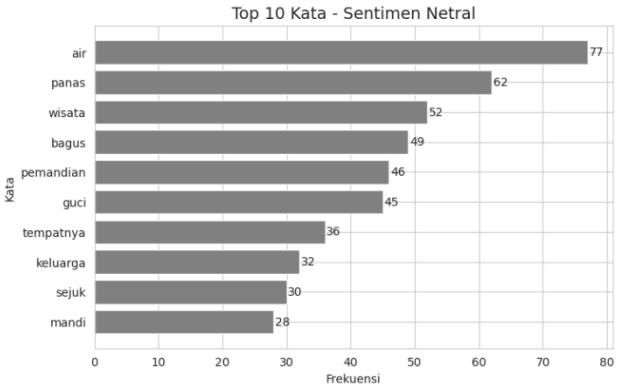
Visualisasi frekuensi kata pada Gambar 15, 16, dan 17 menampilkan daftar kata yang paling banyak digunakan dalam ulasan pada setiap kategori sentimen. Diagram batang ini menunjukkan kata-kata yang memiliki kontribusi besar terhadap pembentukan opini pengunjung, sehingga dapat memberikan informasi yang lebih terarah mengenai faktor yang memengaruhi sentimen positif, negatif, maupun netral pada ulasan Wisata Pemandian Air Panas Guci.



Gambar 15. Frekuensi Kata Negatif



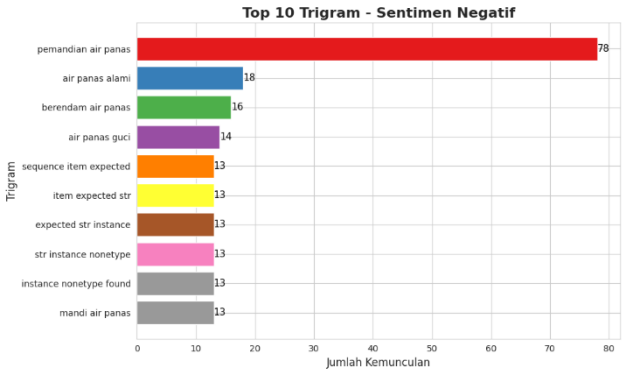
Gambar 16. Frekuensi Kata Positif



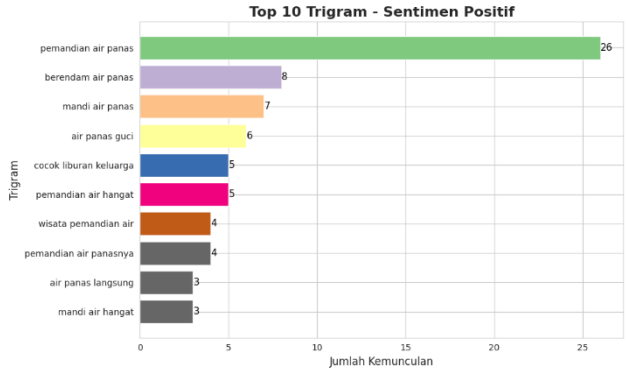
Gambar 17. Frekuensi Kata Netral

3.10.3. *N-Gram*

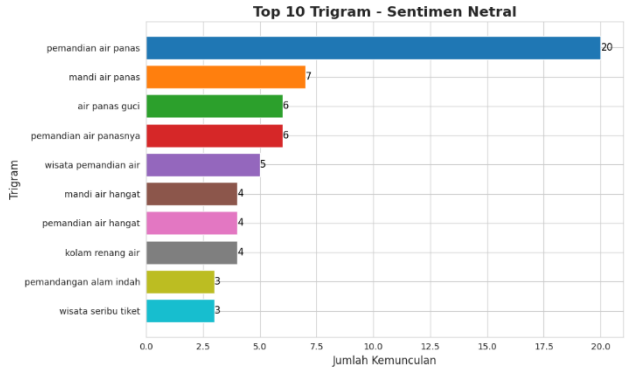
Visualisasi *N-Gram* pada Gambar 18, 19, dan 20 menampilkan pola frasa yang sering muncul secara berurutan dalam ulasan. Dengan menganalisis kombinasi kata dua atau lebih, *N-Gram* memberikan konteks yang lebih jelas terhadap opini yang disampaikan pengunjung. Hasil ini membantu mengidentifikasi ekspresi yang sering digunakan untuk menggambarkan pengalaman wisata di Guci, baik dari sisi fasilitas, pelayanan, maupun kenyamanan selama berkunjung.



Gambar 18. *N-Gram* - Sentimen Negatif



Gambar 19. *N-Gram* - Sentimen Positif



Gambar 20. *N-Gram* - Sentimen Netral

4. KESIMPULAN

Berdasarkan hasil analisis sentimen terhadap ulasan Google Maps pada Wisata Pemandian Air Panas Guci menggunakan algoritma *Support Vector Machine* (SVM) dan *Decision Tree*, dapat disimpulkan bahwa:

a. Mayoritas pengunjung memberikan sentimen positif terhadap wisata Guci, meskipun jumlah sentimen negatif juga cukup tinggi sehingga dapat menjadi masukan bagi pengelola wisata dalam meningkatkan kualitas fasilitas dan pelayanan.

b. Model klasifikasi dengan algoritma *Support Vector Machine* (SVM) menunjukkan performa terbaik dibandingkan *Decision Tree*, dengan akurasi sebesar 69,9% dan *F1-Score* 69,9%, sehingga lebih direkomendasikan untuk digunakan pada analisis sentimen berbasis teks ulasan wisata.

c. Hasil visualisasi *WordCloud*, frekuensi kata, dan *N-Gram* menunjukkan bahwa aspek yang paling sering menjadi sorotan wisatawan meliputi kebersihan, fasilitas, serta harga dan aksesibilitas, yang dapat dijadikan perhatian utama dalam pengembangan destinasi wisata Guci.

Secara keseluruhan, penelitian ini menunjukkan bahwa penerapan analisis sentimen berbasis *machine learning* dapat membantu pengelola dalam memahami persepsi wisatawan dan meningkatkan kualitas layanan pariwisata.

Ucapan Terima Kasih

Penulis G.F.A. mengucapkan terima kasih kepada Universitas Muhammadiyah Tegal dan Program Studi Informatika yang telah memberikan dukungan dalam penyelesaian penelitian ini. Penulis juga menyampaikan apresiasi kepada seluruh pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung, terutama kepada para peninjau dan responden yang memberikan ulasan pada *platform* Google Maps sehingga data dapat digunakan dalam penelitian ini. Segala bentuk bantuan, dukungan moral, serta motivasi yang diberikan telah menjadi bagian penting dalam terselesaikannya penelitian ini.

DAFTAR PUSTAKA

[1] R. Onsu, D. S. Febrian, and F. K. Diane, "Implementasi BI-LSTM dengan Ekstraksi Fitur WORD2VEC untuk Pengembangan Analisis Sentimen Aplikasi Identitas Kependudukan Digital," *Jurnal Teknologi Terpadu*, vol. 10, no. 1, pp. 46–55, 2024.

[2] R. D. R. Apriliansyah, R. Astuti, W. Prihartono, and R. Hamonangan, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Pengunjung di Pantai Kejawanan," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5774.

[3] R. Y. Hayuningtyas and R. Sari, "Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata TMII Menggunakan Naïve Bayes dan PSO," *Jurnal TECHNO Nusa Mandiri*, vol. 16, no. 1, p. 37, 2019, [Online]. Available: http://nusamandiri.ac.id/

[4] O. Somantri and Dairoh, "Analisis sentimen penilaian tempat tujuan wisata Kota Tegal berbasis text mining," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 2, pp. 191–196, 2019.

[5] R. Oktafiani and R. Rianto, "Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree untuk Sistem Rekomendasi Tempat Wisata," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 113–121, Aug. 2023, doi: 10.25077/teknosi.v9i2.2023.113-121.

[6] D. Arianto and I. Budi, "Aspect-based sentiment analysis on Indonesia's tourism destinations based on Google Maps user code-mixed reviews (study case: Borobudur and Prambanan Temples)," in *Proc. 34th Pacific Asia Conf. Language, Information and Computation (PACLIC)*, Hanoi, Vietnam, pp. 359–367, Oct. 2020.

[7] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian*

Journal of Machine Learning and Computer Science, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.

[8] A. R. Hakim, "Analisis Sentimen Objek Wisata di Google Maps Menggunakan Metode Decision Tree," *CBIS JOURNAL*, vol. 12, no. 01, 2024, [Online]. Available: http://ejournal.upbatam.ac.id/index.php/cbishttp://ejournal.upbatam.ac.id/index.php/cbis

[9] A. Ulfah and I. Najiah, "Implementasi Web Scraping pada Situs Jurnal Sinta Menggunakan Framework Selenium Webdriver Python," *JIKA (Jurnal Informatika)*, vol. 7, no. 1, p. 29, Feb. 2023, doi: 10.31000/jika.v7i1.7037.

[10] S. Satriajati, S. B. Panuntun, and S. Pramana, "Implementasi web scraping dalam pengumpulan berita kriminal pada masa pandemi COVID-19 (studi kasus: situs berita detik.com)," in *Proc. Semin. Nas. Official Statistics 2020: Statistics in the New Normal—A Challenge of Big Data and Official Statistics*, pp. 300–308, 2020.

[11] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.

[12] D. Bilianos and G. Mikros, "Sentiment analysis in cross-linguistic context: How can machine translation influence sentiment classification?," *Digital Scholarship in the Humanities*, vol. 38, no. 1, pp. 23–33, Apr. 2023, doi: 10.1093/llc/fqac053.

[13] C. P. Chai, "Comparison of text preprocessing methods," *Nat Lang Eng*, vol. 29, no. 3, pp. 509–553, May 2023, doi: 10.1017/S1351324922000213.

[14] A. Nigam, A. Dhruv, and F. J. Josephin S, "Text Pre-Processing And Feature Extraction Using NLP," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 03, no. 06, pp. 1550–1551, 2021, [Online]. Available: www.irjmets.com

[15] A. Albladi, M. K. Uddin, M. Islam, and C. Seals, "TWSSenti: A Novel Hybrid Framework for Topic-Wise Sentiment Analysis on Social Media Using Transformer Models," *arxiv logo Login*, pp. 1–41, Apr. 2025, [Online]. Available: http://arxiv.org/abs/2504.09896

[16] D. Prasetia, N. Rahaningsih, R. D. Dana, and C. L. Rohmat, "Analisis Sentimen Pengguna Aplikasi Mybluebird dengan Algoritma Naïve Bayes di Playstore," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5687.

[17] T. M. Iryana and P. P. Adikara, "Analisis Sentimen Masyarakat Terhadap Mass Rapid Transit Jakarta Menggunakan Metode Naïve Bayes Dengan Normalisasi Kata," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 6, pp. 2548–964, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[18] S. Ratnaswari, N. C. Wibowo, and D. S. Y. Kartika, "Analisis Sentimen Menggunakan Metode Lexicon-Based Dan Support Vector Machine pada Presiden dan Wakil Presiden Indonesia Periode 2024–2029," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5604.

[19] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, "Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling pada Analisis Sentimen dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 4, pp. 717–723, 2022, doi: 10.25126/jtiik.202294707.

[20] S. A. Nugraha, "Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4949–4957, 2025.

[21] C. A. Bahri and L. H. Suadaa, "Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 1, p. 79, Feb. 2023, doi: 10.22146/ijccs.77354.

[22] S. Fatimah, E. Purwanto, and H. Permatasari, "Analisis Sentimen Twitter Tentang Wisata di Kota Solo Twitter Sentiment Analysis About Tourism in the City of Solo," *Techno.COM*, vol. 22, no. 4, pp. 854–869, 2023.

[23] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: http://j-ptiik.ub.ac.id

[24] D. E. Safitri and A. S. Fitrani, "Implementasi Metode Klasifikasi dengan Algoritma Support Vector Machine Kernel Gaussian RBF untuk Prediksi Partisipasi Pemilu Terhadap Demografi Kota Surabaya," *Indonesian Journal of Business*

Intelligence (IJUBI), vol. 5, no. 1, p. 36, Jun. 2022, doi: 10.21927/ijubi.v5i1.2259.

[25] B. Sugito and M. Iqbal, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Penentuan Bibit Unggul," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 1751–1759, Dec. 2024, doi: 10.47065/bits.v6i3.6199.

[26] D. Hardiansyah, R. A. Aziz, and M. S. Hasibuan, "The Classification Method is Used for Sentiment Analysis in My Telkomsel," *International Journal of Artificial Intelligence Research*, vol. 8, no. 2, p. 169, Dec. 2024, doi: 10.29099/ijair.v8i2.1229.

[27] Henderi, Asro, A. Sulaiman, T. B. Kurniawan, D. A. Dewi, and M. Alqudah, "Utilizing Sentiment Analysis for Reflect and Improve Education in Indonesia," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 189–200, Jan. 2025, doi: 10.47738/jads.v6i1.527.

[28] N. R. A. Putri, T. Trimono, and A. T. Damaliana, "Sentiment Analysis on Digital Korlantas POLRI Application Reviews Using the Distilbert Model," *Journal of Renewable Energy, Electrical, and Computer Engineering*, vol. 4, no. 2, pp. 83–89, Oct. 2024, doi: 10.29103/jreece.v4i2.17197.

[29] J. A. Wibowo, V. C. Mawardi, and T. Sutrisno, "Visualisasi Word Cloud Hasil Analisis Sentimen Berbasis Fitur Layanan Aplikasi Gojek dengan Support Vector Machine," *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 2, no. 1, pp. 61–70, Mar. 2024, doi: 10.24912/jsstk.v2i1.32058.

[30] E. N. Hamdana, A. O. Nur Wardani, and A. R. Tri Hayati Ririd, "Sentiment Analysis of Visitor Reviews on Google Maps at Kampung Coklat Tourism," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 1, p. 274, Mar. 2025, doi: 10.30811/jaise.v5i1.6488.