

**ANALISIS SENTIMEN ULASAN PENGUNJUNG OBJEK
WISATA KAWASAN GUNUNG MERAPI MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh

FERNANDO WIDIANSYAH

19.11.3217

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SENTIMEN ULASAN PENGUNJUNG OBJEK
WISATA KAWASAN GUNUNG MERAPI MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh

FERNANDO WIDIANSYAH

19.11.3217

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN PENGUNJUNG OBJEK WISATA
KAWASAN GUNUNG MERAPI MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE**

yang disusun dan diajukan oleh

Fernando Widiansyah

19.11.3217

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Februari 2026

Dosen Pembimbing,



Anggit Dwi Hartanto, M.Kom
NIK. 190302163

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN PENGUNJUNG OBJEK WISATA
KAWASAN GUNUNG MERAPI MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE**

yang disusun dan diajukan oleh

Fernando Widiandyah

19.11.3217

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192

Irma Rofni Wulandari, S.Pd., M.Eng.
NIK. 190302329

Anggit Dwi Hartanto, S.Kom., M.Kom.
NIK. 190302163



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fernando Widiansyah
NIM : 19.11.3217

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Ulasan Pengunjung Objek Wisata Kawasan Gunung Merapi Menggunakan Metode Support Vector Machine

Dosen Pembimbing : Anggit Dwi Hartanto, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Februari 2026

Yang Menyatakan,



Fernando Widiansyah

HALAMAN PERSEMBAHAN

Alhamdulillah rabbil ‘alamin. Dengan rasa hormat dan penuh syukur, skripsi ini saya persembahkan kepada :

1. Kedua orang tua tercinta, Almarhumah Ibu Sumaryati dan Bapak Suroyo, terima kasih atas rasa cinta, kasih sayang, dan do’a yang tidak pernah putus. Meskipun ibu tidak menyaksikan secara langsung anak-anaknya menapaki langkah demi langkah dalam menempuh pendidikan sarjana, saya persembahkan gelar sarjana ini untuk ibu dan bapak tercinta. Kesabaran bapak dalam mendidik saya sungguh luar biasa. Semoga Allah memberikan kebahagiaan selalu kepada bapak dan ibu baik di dunia maupun di akhirat.
2. Kakak-kakak dan Adikku tersayang, mas Moko, mbak Dian, mbak Nita, dan Tika yang telah memberikan do’a dan dukungan dalam berbagai bidang hingga saat ini.
3. Keluarga besar Trah Harjodinomo dan Trah Pawiroyoso tersayang, yang telah membantu dalam setiap langkah perjalanan di Jogja.
4. Teman-teman yang senantiasa menemani dan berbagi dalam suka maupun duka yang tidak bisa saya sebutkan satu persatu. Terima kasih atas kebersamaan, dukungan, dan pengalaman yang tidak ternilai sepanjang perjalanan ini.

KATA PENGANTAR

Alhamdulillah rabbil ‘alamin. Segala puji penulis haturkan hanya kepada Allah SWT Tuhan seluruh alam, atas berkat rahmat, taufik, dan hidayah-Nya sehingga penulis diberikan kemudahan, kelancaran, kesabaran, dan semangat dalam menyelesaikan Skripsi dengan judul “Analisis Sentimen Ulasan Pengunjung Objek Wisata Kawasan Gunung Merapi Menggunakan Metode Support Vector Machine”.

Skripsi ini tidak akan pernah selesai tanpa adanya pihak-pihak luar biasa dan berjasa dalam memberikan dukungan, motivasi, serta bantuan untuk menyelesaikan tugas akhir ini. Untuk itu, dari lubuk hati yang terdalam penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Eli Pujiastuti, M.Kom selaku Ketua Program Studi S1-Informatika Universita Amikom Yogyakarta.
4. Bapak Anggit Dwi Hartanto, M.Kom selaku dosen pembimbing yang telah memberikan arahan, saran, dan motivasi kepada penulis.
5. Kedua orang tua, kakak, adik, dan keluarga besar tercinta yang telah memberikan do’a dan dukungan sangat besar kepada penulis.
6. Semua pihak yang tidak dapat disebutkan satu persatu yang telah memberikan berbagai dukungan kepada penulis.

Penulis menyadari bahwa dalam penulisan skripsi ini tidak sempurna serta masih memiliki banyak kekurangan sehingga kritik dan saran sangat diharapkan. Semoga skripsi ini bermanfaat bagi penulis pada khususnya dan bagi pembaca pada umumnya.

Yogyakarta, 24 Februari 2026

Penulis

DAFTAR ISI

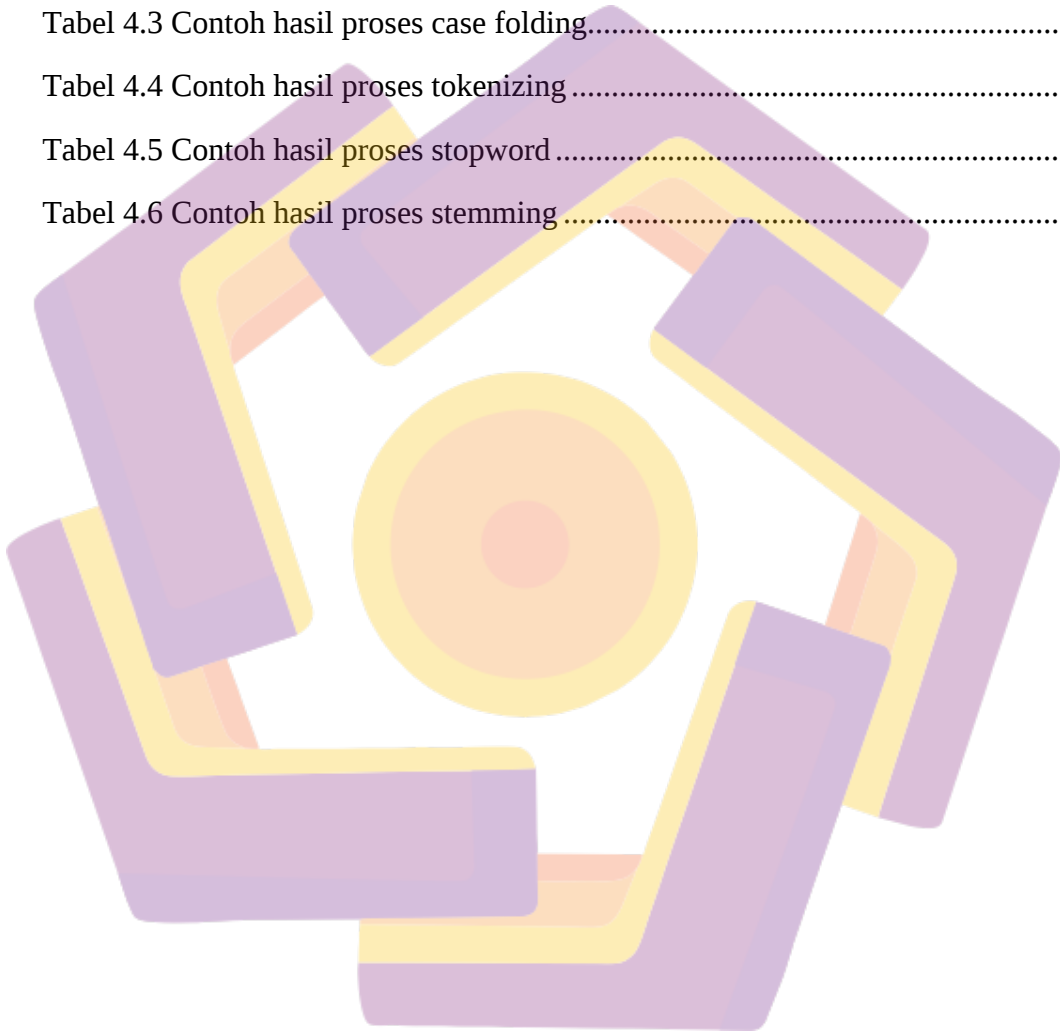
HALAMAN JUDUL	i
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	13
2.2.1 Pariwisata Kawasan Gunung Merapi.....	13
2.2.2 Machine Learning	13
2.2.3 Analisis Sentimen	14

2.2.4 Tripadvisor	14
2.2.5 Web Scraping.....	15
2.2.6 Preprocessing Data.....	15
2.2.7 TF-IDF	16
2.2.8 Text Summary.....	17
2.2.9 Latent Semantic Analysis	17
2.2.10 Support Vector Machine	18
2.2.11 Evaluasi Kinerja Model	19
BAB III METODE PENELITIAN	21
3.1 Objek Penelitian	21
3.2 Alur Penelitian.....	21
1. Pengumpulan Data.....	21
2. Preprocessing Data	22
3. Pelabelan Data	23
4. Pembagian Data	24
5. Ekstraksi Fitur TF-IDF	24
6. Penerapan Latent Semantic Analysis.....	24
7. Pembentukan Model Klasifikasi.....	25
8. Evaluasi Model	25
3.3 Alat dan Bahan	25
3.3.1 Alat.....	25
3.3.2 Bahan	27
BAB IV HASIL DAN PEMBAHASAN	28
4.1 Hasil Pengumpulan Data.....	28
4.2 Import dan Install Library	29

4.3 Translate Data	30
4.4 Tahapan Normalisasi Kata	32
4.5 Preprocessing Data.....	33
4.5.1 Cleaning	34
4.5.2 Case Folding	37
4.5.3 Tokenizing	38
4.5.4 Stopword Removal.....	39
4.5.5 Stemming	40
4.6 Labelling Data.....	42
4.7 Splitting Dataset.....	46
4.8 Ekstraksi Fitur TF-IDF	47
4.9 Pengujian Sistem.....	48
4.9.1 Skenario Pertama (TF-IDF + SVM)	49
4.9.2 Skenario Kedua (TF-IDF + LSA + SVM)	51
BAB V PENUTUP	56
5.1 Kesimpulan.....	56
5.2 Saran.....	57
REFERENSI	59

DAFTAR TABEL

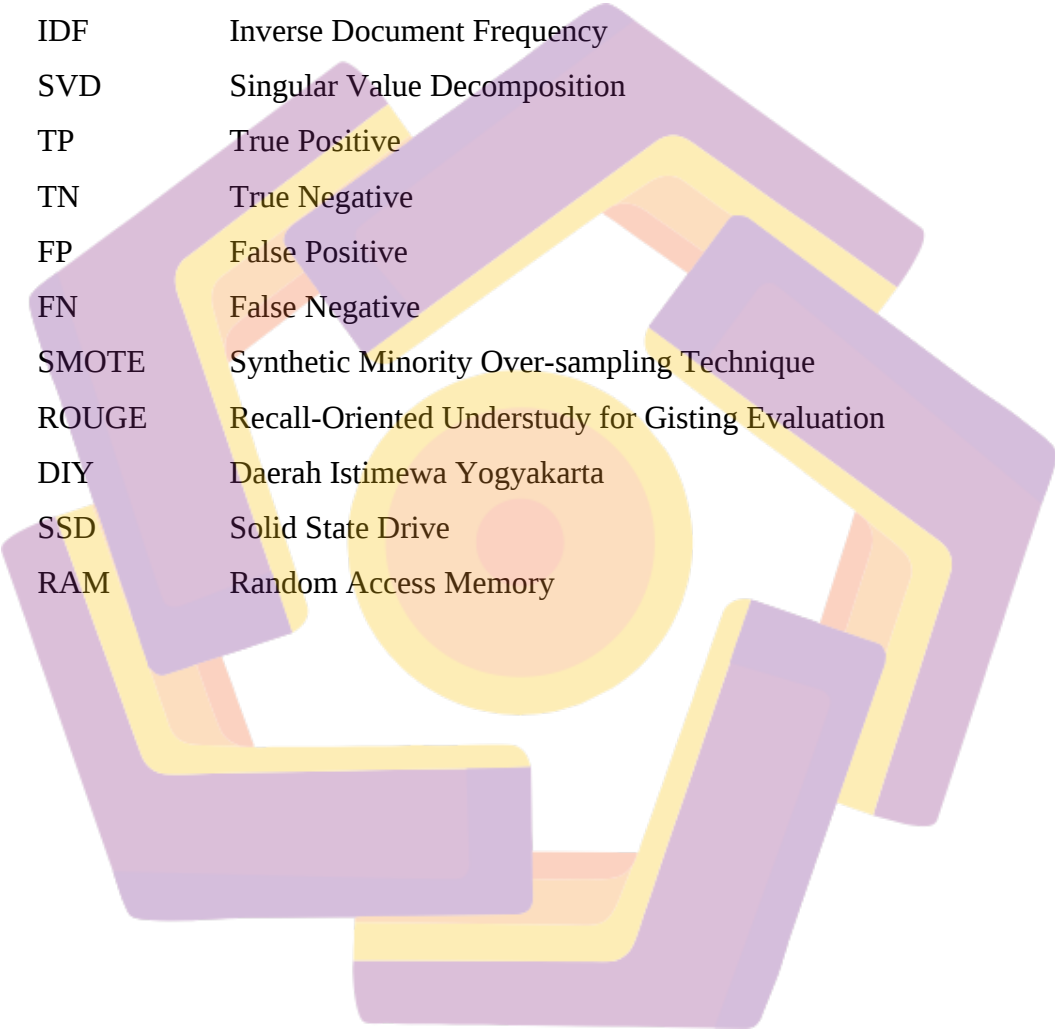
Tabel 2.1 Keaslian Penelitian	8
Tabel 4.1 Contoh hasil proses penerjemahan bahasa.....	32
Tabel 4.2 Contoh hasil proses cleaning	37
Tabel 4.3 Contoh hasil proses case folding.....	38
Tabel 4.4 Contoh hasil proses tokenizing	39
Tabel 4.5 Contoh hasil proses stopwords	40
Tabel 4.6 Contoh hasil proses stemming	41



DAFTAR GAMBAR

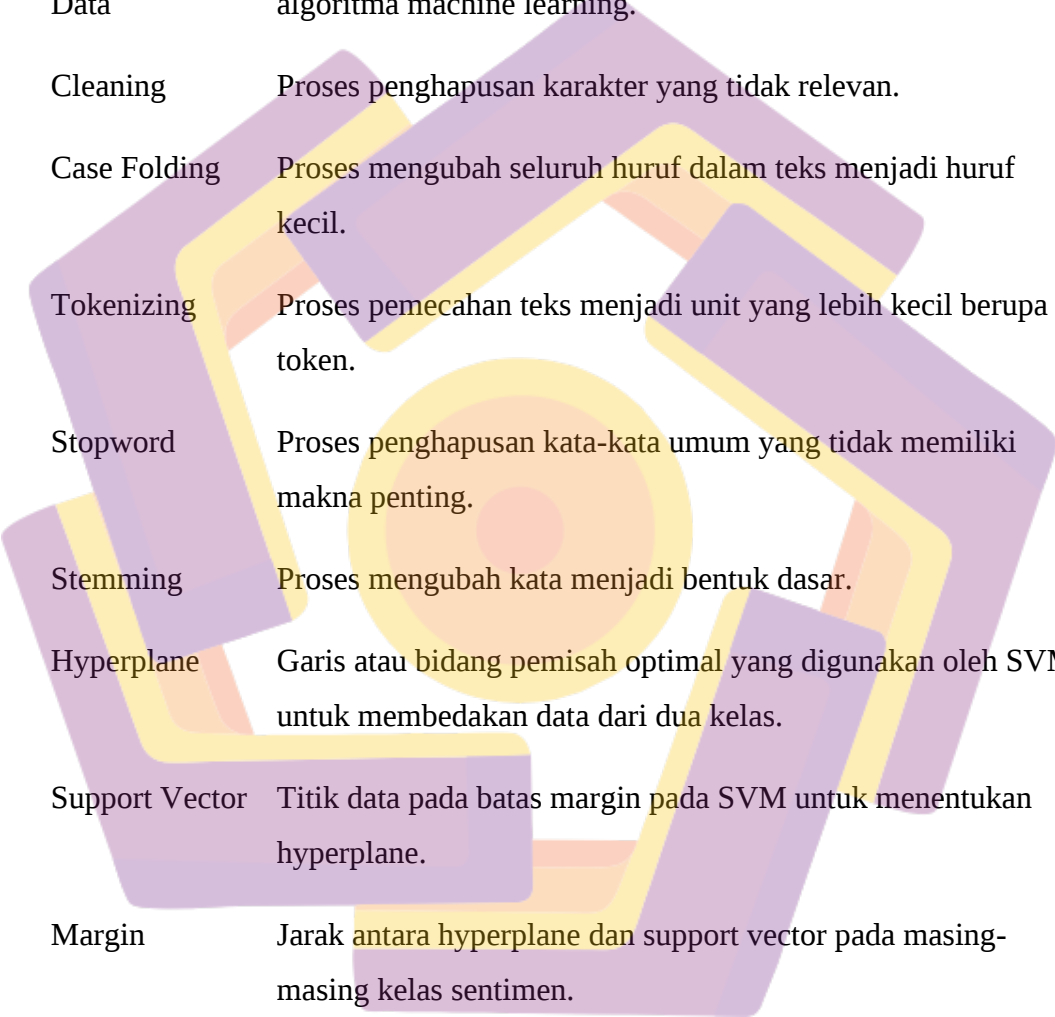
Gambar 2.1 Contoh hyperplane.....	18
Gambar 2.2 Margin maksimal SVM	19
Gambar 2.3 Confusion matrik.....	20
Gambar 3.1 Alur penelitian.....	21
Gambar 4.1 Situs website Tripadvisor Gunung Merapi	28
Gambar 4.2 Scraping data.....	29
Gambar 4.3 Contoh hasil scraping data	29
Gambar 4.4 Visual wordcloud sebelum preprocessing.....	34
Gambar 4.5 Visual wordcloud setelah proses preprocessing.....	41
Gambar 4.6 diagram sepuluh kata terbanyak.....	42
Gambar 4.7 Contoh skor pada sebuah dokumen	45
Gambar 4.8 Contoh hasil proses pelabelan sentimen	45
Gambar 4.9 Distribusi data berlabel pada dataset.....	45
Gambar 4.10 Jumlah pembagian data latih dan uji.....	47
Gambar 4.11 Dimensi fitur matrik hasil TF-IDF	47
Gambar 4.12 Contoh output matrik representasi TF-IDF.....	48
Gambar 4.13 Confusion matrix skenario pertama	51
Gambar 4.14 Dimensi fitur matrik LSA	52
Gambar 4.15 Confusion matrix skenario kedua	54
Gambar 4.16 Wordcloud sentimen positif	55
Gambar 4.17 Wordcloud sentimen negatif	55

DAFTAR LAMBANG DAN SINGKATAN

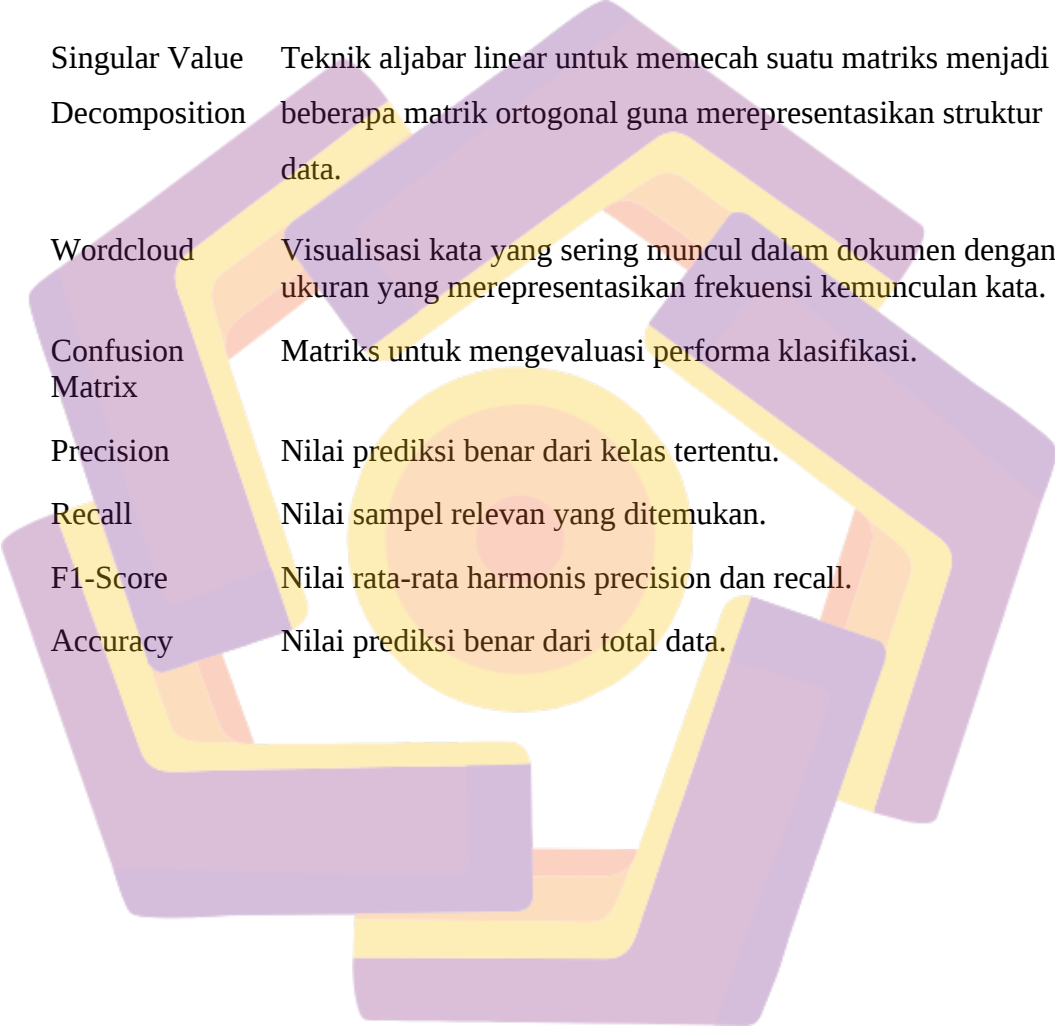


SVM	Support Vector Machines
LSA	Latent Semantic Analysis
TF-IDF	Term Frequency-Inverse Document Frequency
TF	Term Frequency
IDF	Inverse Document Frequency
SVD	Singular Value Decomposition
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
SMOTE	Synthetic Minority Over-sampling Technique
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
DIY	Daerah Istimewa Yogyakarta
SSD	Solid State Drive
RAM	Random Access Memory

DAFTAR ISTILAH



Analisis Sentimen	Proses untuk mengidentifikasi dan mengklasifikasikan opini atau perasaan dalam teks ke dalam kategori sentimen tertentu.
Preprocessing Data	Tahapan awal pengolahan data teks agar dapat diproses oleh algoritma machine learning.
Cleaning	Proses penghapusan karakter yang tidak relevan.
Case Folding	Proses mengubah seluruh huruf dalam teks menjadi huruf kecil.
Tokenizing	Proses pemecahan teks menjadi unit yang lebih kecil berupa token.
Stopword	Proses penghapusan kata-kata umum yang tidak memiliki makna penting.
Stemming	Proses mengubah kata menjadi bentuk dasar.
Hyperplane	Garis atau bidang pemisah optimal yang digunakan oleh SVM untuk membedakan data dari dua kelas.
Support Vector	Titik data pada batas margin pada SVM untuk menentukan hyperplane.
Margin	Jarak antara hyperplane dan support vector pada masing-masing kelas sentimen.
Text Summarization	Proses pemadatan dokumen teks menjadi lebih ringkas dengan tetap mempertahankan informasi utama.
Extractive Summarization	Metode peringkasan dengan memilih kalimat penting dari dokumen tanpa mengubah struktur kalimat aslinya.



Abstractive Summarization	Metode peringkasan dengan membentuk kalimat baru yang lebih ringkas namun tetap memuat informasi penting.
Latent Semantic Analysis	Metode dalam text mining yang memetakan hubungan semantik laten antar kata dan dokumen melalui perhitungan matematis.
Singular Value Decomposition	Teknik aljabar linear untuk memecah suatu matriks menjadi beberapa matrik ortogonal guna merepresentasikan struktur data.
Wordcloud	Visualisasi kata yang sering muncul dalam dokumen dengan ukuran yang merepresentasikan frekuensi kemunculan kata.
Confusion Matrix	Matriks untuk mengevaluasi performa klasifikasi.
Precision	Nilai prediksi benar dari kelas tertentu.
Recall	Nilai sampel relevan yang ditemukan.
F1-Score	Nilai rata-rata harmonis precision dan recall.
Accuracy	Nilai prediksi benar dari total data.

INTISARI

Perkembangan teknologi digital mendorong meningkatnya jumlah ulasan wisatawan pada berbagai platform daring, termasuk TripAdvisor, yang menjadi sumber penting bagi calon pengunjung dalam menentukan pilihan destinasi wisata, khususnya pada kawasan wisata Gunung Merapi. Namun, besarnya volume ulasan menyebabkan proses analisis sentimen secara manual menjadi tidak efektif, sehingga diperlukan pendekatan otomatis untuk mengidentifikasi persepsi positif dan negatif wisatawan. Penelitian ini bertujuan untuk menerapkan metode Support Vector Machine (SVM) dalam mengklasifikasikan sentimen ulasan pengunjung Kawasan Gunung Merapi serta membandingkan kinerja model sebelum dan sesudah penerapan Latent Semantic Analysis (LSA) berbasis Truncated Singular Value Decomposition. Dataset yang digunakan sebanyak 2.264 ulasan berbahasa Indonesia yang diperoleh dari TripAdvisor melalui teknik web scraping. Tahapan penelitian meliputi preprocessing teks, pelabelan sentimen menggunakan pendekatan lexicon-based, pembagian data latih dan uji, ekstraksi fitur menggunakan TF-IDF, serta pengujian model SVM dengan dua skenario, yaitu tanpa LSA dan dengan LSA. Hasil penelitian menunjukkan bahwa kombinasi TF-IDF dan SVM menghasilkan akurasi terbaik sebesar 87,91%, sedangkan penerapan LSA sebelum klasifikasi menghasilkan akurasi sebesar 85,58%, sehingga pada penelitian ini LSA belum mampu meningkatkan kinerja klasifikasi sentimen. Hasil penelitian ini diharapkan dapat dimanfaatkan sebagai referensi pengembangan analisis sentimen pada bidang pariwisata berbasis data ulasan daring.

Kata kunci: analisis sentimen, support vector machine, latent semantic analysis, ulasan wisata, gunung merapi

ABSTRACT

The development of digital technology has driven an increasing number of traveler reviews on various platforms, including TripAdvisor, which has become an important resource for potential visitors in deciding on their destinations, particularly in the Mount Merapi tourist area. However, the large volume of reviews makes manual sentiment analysis ineffective, necessitating an automated approach to identify positive and negative tourist perceptions. This study aims to apply the Support Vector Machine (SVM) method to classify the sentiment of visitor reviews in the Mount Merapi area and compare the model's performance before and after applying Latent Semantic Analysis (LSA) based on Truncated Singular Value Decomposition. The dataset used is 2,264 Indonesian-language reviews obtained from TripAdvisor through web scraping techniques. The research stages include text preprocessing, sentiment labeling using a lexicon-based approach, training and testing data partitioning, feature extraction using TF-IDF, and testing the SVM model in two scenarios: without LSA and with LSA. The results showed that the combination of TF-IDF and SVM produced the best accuracy of 87.91%, while applying LSA before classification produced an accuracy of 85.58%. Therefore, in this study, LSA was unable to improve sentiment classification performance. The results of this study are expected to be used as a reference for developing sentiment analysis in the tourism sector based on online review data.

Keyword: sentiment analysis, support vector machine, latent semantic analysis, tourism reviews, mount merapi