

**ANALISIS DINAMIKA SENTIMEN ULASAN DARING DAN
KORELASINYA TERHADAP PENDAPATAN PARIWISATA
MENGUNAKAN SVM DENGAN OPTIMASI ADASYN-PSO**

SKRIPSI

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



disusun oleh

IKHWAN NURRAMADHAN

22.11.5129

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS DINAMIKA SENTIMEN ULASAN DARING DAN
KORELASINYA TERHADAP PENDAPATAN PARIWISATA
MENGUNAKAN SVM DENGAN OPTIMASI ADASYN-PSO**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

IKHWAN NURRAMADHAN

22.11.5129

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS DINAMIKA SENTIMEN ULASAN DARING DAN
KORELASINYA TERHADAP PENDAPATAN PARIWISATA
MENGUNAKAN SVM DENGAN OPTIMASI ADASYN-PSO**

yang disusun dan diajukan oleh

Ikhwan Nurramadhan

22.11.5129

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Februari 2026

Dosen Pembimbing,


Nuri Cahyono, M.Kom.

NIK. 190302278

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS DINAMIKA SENTIMEN ULASAN DARING DAN
KORELASINYA TERHADAP PENDAPATAN PARIWISATA
MENGUNAKAN SVM DENGAN OPTIMASI ADASYN-PSO**

yang disusun dan diajukan oleh

Ikhwan Nurramadhan

22.11.5129

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Muhammad Rudyanto Arief, S.T., M.T.
NIK. 190302098

Ade Pujiyanto, M.Kom.
NIK. 190302494

Nuri Cahyono, M.Kom.
NIK. 190302278

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ikhwan Nurramadhan
NIM : 22.11.5129

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Dinamika Sentimen Ulasan Daring dan Korelasinya Terhadap Pendapatan Pariwisata Menggunakan SVM dengan Optimasi ADASYN-PSO

Dosen Pembimbing : Nuri Cahyono, 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, 24 Februari 2026

Yang Menyatakan,



Ikhwan Nurramadhan

HALAMAN PERSEMBAHAN

...فَإِذَا عَزَمْتَ فَتَوَكَّلْ عَلَى اللَّهِ إِنَّ اللَّهَ يُحِبُّ الْمُتَوَكِّلِينَ

“...Kemudian, apabila kamu telah membulatkan tekad, bertawakallah kepada Allah. Sesungguhnya Allah mencintai orang-orang yang bertawakal”

(QS. Ali ‘Imran: 159)

Puji syukur senantiasa penulis panjatkan kepada Allah SWT atas semua rahmat, hidayah, dan kekuatan-Nya sehingga karya sederhana ini dapat diselesaikan tepat waktu. Shalawat serta salam senantiasa tercurahkan kepada junjungan agung, Nabi Muhammad SAW, beserta keluarga, sahabat, dan umatnya yang senantiasa menantikan syafaatnya di yaumul akhir.

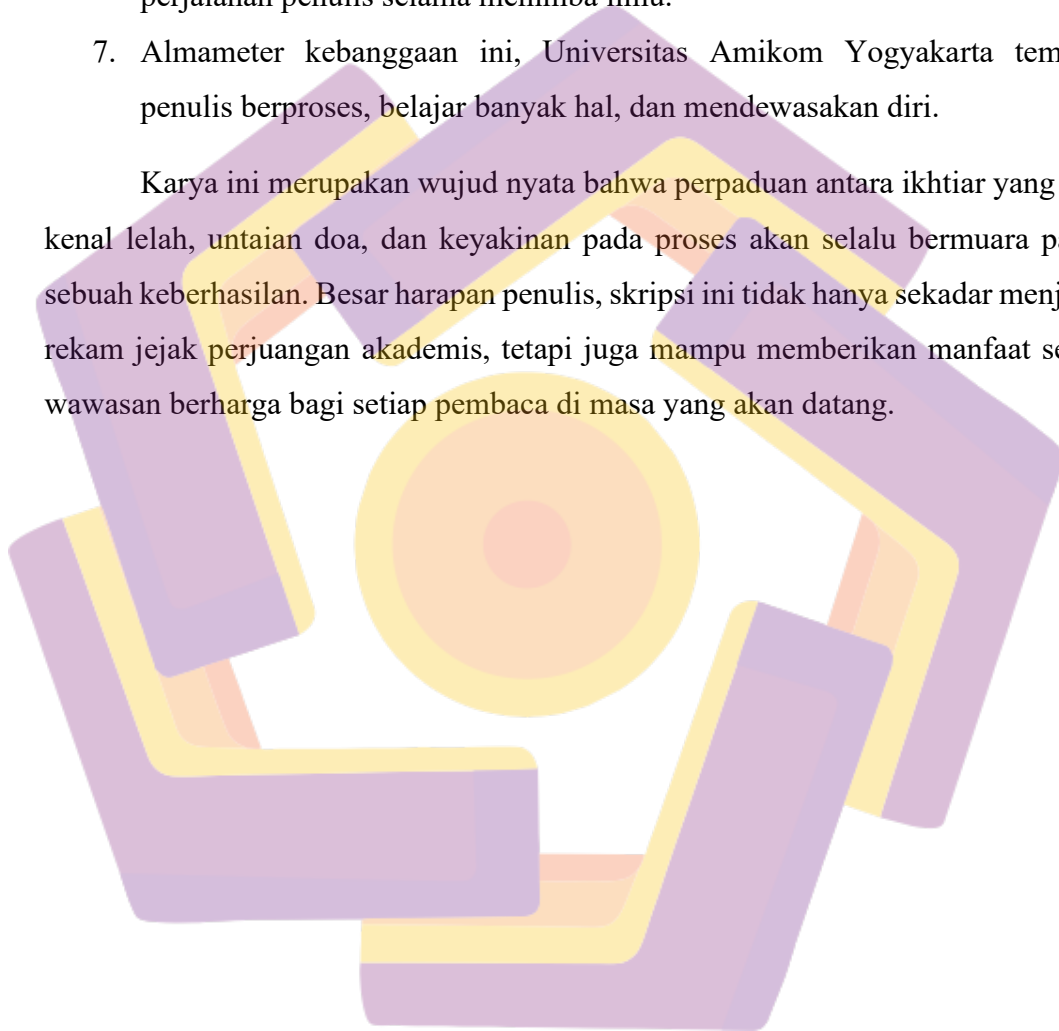
Dengan segala kerendahan hati dan rasa syukur yang mendalam, skripsi ini penulis persembahkan kepada :

1. Kedua orang tua penulis, Bapak Mulyanto dan Ibu Lilis Suryani tercinta. Terimakasih banyak atas doa yang tidak pernah terputus, usaha dan tetes keringat yang diberikan, dan kasih sayang yang tak terhingga. Karya ini adalah bentuk bakti kecil dan rasa terimakasih penulis atas segala usaha dan pengorbanan yang telah diberikan demi pendidikan dan karir penulis.
2. Adik tersayang, Aisyah Nur Ramadhani yang selalu memberikan semangat dan doa yang terbaik disetiap langkah penulis dimanapun berada.
3. Keluarga besar yang selalu memberi nasihat dan motivasi hidup kepada penulis sehingga tidak hilang arah.
4. Teruntuk Dyah Pramudya Sari, seseorang yang istimewa dan insyaallah menjadi calon teman hidup penulis. Terimakasih atas ketulusan hati yang selalu mendampingi dalam setiap suka maupun duka, serta tanpa henti memberikan bantuan, waktu, tenaga, dan dukungan moral sehingga naskah skripsi ini dapat terselesaikan dengan baik dan tepat waktu.
5. Sahabat-sahabat terdekat penulis, yang senantiasa kebersamai dan menjadi tempat bertukar pikiran. Terima kasih atas bantuan, kesabaran

dalam berdiskusi, serta kebersamaan dalam melewati berbagai rintangan hingga tugas akhir ini selesai.

6. Teman- teman seperjuangan, baik di lingkup kampus maupun di tempat lain yang tidak dapat penulis sebutkan satu per satu. Terimakasih atas solidaritas, canda tawa, serta kenangan berharga yang telah mewarnai perjalanan penulis selama menimba ilmu.
7. Almameter kebanggaan ini, Universitas Amikom Yogyakarta tempat penulis berproses, belajar banyak hal, dan mendewasakan diri.

Karya ini merupakan wujud nyata bahwa perpaduan antara ikhtiar yang tak kenal lelah, untaian doa, dan keyakinan pada proses akan selalu bermuara pada sebuah keberhasilan. Besar harapan penulis, skripsi ini tidak hanya sekedar menjadi rekam jejak perjuangan akademis, tetapi juga mampu memberikan manfaat serta wawasan berharga bagi setiap pembaca di masa yang akan datang.



KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala limpahan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyusun dan menyelesaikan skripsi yang berjudul **“Analisis Dinamika Sentimen Ulasan Daring dan Korelasinya Terhadap Pendapatan Pariwisata Menggunakan SVM dengan Optimasi ADASYN-PSO”**. Skripsi ini disusun sebagai salah satu syarat akademik untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penulis menyadari sepenuhnya bahwa dalam penulisan naskah skripsi ini masih terdapat banyak kekurangan dan jauh dari kata sempurna. Hal ini tidak lepas dari keterbatasan ilmu, pengetahuan, serta pengalaman yang penulis miliki. Oleh karena itu, penulis sangat terbuka dan mengharapkan segala bentuk kritik maupun saran yang membangun demi perbaikan karya tulis ini dan penelitian-penelitian selanjutnya.

Proses penyelesaian skripsi ini tentu tidak lepas dari berbagai kendala. Namun, berkat dukungan, bimbingan, serta kerja sama dari berbagai pihak skripsi ini akhirnya dapat diselesaikan dengan baik. Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya dan penghargaan yang setinggi-tingginya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Bapak Nuri Cahyono, M.Kom., selaku Dosen Pembimbing Skripsi yang dengan penuh kesabaran dan keikhlasan telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta motivasi yang sangat berharga selama proses penyusunan skripsi ini.

5. Seluruh jajaran Dosen dan Staf Universitas Amikom Yogyakarta, khususnya di lingkungan Program Studi Informatika, yang telah memberikan ilmu yang tak ternilai harganya serta membantu kelancaran urusan administratif.
6. Kedua orang tua dan keluarga tercinta, yang tiada henti memberikan iringan doa, kasih sayang, serta dukungan baik secara moril maupun finansial selama penulis menempuh pendidikan hingga selesainya tugas akhir ini.
7. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan doa dan dukungan nyata dalam penyelesaian tugas akhir ini.

Semoga Allah SWT senantiasa membalas segala kebaikan dan bantuan yang telah diberikan dengan pahala yang berlipat ganda. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat yang nyata bagi penulis pribadi, pengembangan ilmu pengetahuan, dan para pembaca pada umumnya.

Yogyakarta, 21 Februari 2026

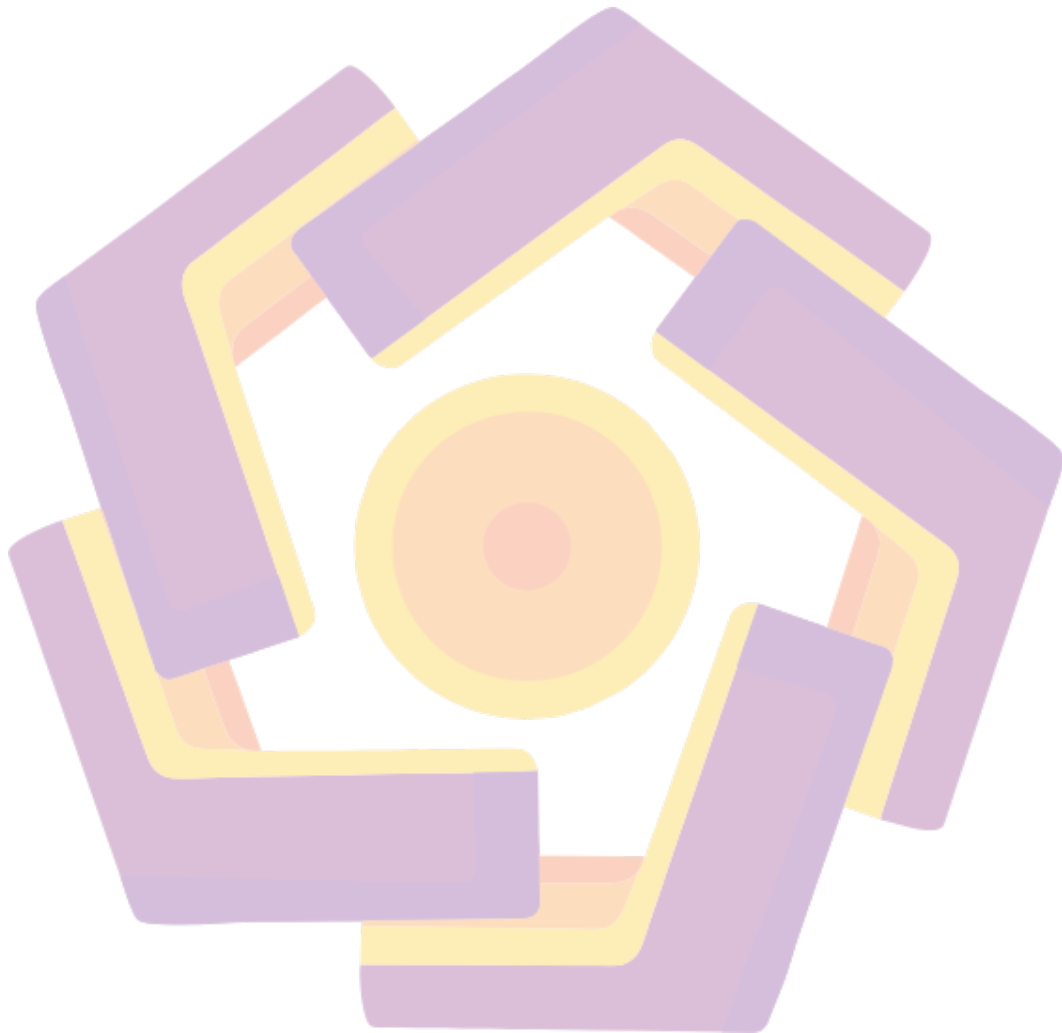
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xviii
INTISARI	xxv
<i>ABSTRACT</i>	xxvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5

BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	13
BAB III METODE PENELITIAN	27
3.1 Objek Penelitian.....	27
3.2 Alur Penelitian	28
3.3 Alat dan Bahan Penelitian.....	35
BAB IV HASIL DAN PEMBAHASAN	38
4.1 Pengumpulan Data	38
4.2 <i>Data Preparation</i>	40
4.3 <i>Preprocessing Data</i>	42
4.4 Pelabelan Data	45
4.5 <i>Exploratory Data Analysis (EDA)</i>	47
4.6 Pembagian Data (<i>Data Splitting</i>).....	50
4.7 Ekstraksi Fitur (<i>Feature Extraction</i>).....	51
4.8 Skenario Validasi Model (Eksperimen).....	52
4.9 Optimasi <i>Hyperparameter</i> dengan PSO	59
4.10 Evaluasi Model Final	62
4.11 Prediksi Sentimen Tahunan dan Analisis Korelasi	63
4.12 Implementasi Sistem (<i>Deployment</i>).....	68
4.13 Pembahasan Hasil Penelitian	72
BAB V PENUTUP	76

5.1	Kesimpulan	76
5.2	Saran	77
REFERENSI		79
LAMPIRAN.....		85



DAFTAR TABEL

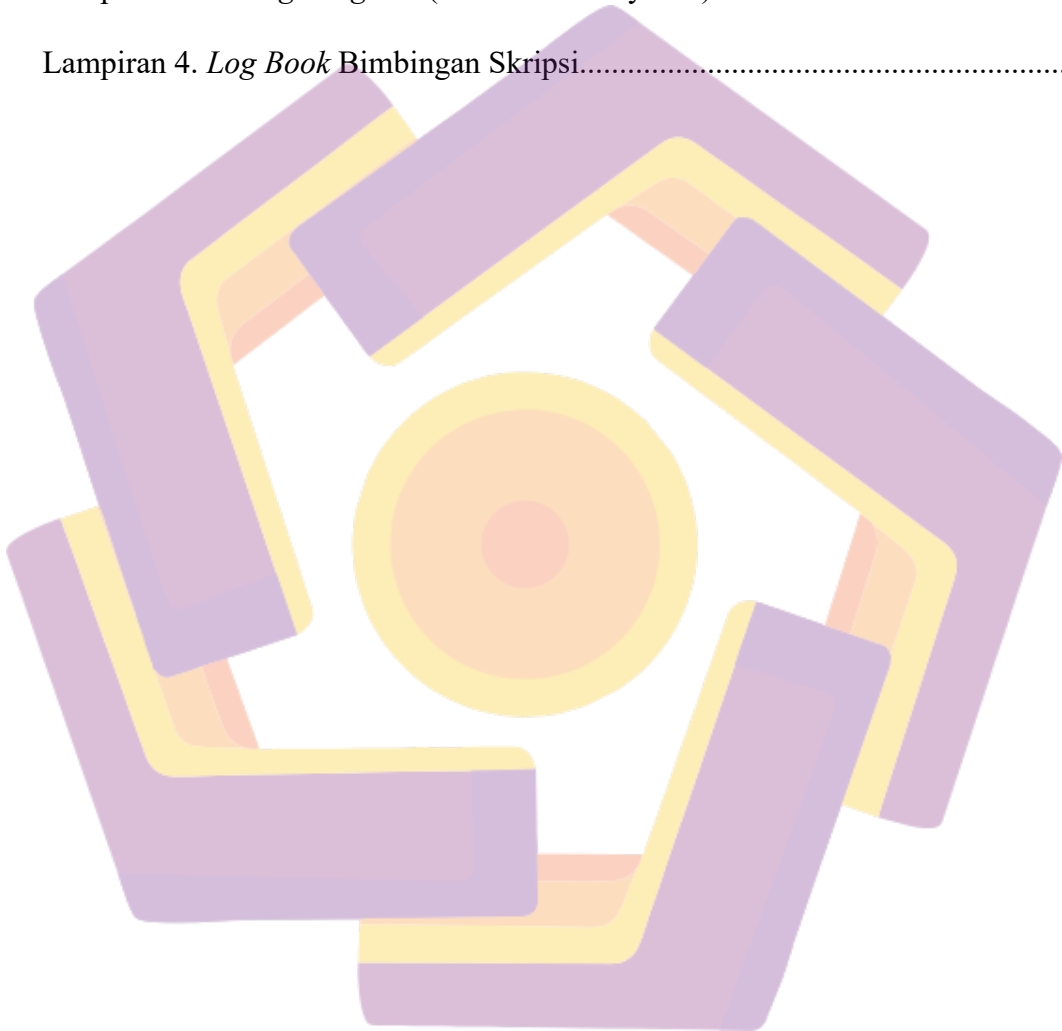
Tabel 2.1. Keaslian Penelitian	8
Tabel 2.2. Confusion Matrix	23
Tabel 2.3. Pedoman Interpretasi Koefisien Korelasi	25
Tabel 3.1. Data Pendapatan Ekowisata Sungai Mudal (2019-2024)	27
Tabel 4.1. Profil <i>Dataset</i> Ulasan Ekowisata Sungai Mudal.....	39
Tabel 4.2. Hasil Pengecekan <i>Missing Values</i>	41
Tabel 4.3. Ilustrasi Hasil Tiap Tahapan <i>Preprocessing</i>	44
Tabel 4.4. Statistik <i>Dataset</i> Final.....	46
Tabel 4.5. Sampel Data Hasil Pelabelan Biner	47
Tabel 4.6. Distribusi Data Latih dan Data Uji	51
Tabel 4.7. Transformasi Dimensi Data (TF-IDF)	52
Tabel 4.8. Hasil Perbandingan Algoritma (Tanpa <i>Balancing</i>).....	53
Tabel 4.9. Hasil Perbandingan Metode Ekstraksi Fitur	55
Tabel 4.10. Hasil Perbandingan Teknik <i>Balancing</i>	57
Tabel 4.11. Peningkatan Performa Model Pasca-Optimasi PSO	61
Tabel 4.12. Performa Model Akhir pada Data Uji.....	63
Tabel 4.13. Rekapitulasi Distribusi Sentimen (Hasil Prediksi)	64
Tabel 4.14. Data Komparasi Sentimen vs Pendapatan Rata-rata.....	66
Tabel 4.15. Perbandingan Metodologi dan Hasil dengan Penelitian Terdahulu....	73

DAFTAR GAMBAR

Gambar 2.1. Ilustrasi Tingkatan Analisis Sentimen pada Ulasan Wisata.....	15
Gambar 2.2. Konsep <i>Hyperplane</i> dan <i>Margin</i> pada SVM.....	21
Gambar 3.1. Alur Penelitian	29
Gambar 4.1. Struktur <i>Dataset</i> Hasil <i>Scraping</i>	40
Gambar 4.2. Distribusi Sentimen Awal (3 Kelas)	45
Gambar 4.3. Distribusi Sentimen Final (Biner)	46
Gambar 4.4. Visualisasi WordCloud Sentimen Positif (Kiri) dan Non-Positif (Kanan)	48
Gambar 4.5. Tren Jumlah Ulasan Pengunjung (2019-2024)	59
Gambar 4.6. Grafik Perbandingan F1-Score pada Kelas Minoritas	55
Gambar 4.7. Peningkatan <i>Recall</i> Kelas Minoritas Pasca- <i>Balancing</i>	58
Gambar 4.8. Grafik Konvergensi Penurunan <i>Loss</i> pada PSO.....	60
Gambar 4.9. Confusion Matrix Model Final	62
Gambar 4.10. Tren Distribusi Sentimen Ulasan per Tahun (2019-2024).....	65
Gambar 4.11. Grafik Korelasi Tren Sentimen vs Pendapatan Rata-rata	67
Gambar 4.12. Tampilan Antarmuka Halaman Utama dan <i>Sidebar</i>	70
Gambar 4.13. Visualisasi Sentimen dan Rekomendasi Manajerial (<i>Insight</i>)	71
Gambar 4.14. Antarmuka Uji Korelasi Sentimen dan Pendapatan.....	72

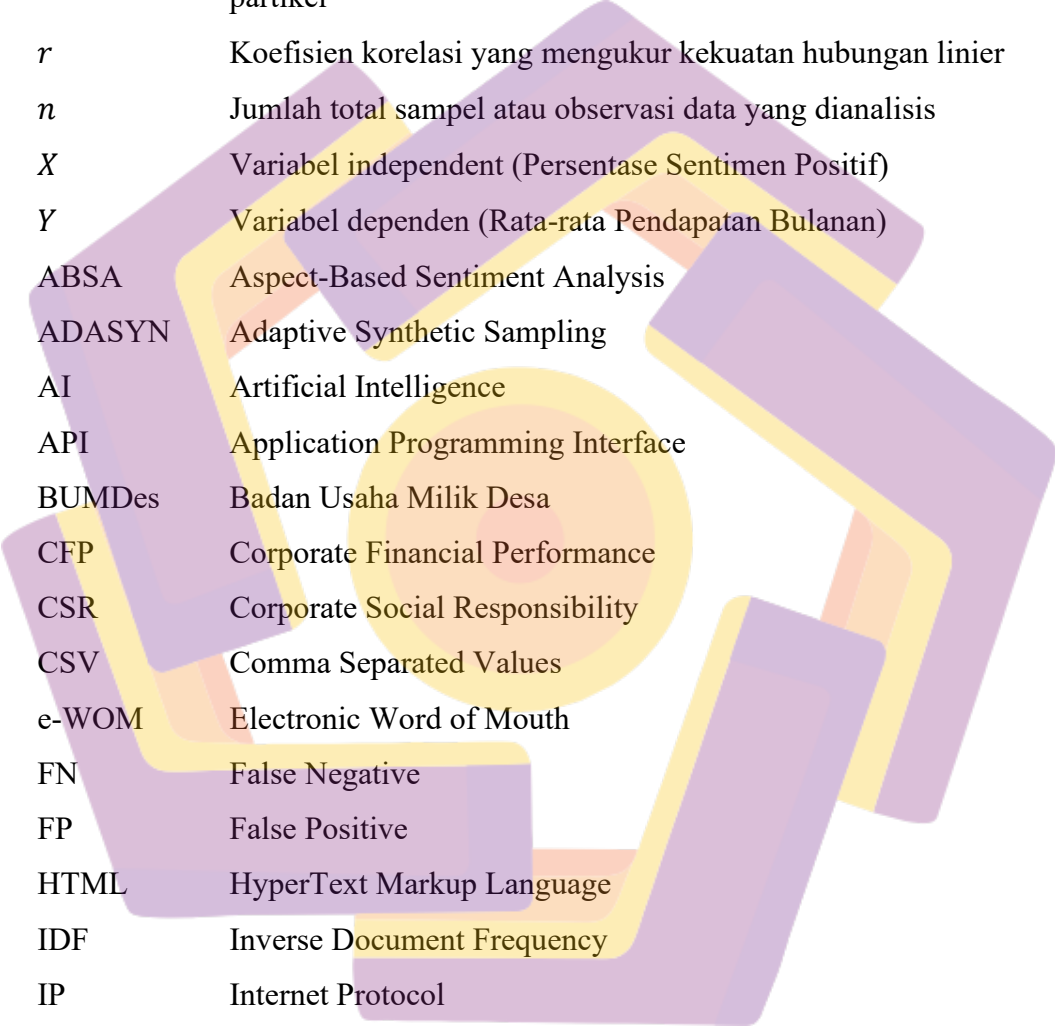
DAFTAR LAMPIRAN

Lampiran 1. Dokumentasi Ekstraksi Data	85
Lampiran 2. Hasil Data Teks (<i>Preprocessing</i>)	87
Lampiran 3. <i>Listing Program</i> (Source Code Python).....	88
Lampiran 4. <i>Log Book</i> Bimbingan Skripsi.....	98



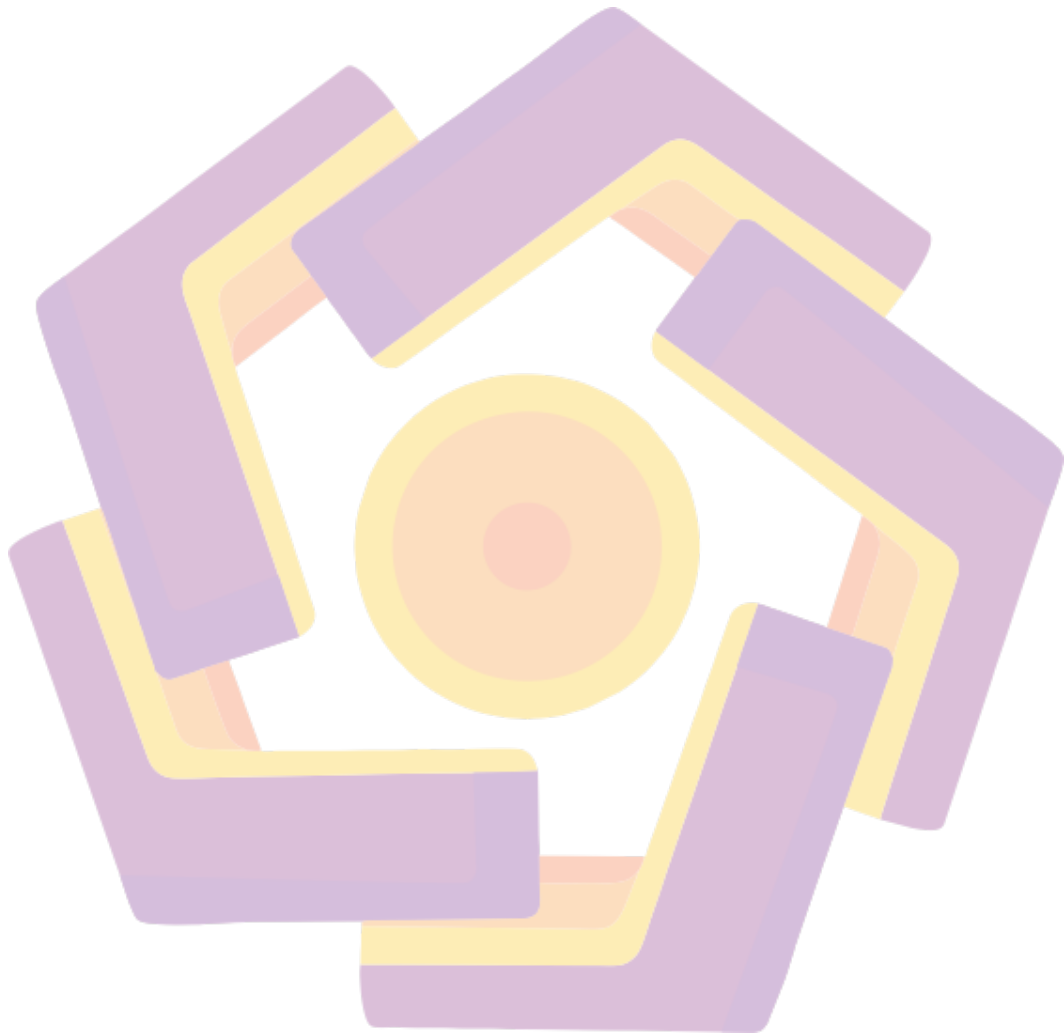
DAFTAR LAMBANG DAN SINGKATAN

t	Periode waktu pengamatan (time-series)
P_t	Persentase sentiment positif pada periode t
$N_{pos,t}$	Jumlah dokumen bersentimen positif pada periode t
$N_{total,t}$	Total volume ulasan yang masuk pada periode t
$W_{d,t}$	Bobot kata t dalam dokumen d
$TF_{d,t}$	Frekuensi kemunculan kata t dalam dokumen d
IDF_t	Nilai Inverse Document Frequency untuk kata t
df_t	Jumlah dokumen yang mengandung kata t
N	Total jumlah seluruh dokumen dalam dataset
d	Derajat ketidakseimbangan data
m_s	Jumlah sampel pada kelas minoritas (Non-Positif)
m_l	Jumlah sampel pada kelas mayoritas (Positif)
r_i	Rasio kesulitan untuk sampel minoritas ke- i
Δ_i	Jumlah tetangga terdekat yang berasal dari kelas mayoritas
K	Jumlah total tetangga terdekat yang digunakan
$\hat{r}_i$	Bobot densitas yang telah dinormalisasi untuk sampel ke- i
g_i	Jumlah data sintetis yang akan dibuat untuk sampel ke- i
G	Total data sintetis yang dibutuhkan untuk menyeimbangkan dataset
s_i	Sampel data sintetis baru yang dihasilkan
x_i	Data minoritas asli yang sedang diproses
x_{zi}	Tetangga minoritas dari x_i yang dipilih secara acak
λ	Vektor acak dengan nilai antara 0 dan 1
x	Vektor input; representasi fitur data ulasan
w	Vektor bobot; penentu orientasi hyperplane
b	Bias atau intercept; konstanta penggeser batas keputusan
$w^T x$	Produk titik antara vektor bobot dan input
$v_i(t + 1)$	Kecepatan partikel
$x_i(t + 1)$	Posisi partikel



w	Bobot inersia untuk menjaga momentum pergerakan
c_1c_2	Konstanta percepatan yang mengatur pengaruh pengalaman partikel
r_1r_2	Bilangan acak antara 0 dan 1 untuk memberikan variasi stokastik
P_{best}	Personal Best; posisi terbaik yang dicapai oleh partikel itu sendiri.
G_{best}	Global Best; posisi terbaik yang dicapai oleh seluruh populasi partikel
r	Koefisien korelasi yang mengukur kekuatan hubungan linier
n	Jumlah total sampel atau observasi data yang dianalisis
X	Variabel independent (Persentase Sentimen Positif)
Y	Variabel dependen (Rata-rata Pendapatan Bulanan)
ABSA	Aspect-Based Sentiment Analysis
ADASYN	Adaptive Synthetic Sampling
AI	Artificial Intelligence
API	Application Programming Interface
BUMDes	Badan Usaha Milik Desa
CFP	Corporate Financial Performance
CSR	Corporate Social Responsibility
CSV	Comma Separated Values
e-WOM	Electronic Word of Mouth
FN	False Negative
FP	False Positive
HTML	HyperText Markup Language
IDF	Inverse Document Frequency
IP	Internet Protocol
JSON	JavaScript Object Notation
KNN	K-Nearest Neighbors
NLP	Natural Language Processing
PSO	Particle Swarm Optimization
SMOTE	Synthetic Minority Over-sampling Technique
SMS	Sentimen Media Sosial
SVM	Support Vector Machine

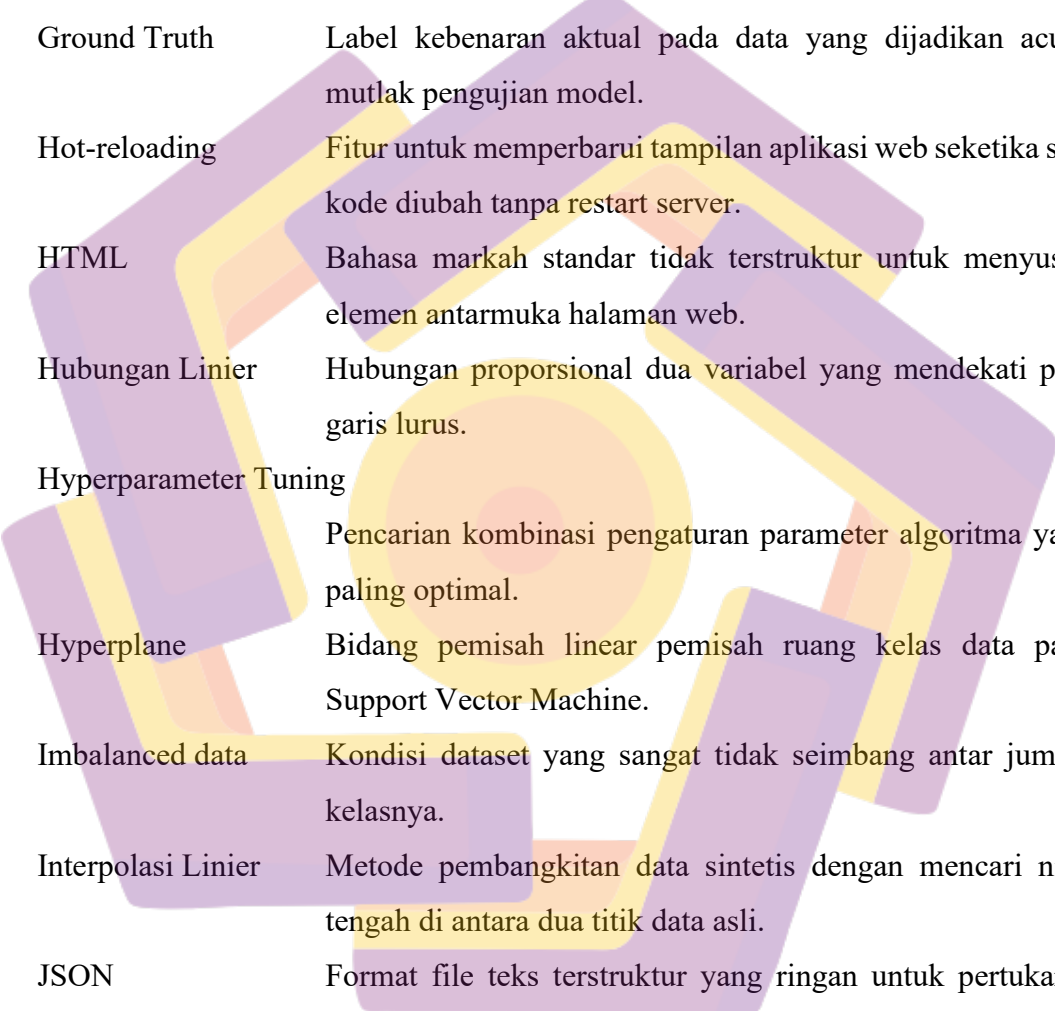
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator



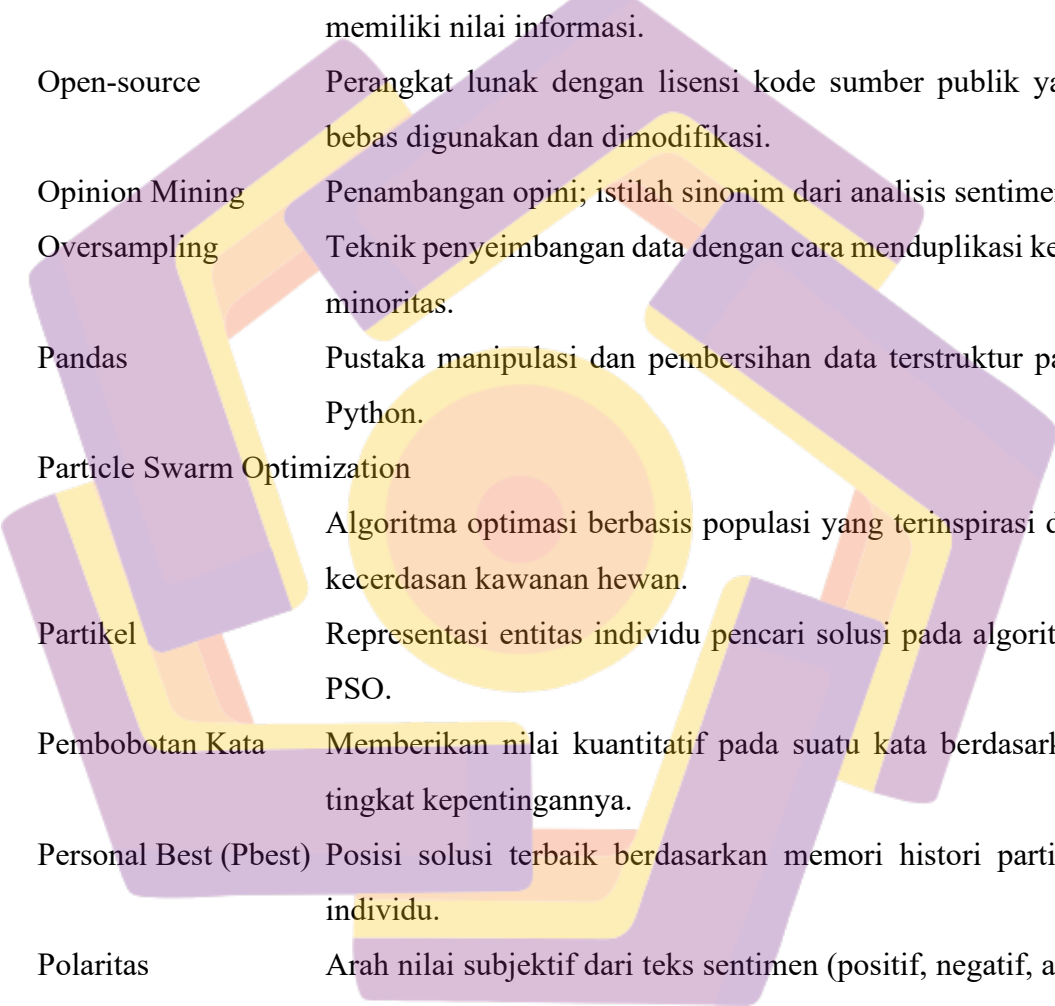
DAFTAR ISTILAH

Accuracy	metrik evaluasi yang menghitung rasio total prediksi yang benar terhadap keseluruhan jumlah data.
Actors	Modul atau skrip serverless pada platform Apify yang beroperasi di komputasi awan untuk memproses tugas ekstraksi data web.
Apify	Platform komputasi awan yang dirancang khusus untuk melakukan otomatisasi penjelajahan web dan data scraping skala besar secara efisien.
Artificial Intelligence	Ilmu yang membekali program komputer dengan kemampuan mensimulasikan kecerdasan manusia.
Aspect-Based Sentiment Analysis	Teknik analisis yang mengklasifikasikan opini berdasarkan kategori atau fitur spesifik dari suatu entitas.
Aspect-Level	Tingkatan analisis sentimen yang mengekstraksi opini berdasarkan aspek spesifik.
Bias / Intercept	Konstanta yang menggeser batas keputusan (hyperplane) agar tidak selalu melewati titik pusat.
Bias Mayoritas	Kegagalan model machine learning yang selalu memprediksi kelas mayoritas akibat ketimpangan data.
Big Data	Kumpulan data dalam volume yang sangat besar, kompleks, dan bertumbuh dengan cepat.
Case Folding	Tahapan preprocessing untuk mengubah seluruh karakter huruf menjadi huruf kecil (lowercase).
Cleaning	Tahapan pembersihan data teks dari karakter yang tidak relevan (seperti simbol, emotikon, URL).
Community Based Tourism	Model pengelolaan pariwisata yang digerakkan dan dikelola oleh komunitas masyarakat setempat.

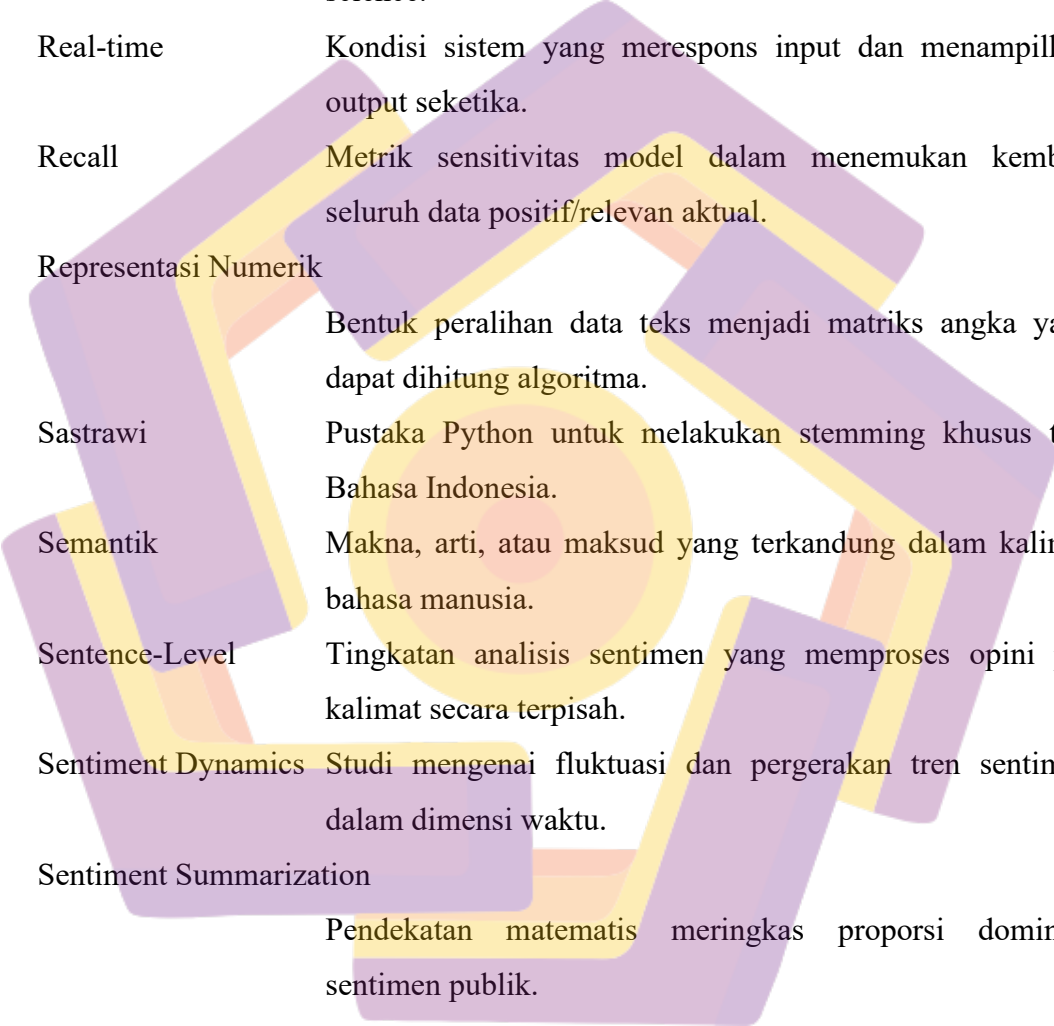
Confusion Matrix	Tabel matriks yang membandingkan hasil prediksi model klasifikasi terhadap label kebenaran data aktual.
Corporate Financial Performance	Kinerja keuangan perusahaan yang mengukur kemampuan entitas dalam menghasilkan pendapatan.
Dataset	Himpunan atau kumpulan data terstruktur yang dikumpulkan dari berbagai sumber untuk diproses lebih lanjut menggunakan algoritma analitik.
Data Mining	Proses penambangan data terstruktur (angka) untuk menemukan pola dan korelasi.
Data-driven insights	Wawasan atau keputusan strategis yang dirumuskan berdasarkan hasil analisis data yang valid.
Decision Boundary	Perbatasan keputusan; area transisi pemisah antara dua kelas data yang berbeda.
Degree of Imbalance	Derajat ketidakseimbangan; rasio tingkat ketimpangan jumlah sampel antar kelas.
Deployment	Tahap penerapan model komputasi menjadi sebuah sistem aplikasi (web) yang siap digunakan.
Document-Level	Tingkatan analisis sentimen yang mengklasifikasikan opini keseluruhan teks sebagai satu kesatuan.
Ekowisata	Perjalanan wisata yang bertanggung jawab terhadap kelestarian alam dan ekonomi masyarakat setempat.
Ekstraksi Fitur	Proses mentransformasikan teks mentah menjadi representasi numerik.
F1-Score	Metrik evaluasi rata-rata harmonik antara Precision dan Recall untuk data tidak seimbang.
False Negative (FN)	Prediksi salah di mana data aktual positif ditebak sebagai negatif.
False Positive (FP)	Prediksi salah di mana data aktual negatif ditebak sebagai positif.



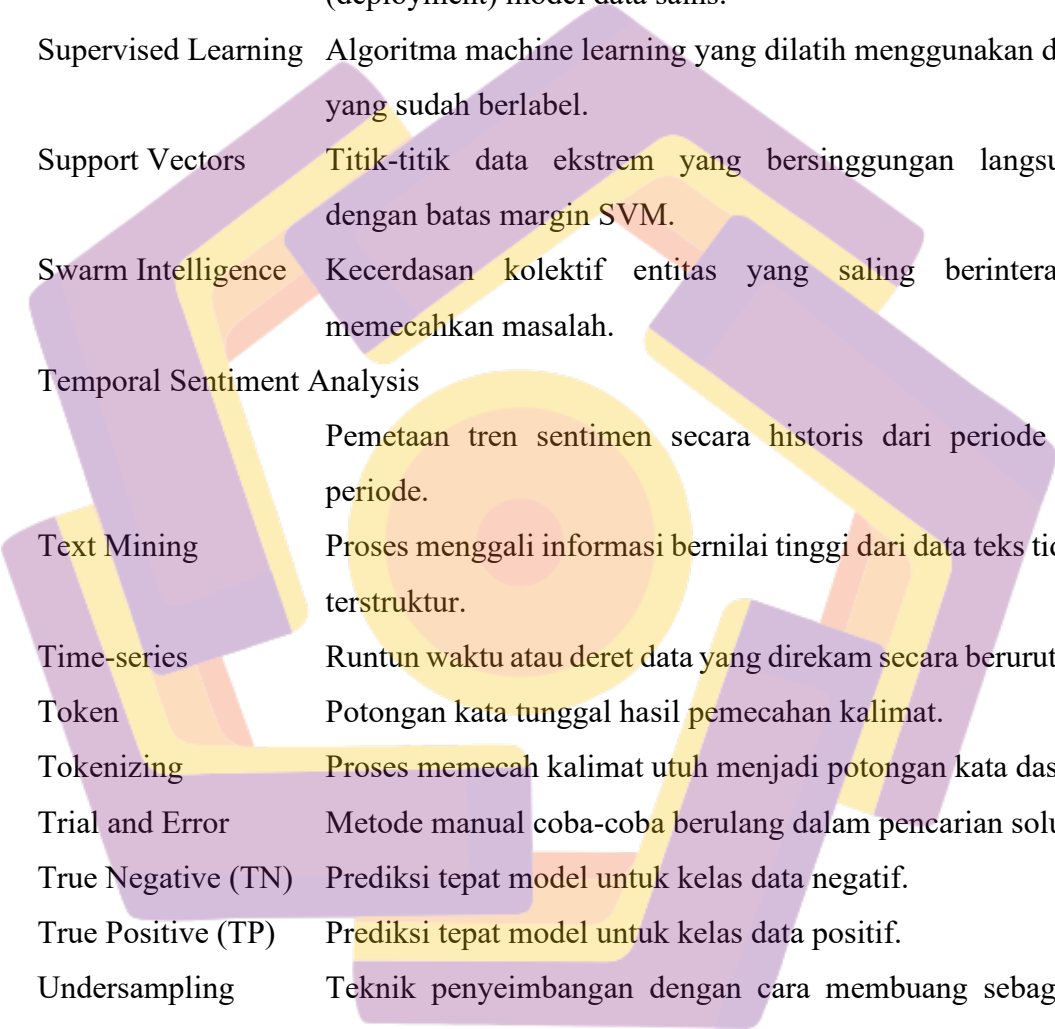
Framework	Kerangka kerja pemrograman yang berisi kumpulan pustaka untuk memfasilitasi pengembangan aplikasi.
Front-end	Bagian antarmuka visual aplikasi yang berhadapan langsung dengan pengguna.
Global Best (Gbest)	Posisi solusi paling optimal yang ditemukan oleh seluruh kawanan partikel PSO.
Ground Truth	Label kebenaran aktual pada data yang dijadikan acuan mutlak pengujian model.
Hot-reloading	Fitur untuk memperbarui tampilan aplikasi web seketika saat kode diubah tanpa restart server.
HTML	Bahasa markah standar tidak terstruktur untuk menyusun elemen antarmuka halaman web.
Hubungan Linier	Hubungan proporsional dua variabel yang mendekati pola garis lurus.
Hyperparameter Tuning	Pencarian kombinasi pengaturan parameter algoritma yang paling optimal.
Hyperplane	Bidang pemisah linear pemisah ruang kelas data pada Support Vector Machine.
Imbalanced data	Kondisi dataset yang sangat tidak seimbang antar jumlah kelasnya.
Interpolasi Linier	Metode pembangkitan data sintetis dengan mencari nilai tengah di antara dua titik data asli.
JSON	Format file teks terstruktur yang ringan untuk pertukaran data.
K-Nearest Neighbors	Metode untuk mencari sampel tetangga terdekat dari suatu titik data.
Koefisien Korelasi (r)	Nilai matematis yang mengukur tingkat kekuatan hubungan linier dua variabel.
Korelasi Pearson Product Moment	



	Teknik statistik parametrik untuk mengukur hubungan linier dua variabel kontinu.
Margin	Jarak maksimum antara hyperplane dengan data terdekat dari masing-masing kelas.
Matplotlib / Plotly	Pustaka visualisasi data grafis interaktif pada Python.
Noise	Karakter pengganggu dalam teks mentah yang tidak memiliki nilai informasi.
Open-source	Perangkat lunak dengan lisensi kode sumber publik yang bebas digunakan dan dimodifikasi.
Opinion Mining	Penambangan opini; istilah sinonim dari analisis sentimen.
Oversampling	Teknik penyeimbangan data dengan cara menduplikasi kelas minoritas.
Pandas	Pustaka manipulasi dan pembersihan data terstruktur pada Python.
Particle Swarm Optimization	Algoritma optimasi berbasis populasi yang terinspirasi dari kecerdasan kawanan hewan.
Partikel	Representasi entitas individu pencari solusi pada algoritma PSO.
Pembobotan Kata	Memberikan nilai kuantitatif pada suatu kata berdasarkan tingkat kepentingannya.
Personal Best (Pbest)	Posisi solusi terbaik berdasarkan memori histori partikel individu.
Polaritas	Arah nilai subjektif dari teks sentimen (positif, negatif, atau netral).
Precision	Metrik ketepatan yang mengukur persentase kebenaran prediksi positif model.
Preprocessing	Tahapan pembersihan dan penyiapan data teks mentah sebelum pemodelan.
Prototyping	Proses pembuatan purwarupa sistem perangkat lunak secara cepat.



Proxy	Server perantara untuk menghindari pemblokiran alamat IP saat web scraping.
Python	Bahasa pemrograman tingkat tinggi yang serbaguna (general-purpose) dan sangat populer digunakan dalam ekosistem pengembangan machine learning dan data science.
Real-time	Kondisi sistem yang merespons input dan menampilkan output seketika.
Recall	Metrik sensitivitas model dalam menemukan kembali seluruh data positif/relevan aktual.
Representasi Numerik	Bentuk peralihan data teks menjadi matriks angka yang dapat dihitung algoritma.
Sastrawi	Pustaka Python untuk melakukan stemming khusus tata Bahasa Indonesia.
Semantik	Makna, arti, atau maksud yang terkandung dalam kalimat bahasa manusia.
Sentence-Level	Tingkatan analisis sentimen yang memproses opini per kalimat secara terpisah.
Sentiment Dynamics	Studi mengenai fluktuasi dan pergerakan tren sentimen dalam dimensi waktu.
Sentiment Summarization	Pendekatan matematis meringkas proporsi dominasi sentimen publik.
Serverless	Model komputasi awan yang membebaskan pengembang dari pengelolaan infrastruktur fisik.
Statistik Parametrik	Cabang statistik untuk menguji data berskala kontinu dan berdistribusi tertentu.
Stemming	Tahapan pemotongan kata berimbuhan menjadi bentuk kata dasarnya.

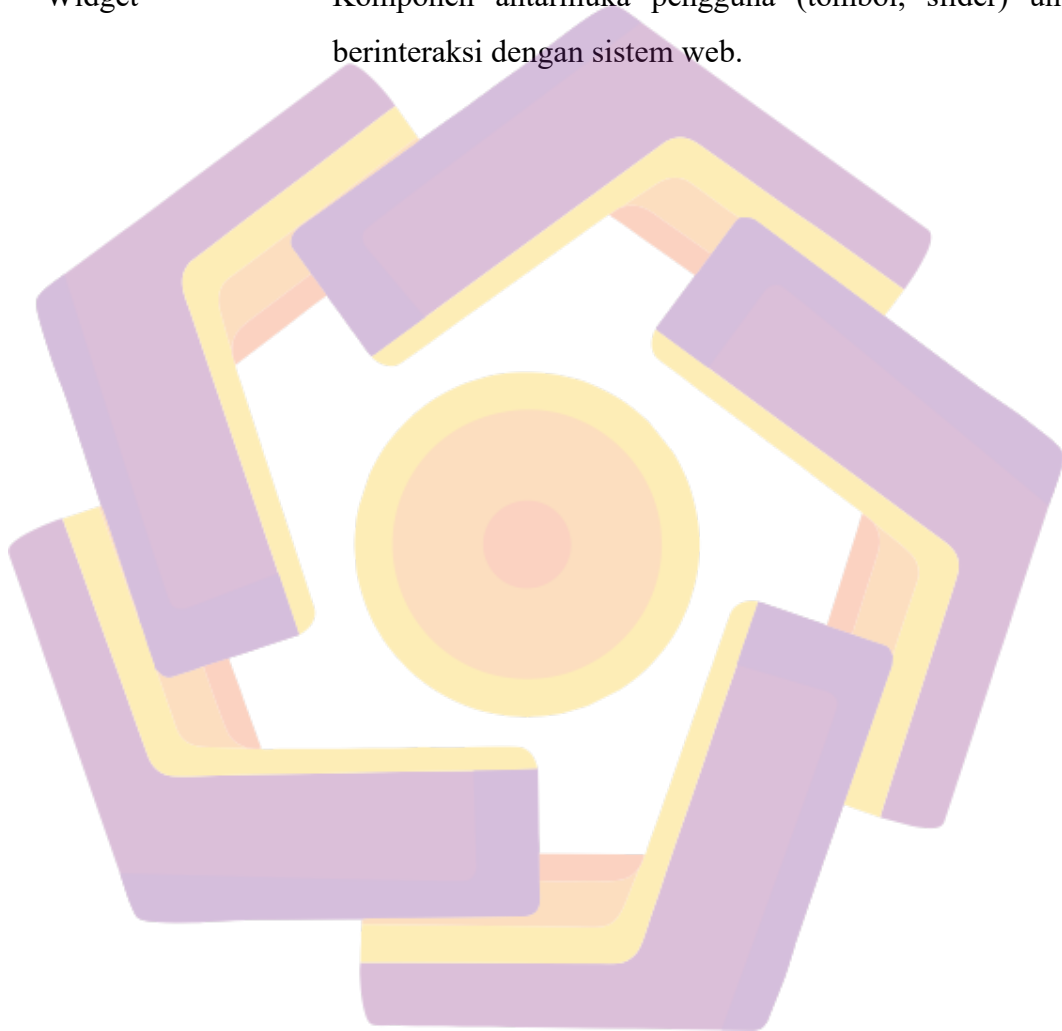


Stokastik	Sesuatu yang bersifat acak (random) atau memiliki elemen probabilitas yang tidak dapat diprediksi secara pasti.
Stopword Removal	Tahapan penghapusan kata-kata umum yang tidak bermakna sentimen kuat.
Streamlit	Framework web interaktif Python untuk implementasi (deployment) model data sains.
Supervised Learning	Algoritma machine learning yang dilatih menggunakan data yang sudah berlabel.
Support Vectors	Titik-titik data ekstrem yang bersinggungan langsung dengan batas margin SVM.
Swarm Intelligence	Kecerdasan kolektif entitas yang saling berinteraksi memecahkan masalah.
Temporal Sentiment Analysis	Pemetaan tren sentimen secara historis dari periode ke periode.
Text Mining	Proses menggali informasi bernilai tinggi dari data teks tidak terstruktur.
Time-series	Runtun waktu atau deret data yang direkam secara berurutan.
Token	Potongan kata tunggal hasil pemecahan kalimat.
Tokenizing	Proses memecah kalimat utuh menjadi potongan kata dasar.
Trial and Error	Metode manual coba-coba berulang dalam pencarian solusi.
True Negative (TN)	Prediksi tepat model untuk kelas data negatif.
True Positive (TP)	Prediksi tepat model untuk kelas data positif.
Undersampling	Teknik penyeimbangan dengan cara membuang sebagian data kelas mayoritas.
User-generated content	Konten ulasan daring yang diciptakan dan dipublikasikan mandiri oleh pengguna akhir.
Variabel Dependen	Variabel terikat yang nilainya dipengaruhi oleh variabel lain.
Variabel Independen	Variabel bebas yang memengaruhi atau menyebabkan perubahan variabel lain.

Web Scraping Teknik ekstraksi otomatis untuk mengunduh data web tidak terstruktur.

Weighted Distribution Distribusi probabilitas densitas untuk menentukan porsi data sintetis.

Widget Komponen antarmuka pengguna (tombol, slider) untuk berinteraksi dengan sistem web.



INTISARI

Pergeseran tren pariwisata digital menjadikan ulasan daring (*Electronic Word of Mouth*) di Google Maps sebagai aset vital bagi Ekowisata Sungai Mudal. Sentimen ulasan pengunjung berdampak langsung pada citra destinasi dan stabilitas pendapatan pasca-pandemi. Namun, pengelola kesulitan memetakan ribuan data teks tidak terstruktur yang memiliki karakteristik sangat timpang (*imbalanced data*), di mana ulasan positif mendominasi. Kondisi ini memicu bias mayoritas pada algoritma standar, sehingga model gagal mendeteksi ulasan keluhan (minoritas) yang sebenarnya sangat krusial untuk evaluasi layanan.

Penelitian ini mengusulkan model klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM) berbasis fitur TF-IDF. Untuk mengatasi bias mayoritas, teknik *Adaptive Synthetic Sampling* (ADASYN) diterapkan pada data latih guna membangkitkan data minoritas secara adaptif, dilanjutkan optimasi *hyperparameter* menggunakan *Particle Swarm Optimization* (PSO). Model diimplementasikan untuk memprediksi 4.037 ulasan periode 2019–2024. Selanjutnya, hasil prediksi tren persentase kepuasan diuji korelasinya terhadap fluktuasi rata-rata pendapatan bulanan menggunakan metode *Pearson Product Moment*.

Hasil eksperimen menunjukkan bahwa integrasi ADASYN dan PSO berhasil mendongkrak sensitivitas model (*Recall*) dalam mendeteksi keluhan menjadi 29% dengan Akurasi Global terjaga di 87%. Uji statistik membuktikan adanya korelasi linier positif yang sangat kuat ($r = 0,92$) antara reputasi digital dan kinerja finansial. Artefak penelitian ini telah diimplementasikan menjadi aplikasi web interaktif berbasis *cloud* (Streamlit) yang dapat dimanfaatkan langsung oleh pengelola ekowisata untuk memantau kepuasan pengunjung secara mandiri.

Kata kunci: Analisis Sentimen, *Imbalanced Data*, SVM, ADASYN-PSO, Korelasi Pearson

ABSTRACT

The shift in digital tourism trends makes online reviews (Electronic Word of Mouth) on Google Maps a vital asset for Ekowisata Sungai Mudal. Visitor sentiment directly impacts the destination's image and post-pandemic revenue stability. However, the management struggles to map thousands of unstructured text data that possess highly imbalanced characteristics, where positive reviews dominate. This condition triggers majority bias in standard algorithms, causing the model to fail in detecting complaint reviews (minority class) which are actually crucial for service evaluation.

This study proposes a sentiment classification model using the Support Vector Machine (SVM) algorithm based on TF-IDF features. To overcome the majority bias, the Adaptive Synthetic Sampling (ADASYN) technique was applied to the training data to generate minority data adaptively, followed by hyperparameter tuning using Particle Swarm Optimization (PSO). The model was implemented to predict 4,037 reviews from the 2019–2024 period. Furthermore, the predicted trend of satisfaction percentage was tested for its correlation with monthly average revenue fluctuations using the Pearson Product Moment method.

The experimental results showed that the integration of ADASYN and PSO successfully boosted the model's sensitivity (Recall) in detecting complaints to 29% while maintaining a Global Accuracy of 87%. Statistical tests proved a very strong positive linear correlation ($r = 0.92$) between digital reputation and financial performance. This research artifact has been deployed into a cloud-based interactive web application (Streamlit) that can be directly utilized by ecotourism management to monitor visitor satisfaction independently.

Keyword: Sentiment Analysis, Imbalanced Data, SVM, ADASYN-PSO, Pearson Correlation