

**PENERAPAN GRID SEARCH UNTUK OPTIMASI MODEL
MACHINE LEARNING DALAM KLASIFIKASI SENTIMEN
KOMENTAR YOUTUBE**

SKRIPSI

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



disusun oleh

MAHENDRA BAYU PRAYOGA

21.11.4117

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PENERAPAN GRID SEARCH UNTUK OPTIMASI MODEL
MACHINE LEARNING DALAM KLASIFIKASI SENTIMEN
KOMENTAR YOUTUBE**

SKRIPSI

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



disusun oleh

MAHENDRA BAYU PRAYOGA

21.11.4117

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN GRID SEARCH UNTUK OPTIMASI MODEL MACHINE
LEARNING DALAM KLASIFIKASI SENTIMEN KOMENTAR
YOUTUBE**

yang disusun dan diajukan oleh

Mahendra Bayu Prayoga

21.11.4117

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Januari 2025

Dosen Pembimbing,



Nuri Cahyono, M.Kom.

NIK. 190302278

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN GRID SEARCH UNTUK OPTIMASI MODEL MACHINE
LEARNING DALAM KLASIFIKASI SENTIMEN KOMENTAR
YOUTUBE

yang disusun dan diajukan oleh

Mahendra Bayu Prayoga

21.11.4117

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Januari 2025

Susunan Dewan Penguji

Nama Penguji

I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302352

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Nuri Cahyono, M.Kom.
NIK. 190302278

Tanda Tangan



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Mahendra Bayu Prayoga
NIM : 21.11.4117

Menyatakan bahwa Skripsi dengan judul berikut:

**Penerapan Grid Search Untuk Optimasi Model Machine Learning Dalam
Klasifikasi Sentimen Komentar YouTube**

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, 21 Januari 2025

Yang Menyatakan,



Mahendra Bayu Prayoga

HALAMAN PERSEMBAHAN

Terwujudnya skripsi ini tidak akan tercapai tanpa dukungan dan bantuan dari berbagai pihak yang senantiasa memberikan dorongan serta motivasi kepada penulis dalam menyelesaikan karya ini. Oleh karena itu, penulis dengan penuh rasa hormat mengucapkan terima kasih kepada:

1. Bapak Nuri Cahyono, M.Kom, selaku dosen pembimbing, yang telah memberikan bimbingan, masukan, dan motivasi dengan penuh kesabaran serta dedikasi hingga skripsi ini dapat diselesaikan dengan baik.
2. Orang tua tercinta, yang selalu memberikan kasih sayang, doa, serta dukungan moral dan materi yang tak pernah putus, menjadi sumber kekuatan utama bagi penulis dalam menyelesaikan skripsi ini. Tak lupa dengan adik tercinta, yang selalu menghibur dan memberikan semangat kepada penulis.
3. Bagas Restya Ermawan, Aldino Marsel Pratama, dan Akmal Fauzan Restu Agung, atas dukungan, bantuan, dan waktu yang diberikan untuk mendampingi penulis dalam menghadapi berbagai kesulitan selama penyusunan skripsi ini.
4. Teman-teman seperjuangan, yang senantiasa memberikan semangat, kebersamaan, dan dorongan selama masa studi hingga penyelesaian karya ini, khususnya kepada: Wayan Surya, Hafid Afnan, Reihansyah Maulana, dan Aji Mustajab.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini diajukan sebagai salah satu syarat untuk menyelesaikan program studi pada tingkat pendidikan tinggi.

Dalam proses penyusunan skripsi ini, penulis mendapatkan banyak dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Nuri Cahyono, M.Kom, selaku Dosen Pembimbing yang telah memberikan arahan, masukan, dan motivasi selama penyusunan skripsi ini. Ucapan terima kasih juga penulis sampaikan kepada Tim Dosen Penguji atas saran dan kritik yang membangun untuk perbaikan karya ini.

Penulis juga mengucapkan terima kasih kepada orang tua dan adik tercinta yang senantiasa memberikan doa, dukungan moral, dan materi selama proses studi hingga penyelesaian skripsi ini. Tidak lupa, ucapan terima kasih disampaikan kepada seluruh teman-teman dan orang-orang terdekat yang telah membantu, baik secara langsung maupun tidak langsung, dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan untuk perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan pembaca sekalian.

Yogyakarta, 21 Januari 2025

Penulis

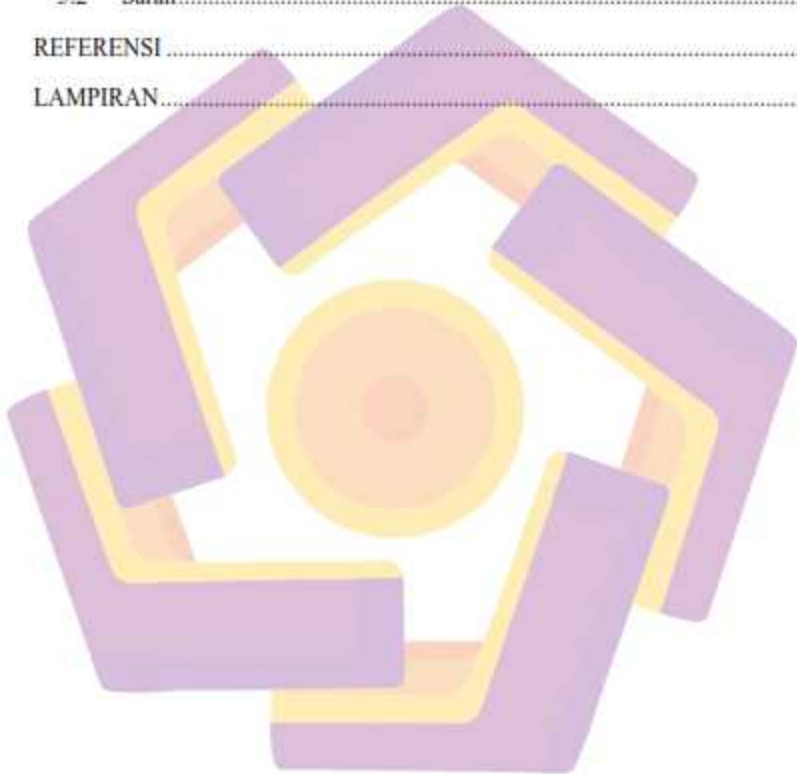
DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xvi
DAFTAR LAMBANG DAN SINGKATAN	xvii
DAFTAR ISTILAH	xviii
INTISARI	xx
<i>ABSTRACT</i>	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5

2.2	Dasar Teori	10
2.2.1	<i>YouTube</i>	10
2.2.2	<i>Google Drive</i>	10
2.2.3	<i>Google Colaboratory</i>	11
2.2.4	<i>Python</i>	11
2.2.5	Klasifikasi Sentimen	12
2.2.6	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	13
2.2.7	<i>Lexicon Based</i>	14
2.2.8	<i>SMOTE</i>	15
2.2.9	<i>Machine Learning</i>	15
2.2.10	<i>Support Vector Machine (SVM)</i>	16
2.2.11	<i>Logistic Regression</i>	18
2.2.12	<i>Random Forest</i>	19
2.2.13	<i>Grid Search</i>	20
2.2.14	<i>Confusion Matrix</i>	21
BAB III METODE PENELITIAN		24
3.1	Alur Penelitian	24
3.1.1	<i>Data Scraping</i>	25
3.1.2	<i>Preprocessing</i>	25
3.1.3	<i>Labelling</i>	29
3.1.4	<i>Feature Extraction with TF-IDF</i>	30
3.1.5	<i>Oversampling with SMOTE</i>	30
3.1.6	<i>Data Splitting</i>	30
3.1.7	<i>Support Vector Machine (SVM)</i>	31
3.1.8	<i>Logistic Regression</i>	31

3.1.9	<i>Random Forest</i>	31
3.1.10	<i>Optimization using Grid Search CV</i>	31
3.1.11	<i>Confusion Matrix</i>	32
3.2	Alat dan Bahan	32
BAB IV HASIL DAN PEMBAHASAN		34
4.1	<i>Data Scraping</i>	34
4.2	<i>Preprocessing</i>	35
4.2.1	<i>Unique Handling</i>	35
4.2.2	<i>Case Folding</i>	36
4.2.3	<i>Text Cleaning</i>	37
4.2.4	<i>Slang Words Normalization</i>	38
4.2.5	<i>Stopwords Removal</i>	39
4.2.6	<i>Tokenizing</i>	40
4.2.7	<i>Stemming</i>	41
4.3	<i>Labelling</i>	42
4.4	<i>Feature Extraction with TF-IDF</i>	44
4.5	<i>Oversampling with SMOTE</i>	46
4.6	<i>Data Splitting</i>	48
4.7	<i>Hasil Modelling</i>	48
4.7.1	Hasil Klasifikasi Algoritma <i>SVM</i>	49
4.7.2	Hasil Klasifikasi Algoritma <i>SVM</i> dengan <i>Grid Search</i>	52
4.7.3	Hasil Klasifikasi Algoritma <i>Logistic Regression</i>	54
4.7.4	Hasil Klasifikasi Algoritma <i>Logistic Regression</i> dengan <i>Grid Search</i> 57	
4.7.5	Hasil Klasifikasi Algoritma <i>Random Forest</i>	59

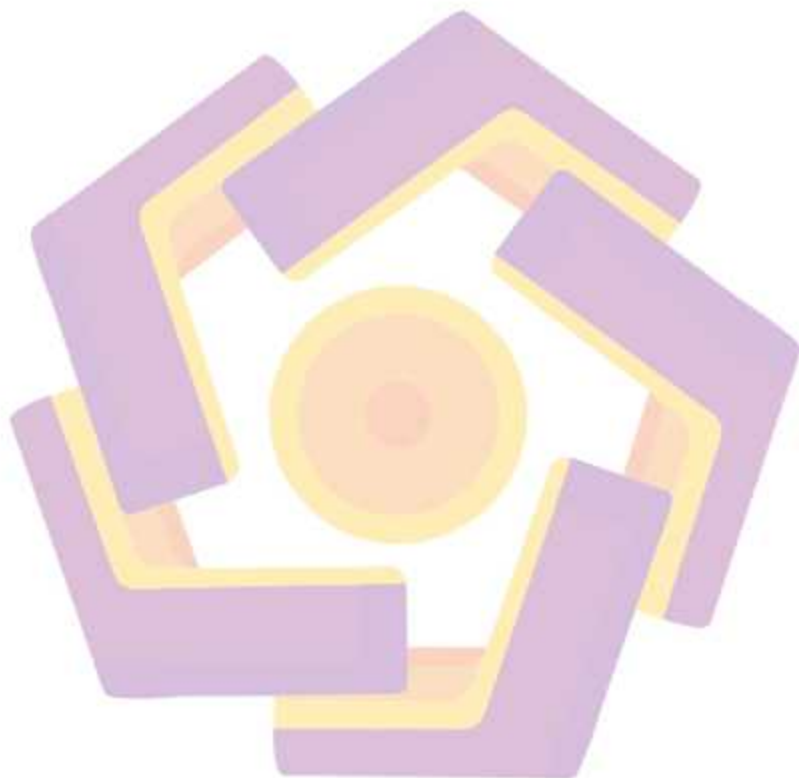
4.7.6	Hasil Klasifikasi Algoritma <i>Random Forest</i> dengan <i>Grid Search</i> ..	62
4.7.7	Hasil Perbandingan Metode Klasifikasi.....	64
BAB V PENUTUP		66
5.1	Kesimpulan.....	66
5.2	Saran.....	66
REFERENSI		67
LAMPIRAN.....		72



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 3.1 Contoh Tahap <i>Case Folding</i>	25
Tabel 3.2 Contoh Tahap <i>Text Cleaning</i>	26
Tabel 3.3 Contoh Tahap <i>Slang Word Normalization</i>	27
Tabel 3.4 Contoh Tahap <i>Stopwords Removal</i>	28
Tabel 3.5 Contoh Tahap <i>Tokenizing</i>	28
Tabel 3.6 Contoh Tahap <i>Stemming</i>	29
Tabel 4.1 Tabel Sebelum Pembersihan Data	36
Tabel 4.2 Tabel Setelah Pembersihan Data	36
Tabel 4.3 Hasil <i>Case Folding</i>	37
Tabel 4.4 Hasil <i>Text Cleaning</i>	38
Tabel 4.5 Hasil <i>Slang Words Normalization</i>	39
Tabel 4.6 Hasil <i>Stopwords Removal</i>	40
Tabel 4.7 Hasil <i>Tokenizing</i>	40
Tabel 4.8 Hasil <i>Stemming</i>	41
Tabel 4.9 Hasil <i>Labelling</i>	43
Tabel 4.10 Dataset Komentar <i>YouTube</i>	45
Tabel 4.11 Perbandingan nilai TF-IDF untuk kata kunci	45
Tabel 4.12 Top 10 kata dengan bobot TF-IDF tertinggi pada komentar ke-1	46
Tabel 4.13 Perbandingan Metrik Evaluasi Tanpa <i>SMOTE</i> dan Dengan <i>SMOTE</i> ..	49
Tabel 4.14 Hasil <i>Grid Search SVM</i>	52
Tabel 4.15 Perbandingan Metrik Evaluasi Tanpa <i>SMOTE</i> dan Dengan <i>SMOTE</i> ..	54
Tabel 4.16 Hasil <i>Grid Search Logistic Regression</i>	57

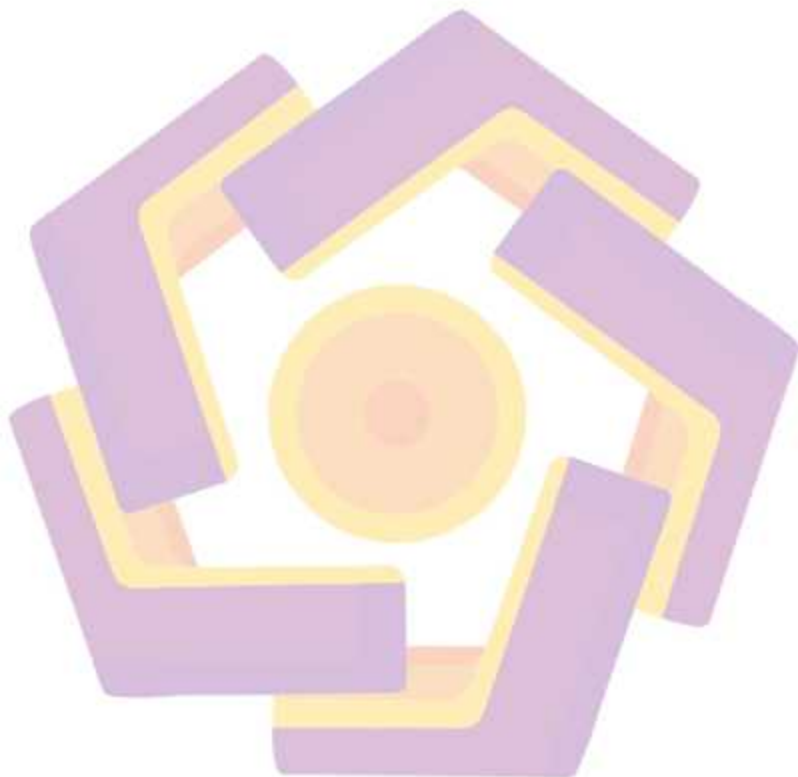
Tabel 4.17 Perbandingan Metrik Evaluasi Tanpa <i>SMOTE</i> dan Dengan <i>SMOTE</i> ..	59
Tabel 4.18 Hasil <i>Grid Search Random Forest</i>	62
Tabel 4.19 Perbandingan Akurasi Sebelum dan Sesudah Optimasi	65



DAFTAR GAMBAR

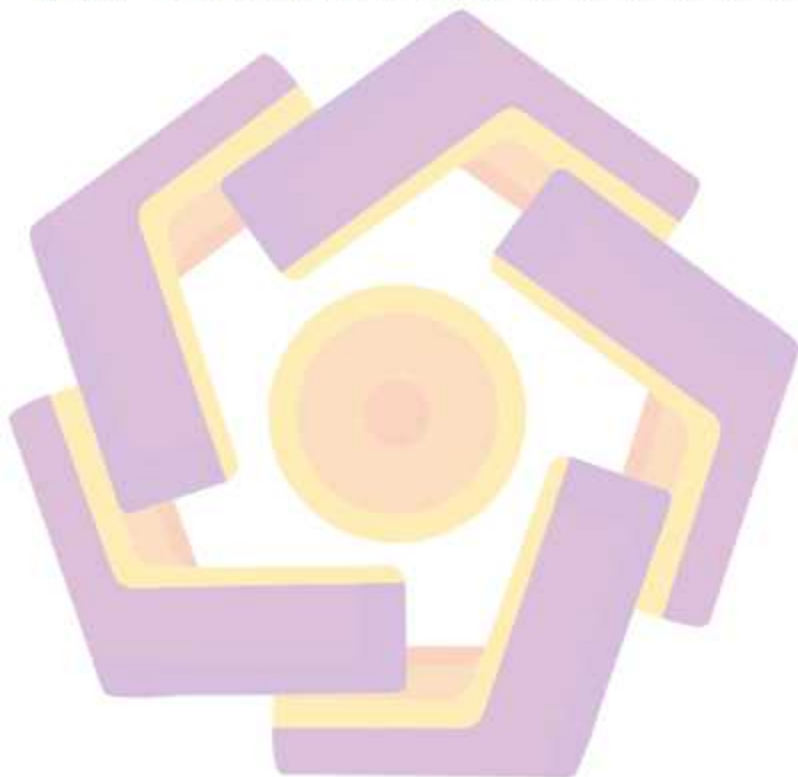
Gambar 2.1 Logo <i>YouTube</i>	10
Gambar 2.2 Logo <i>Google Drive</i>	11
Gambar 2.3 Logo <i>Google Colaboratory</i>	11
Gambar 2.4 Logo <i>Python</i>	12
Gambar 2.5 Representasi <i>Hyperplane</i> pada Algoritma <i>SVM</i>	16
Gambar 2.6 Grafik <i>MedparSlos</i> terhadap Probabilitas (μ)	19
Gambar 2.7 Tingkat <i>Error Random Forest</i> Berdasarkan Jumlah Pohon.....	20
Gambar 2.8 Ilustrasi <i>Grid Search</i>	21
Gambar 2.9 <i>Confusion Matrix 3x3</i>	22
Gambar 3.1 Alur Penelitian	24
Gambar 4.1 Dataset Komentar <i>YouTube</i>	34
Gambar 4.2 Distribusi Data	43
Gambar 4.3 Distribusi Data Setelah Penghapusan Data Duplikat.....	44
Gambar 4.4 Distribusi Dataset Sebelum <i>SMOTE</i>	47
Gambar 4.5 Distribusi Dataset Setelah <i>SMOTE</i>	47
Gambar 4.6 <i>Confusion Matrix</i> Model <i>SVM</i> tanpa <i>SMOTE</i>	51
Gambar 4.7 <i>Confusion Matrix</i> Model <i>SVM</i> dengan <i>SMOTE</i>	51
Gambar 4.8 Hasil Klasifikasi Model <i>SVM</i> dengan <i>Grid Search</i>	53
Gambar 4.9 <i>Confusion Matrix</i> Model <i>SVM</i> dengan <i>Grid Search</i>	54
Gambar 4.10 <i>Confusion Matrix</i> Model <i>Logistic Regression</i> dengan <i>SMOTE</i>	56
Gambar 4.11 <i>Confusion Matrix</i> Model <i>Logistic Regression</i> dengan <i>SMOTE</i>	56
Gambar 4.12 Hasil Klasifikasi Model <i>Logistic Regression</i> dengan <i>Grid Search</i> ..	58
Gambar 4.13 <i>Confusion Matrix</i> Model <i>Logistic Regression</i> dengan <i>Grid Search</i>	59

Gambar 4.14 <i>Confusion Matrix</i> Model <i>Random Forest</i> tanpa <i>SMOTE</i>	61
Gambar 4.15 <i>Confusion Matrix</i> Model <i>Random Forest</i> dengan <i>SMOTE</i>	61
Gambar 4.16 Hasil Klasifikasi Model <i>Random Forest</i> dengan <i>Grid Search</i>	63
Gambar 4.17 <i>Confusion Matrix</i> Model <i>Random Forest</i> dengan <i>Grid Search</i>	64



DAFTAR LAMPIRAN

Lampiran 1 Link YouTube yang Di- <i>Scraping</i>	72
Lampiran 2 Kode <i>Scraping Data</i>	73
Lampiran 3 Kode Klasifikasi Sentimen	75



DAFTAR LAMBANG DAN SINGKATAN

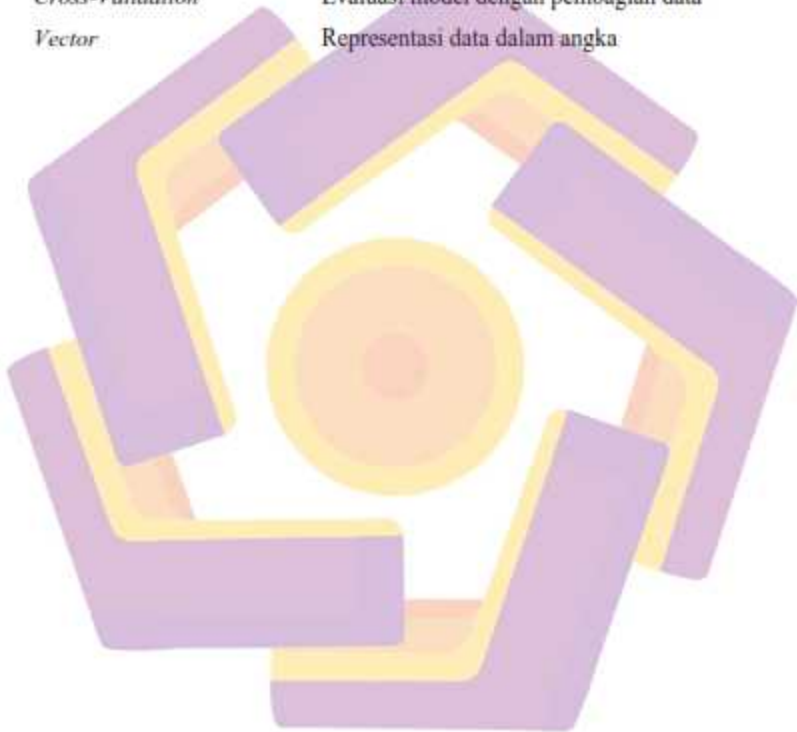


μ	<i>Rata-rata distribusi data</i>
δ	<i>Perubahan kecil pada parameter</i>
γ	<i>Parameter kernel pada SVM</i>
σ	<i>Standar deviasi atau parameter Gaussian</i>
β	<i>Parameter untuk F-Beta Score</i>
Π	<i>Operator perkalian (product)</i>
Σ	<i>Operator penjumlahan (summation)</i>
SVM	<i>Support Vector Machines</i>
LR	<i>Logistic Regression</i>
RF	<i>Random Forest</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
VADER	<i>Valance Aware Dictionary for Sentiment Reasoning</i>
CV	<i>Cross Validation</i>
URL	<i>Uniform Resource Locators</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
NLTK	<i>Natural Language Toolkit, pustaka NLP</i>
ML	<i>Mahine Learning</i>
InSet	<i>Indonesian Sentiment Lexicon</i>
TP	<i>True Positive</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
TN	<i>True Negative</i>

DAFTAR ISTILAH

<i>Hyperparameter</i>	Parameter model sebelum pelatihan.
<i>Tuning</i>	Optimasi hyperparameter
<i>Grid Search</i>	Metode mencari kombinasi hyperparameter terbaik
Klasifikasi	Mengelompokkan data ke kelas tertentu
Parameter	Nilai dalam model yang diatur saat pelatihan
Sentimen	Analisis emosi dalam teks
<i>Supervised Learning</i>	Pembelajaran dengan data berlabel
<i>Hyperplane</i>	Garis pemisah dalam SVM
<i>Support Vectors</i>	Titik data penentu margin di SVM
<i>Kernel</i>	Fungsi pemetaan data di SVM
<i>Linear</i>	Hubungan langsung antar variabel
<i>Library</i>	Kumpulan kode siap pakai
<i>Confusion matrix</i>	Tabel evaluasi prediksi model
<i>Preprocessing</i>	Pemrosesan awal data
<i>Slang</i>	Bahasa informal
<i>Machine learning</i>	Mesin belajar dari data
<i>Sastrawi</i>	Library stemming Bahasa Indonesia
<i>Oversampling</i>	Menambah data kelas minoritas
<i>Platform</i>	Lingkungan untuk menjalankan aplikasi
<i>Corpus</i>	Kumpulan teks untuk analisis
<i>Scikit-learn</i>	Library machine learning Python
<i>Matplotlib</i>	Library visualisasi data
<i>Lexicon</i>	Kumpulan kata berlabel sentimen
<i>Overfitting</i>	Model terlalu sesuai data pelatihan
<i>Accuracy</i>	Persentase prediksi benar
<i>Precision</i>	Akurasi prediksi positif
<i>Recall</i>	Kemampuan mendeteksi data positif
<i>F1-score</i>	Rata-rata precision dan recall
<i>Scraping</i>	Mengambil data dari web

<i>Null</i>	Nilai kosong
<i>JSON</i>	Format data berbasis teks
<i>Noise</i>	Data tidak relevan
<i>Class Imbalance</i>	Ketidakseimbangan data antar kelas
<i>Training</i>	Proses melatih model
<i>Testing</i>	Uji model dengan data baru
<i>Cross-Validation</i>	Evaluasi model dengan pembagian data
<i>Vector</i>	Representasi data dalam angka



INTISARI

YouTube bukan hanya platform berbagi video, tetapi juga ruang interaksi sosial melalui komentar yang mencerminkan opini dan persepsi pengguna. Penelitian ini bertujuan untuk menganalisis kinerja algoritma klasifikasi sentimen, yaitu *Support Vector Machine (SVM)*, *Logistic Regression*, dan *Random Forest*, yang dioptimasi menggunakan *Grid Search* untuk klasifikasi sentimen komentar *YouTube*. Dataset yang digunakan terdiri dari 12.000 komentar dari 12 topik berbeda, meliputi politik, pendidikan, hiburan, teknologi, dan lainnya. Data diproses melalui *preprocessing* yang meliputi *case folding*, *text cleaning*, *slang words normalization*, *stopword removal*, *tokenizing*, dan *stemming* menggunakan *Sastrawi*. Pelabelan sentimen dilakukan dengan menggunakan kamus sentimen dalam tiga kelas: positif, negatif, dan netral. *Oversampling* dengan *SMOTE* diterapkan untuk mengatasi ketidakseimbangan data. Hasil penelitian menunjukkan bahwa optimasi menggunakan *Grid Search* meningkatkan akurasi model: *SVM* meningkat dari 90,97% menjadi 92,55%, *Logistic Regression* dari 86,72% menjadi 92,34%, dan *Random Forest* dari 87,33% menjadi 88,84%. Model *SVM* dengan kombinasi parameter '*C*': 10, '*max_iter*': 1000, dan '*tol*': 0.0001 menunjukkan performa terbaik dibandingkan model lain yang sudah dioptimasi. Ekstraksi fitur menggunakan *TF-IDF* dan tahapan *preprocessing* memberikan kontribusi signifikan terhadap peningkatan kinerja model. Penelitian ini berkontribusi pada pengembangan metode optimasi model untuk klasifikasi sentimen teks, khususnya komentar *YouTube*. Sebagai saran, penelitian lanjutan dapat mengeksplorasi algoritma *machine learning* lainnya, memperkaya kamus pelabelan, serta menggunakan teknik klasifikasi canggih seperti *deep learning* atau *transformer* pada dataset yang lebih besar dan dari berbagai platform media sosial.

Kata kunci: *Grid Search*, Klasifikasi Sentimen, *Machine Learning*, Optimasi Model, *YouTube*

ABSTRACT

YouTube is not only a video-sharing platform but also a space for social interaction through comments that reflect users' opinions and perceptions. This study aims to analyze the performance of sentiment classification algorithms, namely Support Vector Machine (SVM), Logistic Regression, and Random Forest, optimized using Grid Search for classifying YouTube comment sentiments. The dataset consists of 12,000 comments from 12 different topics, including politics, education, entertainment, technology, and others. The data is processed through preprocessing steps such as case folding, text cleaning, slang words normalization, stopword removal, tokenizing, and stemming using Sastrawi. Sentiment labeling is conducted using a sentiment dictionary into three classes: positive, negative, and neutral. Oversampling with SMOTE is applied to address data imbalance. The study results show that optimization using Grid Search improves model accuracy: SVM increased from 90.97% to 92.55%, Logistic Regression from 86.72% to 92.34%, and Random Forest from 87.33% to 88.84%. The SVM model with parameter combinations of 'C': 10, 'max_iter': 1000, and 'tol': 0.0001 demonstrated the best performance compared to other optimized models. Feature extraction using TF-IDF and the preprocessing steps significantly contributed to improving model performance. This study contributes to the development of model optimization methods for text sentiment classification, particularly for YouTube comments. Future research is recommended to explore other machine learning algorithms, enrich sentiment labeling dictionaries, and use advanced classification techniques such as deep learning or transformers on larger datasets from various social media platforms.

Keyword: Grid Search, Sentiment Classification, Machine Learning, Model Optimization, YouTube