

**OPTIMASI HYPERPARAMETER SUPPORT VECTOR
MACHINE (SVM) MENGGUNAKAN GRID SEARCH DAN
CROSS-VALIDATION UNTUK MENINGKATKAN AKURASI
KLASIFIKASI SENTIMEN ULASAN APLIKASI AMIKOM
ONE**

SKRIPSI

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



disusun oleh

SANDI HERNANTO

21.11.4530

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**OPTIMASI HYPERPARAMETER SUPPORT VECTOR
MACHINE (SVM) MENGGUNAKAN GRID SEARCH DAN
CROSS-VALIDATION UNTUK MENINGKATKAN AKURASI
KLASIFIKASI SENTIMEN ULASAN APLIKASI AMIKOM
ONE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

SANDI HERNANTO

21.11.4530

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMASI HYPERPARAMETER SUPPORT VECTOR MACHINE (SVM)
MENGUNAKAN GRID SEARCH DAN CROSS-VALIDATION UNTUK
MENINGKATKAN AKURASI KLASIFIKASI SENTIMEN ULASAN
APLIKASI AMIKOM ONE**

yang disusun dan diajukan oleh

Sandi Hernanto

21.11.4530

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 16 Desember 2024

Dosen Pembimbing,



Windha Mega Pradnya Duhita, M.Kom
NIK. 190302185

HALAMAN PENGESAHAN

SKRIPSI

OPTIMASI HYPERPARAMETER SUPPORT VECTOR MACHINE (SVM)
MENGUNAKAN GRID SEARCH DAN CROSS-VALIDATION UNTUK
MENINGKATKAN AKURASI KLASIFIKASI SENTIMEN ULASAN
APLIKASI AMIKOM ONE

yang disusun dan diajukan oleh

Sandi Hernanto

21.11.4530

Telah dipertahankan di depan Dewan Penguji
pada tanggal 16 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Rifda Faticha Alfa Aziza, S.Kom., M.Kom.
NIK. 190302392

Ainul Yaqin, M.Kom
NIK. 190302255

Windha Meza Pradnya Dhubita, M.Kom.
NIK. 190302185



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 Desember 2024

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 : Sandi Hernanto
NIM : 21.11.4530

Menyatakan bahwa Skripsi dengan judul berikut:

Optimasi Hyperparameter Support Vector Machine (SVM) Menggunakan Grid Search dan Cross-Validation Untuk Meningkatkan Akurasi Klasifikasi Sentimen Ulasan Aplikasi Amikom One

Dosen Pembimbing: Windha Mega Pradnya Duhita, 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, 16 Desember 2024

Yang Menyatakan,



Sandi Hernanto

HALAMAN PERSEMBAHAN

Segala puji dan ucapan rasa syukur penulis panjatkan kepada Allah SWT atas berkat rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini. Shalawat serta salam senantiasa tercurahkan kepada Nabi Muhammad SAW. Dengan segenap ketulusan dan kerendahan hati, penulis dapat sampai pada tahap ini yang merupakan sebuah pencapaian yang tentunya tidak diraih dengan mudah. Berbagai tantangan telah dilalui, namun berkat niat yang tulus, doa-doa yang teriring, serta dukungan dari orang-orang tercinta, skripsi ini dapat terselesaikan dengan baik. Skripsi ini penulis persembahkan dengan penuh cinta dan rasa hormat kepada:

1. Ayah Maniso dan Ibu Sumirah Tercinta, terima kasih atas doa yang tak pernah putus, kasih sayang yang tulus, dan dukungan tanpa batas. Ayah dan Ibu adalah sumber kekuatan penulis untuk terus maju dan menyelesaikan setiap tantangan. Segala pencapaian ini tak akan terwujud tanpa kehadiran dan cinta kalian.
2. Nenek Tersayang, Wakinem, Wakijah, dan Wasiyam yang selalu mendoakan dan memberi semangat. Kehadiran dan doa dari kalian memberi penulis kekuatan dalam menyelesaikan perjalanan ini.
3. Kakak Tercinta, Heni Septiningsih, terima kasih atas dukungan, perhatian, dan kasih sayang yang selalu ada di setiap langkah. Kehadiranmu adalah semangat bagi penulis untuk terus maju.
4. Keponakan Tersayang, Anindra Cahya Nugraha, Keenandra Alfarizky Nugroho, serta untuk alm. Alvarendra Adha Nugroho yang telah berpulang lebih dulu. Kalian adalah anugerah dalam keluarga dan penyemangat dalam hati penulis.
5. Nur Aini yang senantiasa mendengarkan keluh kesah penulis, memberi dukungan, motivasi, pengingat, dan menemani penulis sehingga skripsi ini dapat terselesaikan dengan baik.
6. Teman-teman Seperjuangan, terima kasih kepada sahabat-sahabat yang selalu menemani dan berbagi dalam suka maupun duka. Terima kasih atas

kebersamaan, semangat, dan dukungan yang tak ternilai sepanjang perjalanan ini.

7. Almamater Tercinta, semoga karya sederhana ini menjadi sumbangsih kecil yang bermanfaat, bagi perkembangan ilmu pengetahuan dan kemajuan almamater tercinta, dan dapat memberi manfaat yang lebih luas.

Dengan harapan yang tulus, semoga skripsi ini tidak hanya menjadi bukti perjuangan tetapi juga memberi manfaat dan menjadi awal yang baik untuk masa depan yang lebih gemilang.



KATA PENGANTAR

Dengan mengucapkan puji syukur kehadirat Allah SWT yang telah melimpahkan Rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan skripsi ini yang berjudul “ **Optimasi Hyperparameter Support Vector Machine (SVM) Menggunakan Grid Search dan Cross-Validation Untuk Meningkatkan Akurasi Klasifikasi Sentimen Ulasan Aplikasi Amikom One** “. Penulisan skripsi ini dibuat sebagai salah satu syarat kelulusan program Sarjana Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penulis menyadari bahwa tanpa bimbingan dan dorongan dari semua pihak, maka penulisan skripsi ini tidak akan berjalan lancar. Oleh karena itu, pada kesempatan ini, izinkanlah penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Hanif Al-Fatta, M.Kom, Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
3. Ibu Windha Mega Pradnya Duhita, M.Kom., selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta serta sebagai Dosen Pembimbing Skripsi yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan skripsi ini. Terima kasih atas ilmu dan pengetahuan yang telah diberikan.
4. Bapak M. Rudyanto Arief, S.T, M.T., selaku dosen wali yang telah membantu dan mengarahkan penulis dari semester awal hingga semester akhir.
5. Seluruh Staf Pengajar dan Pegawai Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah memberikan bantuan dan dukungan kepada penulis selama menjalani studi di Universitas Amikom Yogyakarta. Semua kontribusi kalian sangat berarti bagi penulis dalam meraih keberhasilan ini.
6. Keluarga Tercinta, Ayah, Ibu, Kakak, serta seluruh keluarga, terima kasih atas doa, dukungan, dan kasih sayang yang tidak pernah putus. Kalian adalah sumber

motivasi dan kekuatan yang memberikan penulis semangat untuk terus berusaha dan menyelesaikan skripsi ini dengan baik.

Penulis menyadari bahwa skripsi ini jauh dari kesempurnaan, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi dunia akademik, serta menjadi kontribusi kecil dalam pengembangan ilmu pengetahuan, khususnya dalam bidang klasifikasi sentimen dan machine learning.

Yogyakarta, 16 Desember 2024



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	xvi
INTISARI	xvii
<i>ABSTRACT</i>	xviii
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	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5

2.2	Dasar Teori	15
2.2.1	Data Mining	15
2.2.3	Analisis Sentimen	16
2.2.4	Text Mining	16
2.2.5	Web Scraping	16
2.2.6	Preprocessing Data	17
2.2.7	Term Frequency-Inverse Document Frequency (TF-IDF)	18
2.2.8	Synthetic Minority Over-Sampling Technique (SMOTE)	19
2.2.9	Support Vector Machine (SVM)	19
2.2.10	Grid Search	19
2.2.11	K-Fold Cross Validation	20
BAB III METODE PENELITIAN		22
3.1	Objek Penelitian	22
3.2	Alur Penelitian	22
3.2.1	Scraping	23
3.2.2	Pelabelan Data	23
3.2.3	Preprocessing Data	23
3.2.4	TF-IDF (Term Frequency-Inverse Document Frequency)	24
3.2.5	WordCloud	24
3.2.6	SMOTE (Synthetic Minority Over-sampling Technique)	25
3.2.7	Klasifikasi Support Vector Machine (SVM)	25
3.2.8	K-Fold Cross Validation	25
3.2.9	Hyperparameter tuning dengan Grid Search	26
3.2.10	Penerapan Parameter Pada K-Fold Cross Validation	26
3.2.11	Perbandingan Sebelum dan Sesudah Optimasi	27

3.3	Alat dan Bahan	27
3.3.1	Data Penelitian	27
3.3.2	Alat/instrumen	28
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Scraping	30
4.2	Pelabelan manual	30
4.3	Preprocessing Data	32
4.3.1	Case Folding	32
4.3.2	Data Cleaning	33
4.3.3	Stopword Removal	34
4.3.4	Tokenizing	35
4.3.5	Stemming	35
4.4	TF-IDF	36
4.5	SMOTE (Synthetic Minority Over-sampling Technique)	37
4.6	K-Fold Cross Validation	39
4.7	Hyperparameter tuning dengan Grid Search	41
4.8	Penerapan Parameter pada K-Fold Cross Validation	42
4.9	Perbandingan Sebelum dan Sesudah Optimasi	44
BAB V PENUTUP		46
5.1	Kesimpulan	46
5.2	Saran	47
REFERENSI		48
LAMPIRAN		52

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 4. 1 Hasil Scraping	30
Tabel 4. 2 Hasil Pelabelan Manual	31
Tabel 4. 3 Hasil Case Folding	33
Tabel 4. 4 Hasil Data Cleaning	33
Tabel 4. 5 Hasil Stopword Removal	34
Tabel 4. 6 Hasil Tokenizing	35
Tabel 4. 7 Hasil Stemming	36
Tabel 4. 8 Hasil TF-IDF	36
Tabel 4. 9 Skor TF-IDF	37
Tabel 4. 10 Hasil Proses 10-Fold Cross Validation Tanpa Optimasi	39
Tabel 4. 11 Parameter Terbaik SVM	42
Tabel 4. 12 Hasil Optimasi Menggunakan Grid Search	42
Tabel 4. 13 Hasil Proses K-Fold Cross Validation Setelah Optimasi	43
Tabel 4. 14 Perbandingan Hasil Evaluasi Sebelum dan Sesudah Optimasi Hyperparameter	44

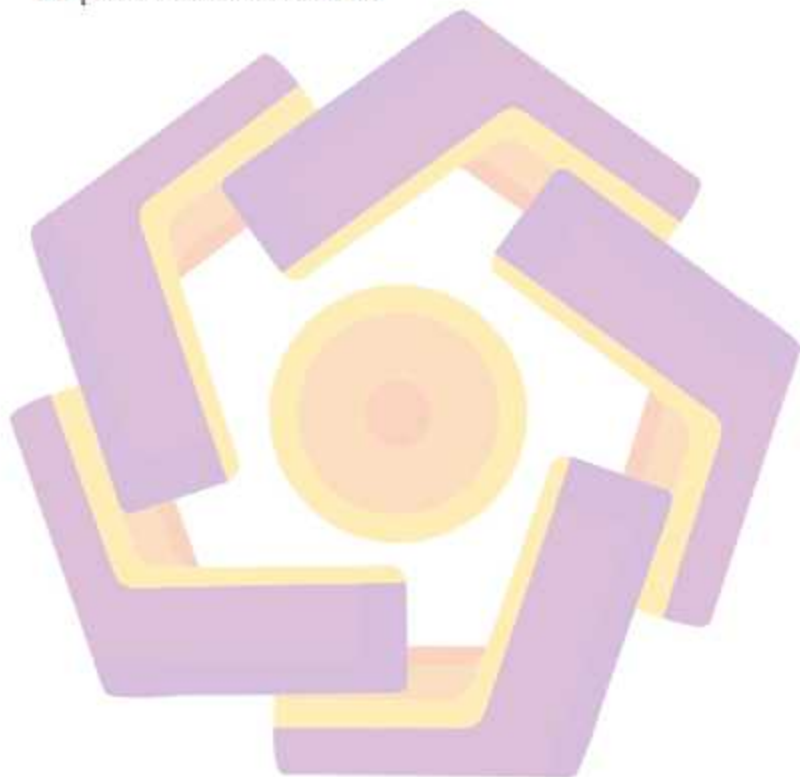
DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Grid Search Cross Validation	20
Gambar 2. 2 k-fold cross-validation dengan k =10	21
Gambar 3. 1 Alur Penelitian	22
Gambar 4. 1 Visualisasi hasil pelabelan manual	32
Gambar 4. 2 Visualisasi dari skor TF-IDF	37
Gambar 4. 3 Sebelum Proses SMOTE	38
Gambar 4. 4 Hasil Setelah Proses SMOTE	39
Gambar 4. 5 Grafik Hasil 10-Fold Cross Validation	41
Gambar 4. 6 Hasil Evaluasi Setelah Proses Optimasi	44



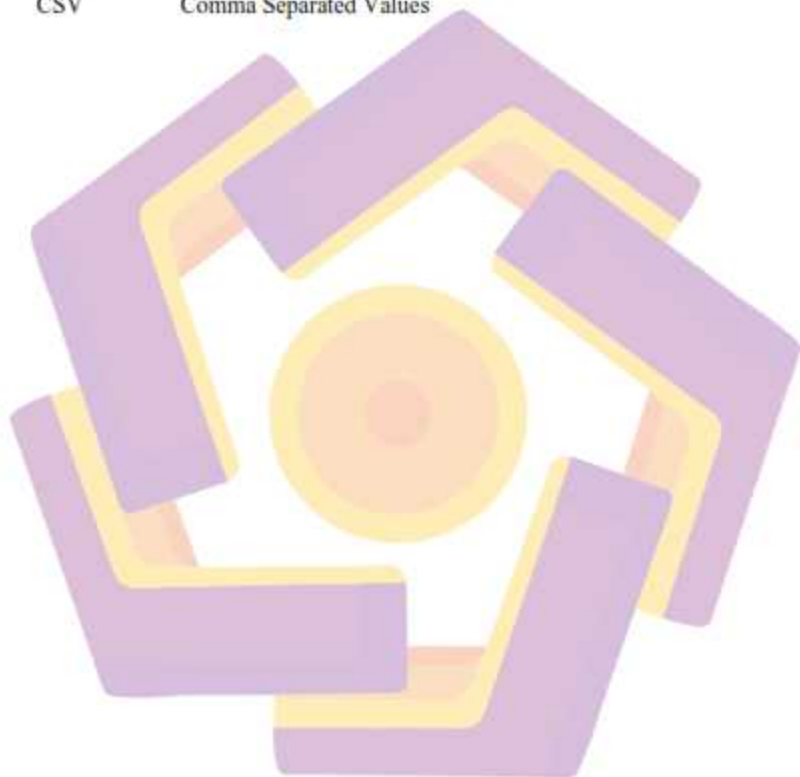
DAFTAR LAMPIRAN

Lampiran 1 Source Code Scraping (Pengumpulan Data)	52
Lampiran 2 Data Hasil Scraping	52
Lampiran 3 Source Code Penelitian	52



DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
SMOTE	Syntethic Minority Over-sampling Technique
CSV	Comma Separated Values



DAFTAR ISTILAH

Optimasi	Proses peningkatan kinerja model.
Hyperparameter	Parameter yang ditetapkan sebelum pelatihan model.
Klasifikasi	Pengelompokan data ke kategori.
Sentimen	Perasaan dalam teks (positif/negatif).
Vektor	besaran yang mempunyai arah
Overfitting	Skor akurasi tinggi pada data training, tetapi skor rendah pada data testing
Underfitting	Skor akurasi yang rendah pada data training dan data testing
C	Hyperparameter SVM yang mengatur keseimbangan antara margin dan kesalahan klasifikasi
Gamma	Hyperparameter yang mengatur jangkauan pengaruh satu data di SVM.
Kernel	Fungsi untuk memetakan data ke ruang dimensi lebih tinggi.
Akurasi	Persentase prediksi yang benar dari total prediksi.
Precision	Proporsi prediksi positif yang benar dari semua prediksi positif.
Recall	Proporsi data positif yang berhasil terdeteksi dari semua data positif.
F1 Score	Rata-rata harmonis antara precision dan recall.

INTISARI

Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi sentimen ulasan aplikasi Amikom One menggunakan model Support Vector Machine (SVM) melalui optimasi hyperparameter. Masalah yang dihadapi adalah akurasi model SVM yang cenderung menurun jika parameter seperti C , kernel, dan γ tidak diatur dengan tepat, yang dapat mempengaruhi kemampuan model dalam mengklasifikasikan sentimen dengan baik. Metode Grid Search digunakan untuk mengevaluasi kombinasi parameter secara sistematis, sementara Cross-Validation diterapkan untuk meningkatkan konsistensi model dan mengurangi overfitting.

Proses penelitian ini meliputi pengumpulan data ulasan aplikasi Amikom One melalui teknik *scraping* menggunakan Google Play Scraper, yang kemudian dilanjutkan dengan proses pelabelan manual data ulasan menjadi kategori sentimen positif dan negatif. Setelah itu, dilakukan *preprocessing data*, yang mencakup *case folding*, *data cleaning*, *stopword removal*, *tokenizing*, dan *stemming* serta pembobotan kata menggunakan TF-IDF. Untuk menangani ketidakseimbangan kelas, teknik SMOTE diterapkan untuk menyeimbangkan distribusi antara kelas positif dan negatif. Model SVM kemudian dioptimalkan menggunakan Grid Search untuk mendapatkan kombinasi hyperparameter terbaik. Evaluasi model dilakukan menggunakan teknik 10-fold Cross-Validation untuk memastikan hasil yang lebih stabil dan mengurangi risiko overfitting.

Hasil penelitian menunjukkan bahwa optimasi hyperparameter pada model SVM berhasil meningkatkan akurasi klasifikasi dari 81,50% menjadi 83,50%. Selain akurasi, *precision*, *recall*, dan *F1-score* meningkat signifikan, dari masing-masing nilai 81,88%, 81,50%, dan 80,71% menjadi 83,72%, 83,50%, dan 83,19% setelah optimasi. Hasil penelitian ini menunjukkan bahwa metode Grid Search terbukti efektif dalam menemukan kombinasi parameter terbaik, sedangkan Cross-Validation memastikan hasil yang lebih stabil dan andal.

Kata kunci: Klasifikasi sentimen, Optimasi hyperparameter, Support Vector Machine, Grid Search, Cross-Validation

ABSTRACT

This study aims to improve the sentiment classification accuracy of Amikom One application reviews using the Support Vector Machine (SVM) model through hyperparameter optimization. The issue faced is that the SVM model's accuracy tends to decrease if parameters such as C, kernel, and gamma are not properly tuned, which can affect the model's ability to classify sentiment accurately. The Grid Search method was used to systematically evaluate parameter combinations, while Cross-Validation was applied to enhance model consistency and reduce overfitting.

The research process involved collecting Amikom One application review data through scraping using Google Play Scraper, followed by manual labeling of the reviews into positive and negative sentiment categories. Data preprocessing was then performed, including case folding, data cleaning, stopword removal, tokenization, and stemming, as well as term weighting using TF-IDF. To address class imbalance, the SMOTE technique was applied to balance the distribution between positive and negative classes. The SVM model was then optimized using Grid Search to find the best hyperparameter combination. Model evaluation was conducted using 10-fold Cross-Validation to ensure more stable results and reduce the risk of overfitting.

The results showed that hyperparameter optimization in the SVM model successfully increased classification accuracy from 81.50% to 83.50%. In addition to accuracy, precision, recall, and F1-score also increased significantly, from 81.88%, 81.50%, and 80.71% to 83.72%, 83.50%, and 83.19% after optimization. The results indicate that the Grid Search method was effective in finding the best parameter combination, while Cross-Validation ensured more stable and reliable results.

Keywords: *Sentiment classification, Hyperparameter optimization, Support Vector Machine, Grid Search, Cross-Validation*