

**IMPLEMENTASI ALGORITMA SVM DAN OPTIMASI
MENGUNAKAN KOMPARASI N-GRAM DAN GLOVE
PADA SENTIMEN PENGGUNA APLIKASI ACCESS BY KAI**

SKRIPSI

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



disusun oleh

RAKA KUSUMA WARDANA

21.11.3896

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**IMPLEMENTASI ALGORITMA SVM DAN OPTIMASI
MENGUNAKAN KOMPARASI N-GRAM DAN GLOVE
PADA SENTIMEN PENGGUNA APLIKASI ACCESS BY KAI**

SKRIPSI

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



disusun oleh

RAKA KUSUMA WARDANA

21.11.3896

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI ALGORITMA SVM DAN OPTIMASI
MENGUNAKAN KOMPARASI N-GRAM DAN GLOVE PADA
SENTIMEN PENGGUNA APLIKASI ACCESS BY KAI**

yang disusun dan diajukan oleh

Raka Kusuma Wardana

21.11.3896

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Januari 2025

Dosen Pembimbing



Nuri Cahyono, M.Kom.

NIK. 190302278

HALAMAN PENGESAHAN

SKRIPSI

**IMPLEMENTASI ALGORITMA SVM DAN OPTIMASI
MENGUNAKAN KOMPARASI N-GRAM DAN GLOVE PADA
SENTIMEN PENGGUNA APLIKASI ACCESS BY KAI**

yang disusun dan diajukan oleh

Raka Kusuma Wardana

21.11.3896

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Januari 2025

Susunan Dewan Penguji

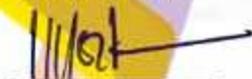
Nama Penguji

Uyock Anggoro Saputro, S.Kom., M.Kom.
NIK. 190302419

Ahmad Sa'di, S.Kom., M.Eng.
NIK. 190302459

Nuri Cahyono, M.Kom.
NIK. 190302278

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 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 : Raka Kusuma Wardana

NIM : 21.11.3896

Menyatakan bahwa Skripsi dengan judul berikut:

Implementasi Algoritma SVM Dan Optimasi Menggunakan Komparasi N-Gram Dan GloVe Pada Sentimen Pengguna Aplikasi Access by KAI

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, 22 Januari 2025

Yang Menyatakan,



Raka Kusuma Wardana

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "IMPLEMENTASI ALGORITMA SVM DAN OPTIMASI MENGGUNAKAN KOMPARASI N-GRAM DAN GLOVE PADA SENTIMEN PENGGUNA APLIKASI ACCESS BY KAI". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi S1-Informatika di Universitas Amikom Yogyakarta.

Penulis menyadari bahwa skripsi ini tidak akan terselesaikan tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

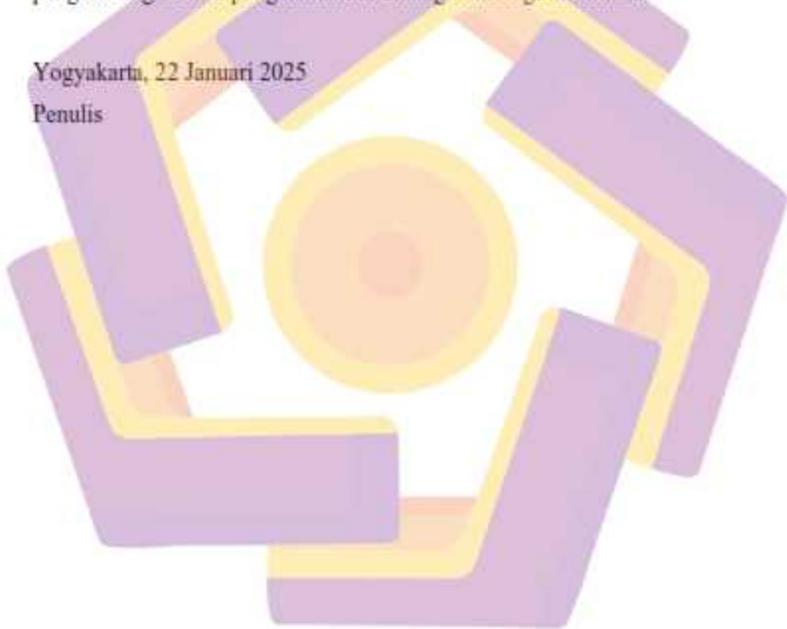
1. Allah SWT yang telah memberikan Rahmat, Hidayah, serta PetunjukNya sehingga penulis bisa melewati segala rintangan dan halangan.
2. Bapak prof. Dr. M. Suyanto,M.M. selaku rektor Universitas Amikom Yogyakarta.
3. Bapak Hanif Al fatta,S.Kom.,M.Kom.,Ph.D. selaku Dekan fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Arif Dwi Laksito, M.Kom. selaku dosen wali yang telah memberikan pengarahan kepada penulis sejak awal hingga akhir masa perkuliahan.
5. Bapak Nuri Cahyono, M.Kom. selaku dosen pembimbing saya yang telah memberikan waktu,saran,kritik,motivasi dan bimbingan dalam skripsi ini sehingga dapat menyelesaikan skripsi selesai dengan baik.
6. Bapak Udiyono dan Ibu Nuning Sri Uji Jumiwati selaku kedua orang tua penulis yang telah mendidik dan memfasilitasi segala keperluan yang diperlukan oleh penulis.

7. Teman-teman kelas 21-IF-02 di Universitas Amikom Yogyakarta, yang selalu memberikan semangat, dukungan, dan kebersamaan selama proses penelitian dan penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, penulis dengan senang hati menerima saran dan kritik yang membangun dari pembaca demi perbaikan di masa mendatang. Semoga skripsi ini bermanfaat bagi semua pihak yang membutuhkan, khususnya dalam pengembangan ilmu pengetahuan di bidang teknologi informasi.

Yogyakarta, 22 Januari 2025

Penulis



DAFTAR ISI

HALAMAN JUDUL	i
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	15
2.2.1 Analisis Sentimen	15
2.2.2 Access by KAI	15
2.2.3 Support Vector Machine (SVM)	16

2.2.4.	N-Gram	18
2.2.5.	Word Embedding GloVe	20
2.2.6.	Evaluasi Akurasi Model Klasifikasi	22
2.2.7.	Draw.io.....	22
BAB III METODE PENELITIAN		23
3.1	Alur Penelitian.....	23
3.1.1	Scraping Data.....	24
3.1.2	Labeling Data.....	24
3.1.3	Exploratory Data Analysis (EDA).....	26
3.1.4	Preprocessing Data.....	26
3.1.5	Ekstraksi Fitur.....	29
3.1.6	Implementasi Model.....	35
3.1.7	Evaluasi Model.....	36
3.1.8	Analisis Hasil.....	37
3.2	Alat dan Bahan.....	38
3.2.1	Alat/instrumen.....	38
3.2.2	Bahan Penelitian.....	39
BAB IV HASIL DAN PEMBAHASAN		41
4.1	Pengumpulan Data.....	41
4.2	Pelabelan Sentimen.....	44
4.3	Exploratory Data Analysis.....	46
4.3.1	Pemeriksaan Struktur Dataset.....	46
4.3.2	Distribusi Label Sentimen.....	47
4.3.3	Visualisasi Distribusi Sentimen.....	47
4.3.4	Kesimpulan dari Analisis Awal.....	48

4.4	Preprocessing Data	49
4.4.1	Normalisasi Data	49
4.4.2	Tokenizing	52
4.4.3	Stopword Removal	53
4.4.4	Stemming	54
4.5	Hasil Evaluasi SVM Tanpa Optimasi	55
4.6	Modeling SVM menggunakan N-Gram TF-IDF	59
4.6.1	Proses Training dan Testing	60
4.6.2	Hasil Evaluasi Model	61
4.6.3	Analisis Hasil N-Gram TF-IDF	63
4.7	Modeling SVM menggunakan GloVe Embedding	64
4.7.1	Proses Training dan Testing	65
4.7.2	Hasil Evaluasi Model	67
4.7.3	Analisis Hasil GloVe	68
4.8	Perbandingan Hasil	70
BAB V PENUTUP		75
5.1	Kesimpulan	75
5.2	Saran	76
REFERENSI		77
LAMPIRAN		80

DAFTAR TABEL

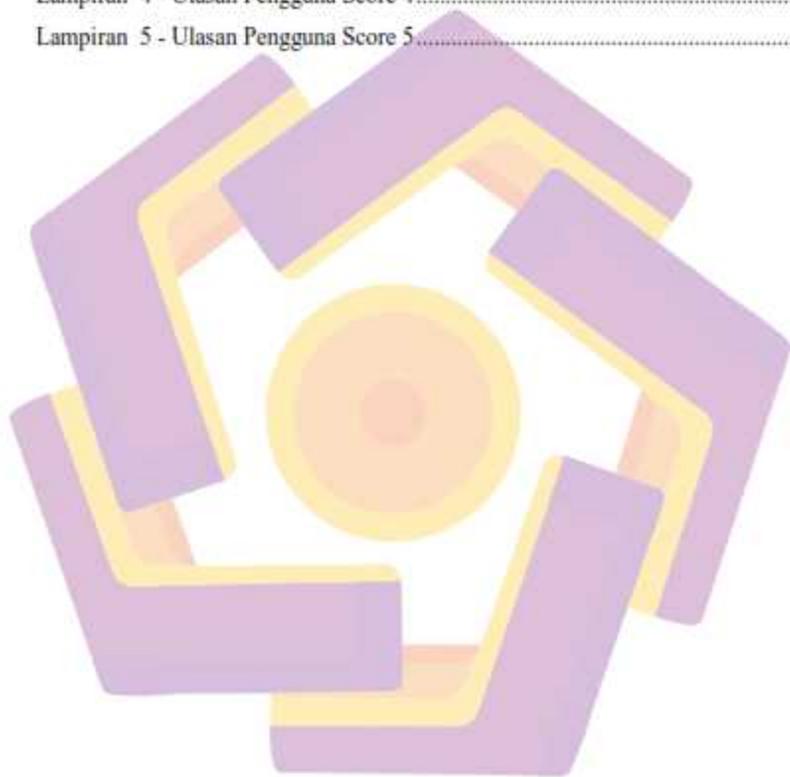
Tabel 2. 1 - Keaslian Penelitian	9
Tabel 3. 1 - Labeling Data	25
Tabel 3. 2 - Normalisasi	27
Tabel 4. 1 - Hasil Scraping	42
Tabel 4. 2 - Jumlah Data yang Dipakai	43
Tabel 4. 3 - Sentimen Positif	44
Tabel 4. 4 - Sentimen Negatif	45
Tabel 4. 5 - Hasil Normalisasi	49
Tabel 4. 6 - Hasil Tokenisasi	53
Tabel 4. 7 - Hasil Stopword Removal	54
Tabel 4. 8 - Hasil Stemming	55
Tabel 4. 9 - Hasil Klasifikasi SVM Tanpa Optimasi	58
Tabel 4. 10 - Hasil Klasifikasi N-Gram	62
Tabel 4. 11 - Hasil Klasifikasi N-Gram TF-IDF	63
Tabel 4. 12 - Hasil Klasifikasi GloVe Kernel Linear	67
Tabel 4. 13 - Hasil Klasifikasi GloVe	69
Tabel 4. 14 - Perbandingan Hasil Evaluasi Model SVM, N-Gram TF-IDF dan GloVe	70

DAFTAR GAMBAR

Gambar 2. 1 - Ilustrasi Support Vector Machine.....	16
Gambar 2. 2 – Seleksi Fitur dengan Kombinasi N-Gram.....	19
Gambar 3. 1 - Alur Penelitian.....	23
Gambar 3. 3 - Alur Tokenisasi.....	28
Gambar 3. 4 - Alur Stopword Removal.....	28
Gambar 3. 5 - Alur Stemming.....	29
Gambar 3. 6 - Kombinasi Nilai C dan Jenis Kernel.....	35
Gambar 4. 1 - Kode Tahap Pengumpulan Data.....	41
Gambar 4. 2 - Diagram Hasil Labeling.....	44
Gambar 4. 3 - Kode Program Tahap EDA.....	46
Gambar 4. 4 - Distribusi Sentimen Ulasan.....	48
Gambar 4. 5 - Kode Program Tahapan Normalisasi.....	51
Gambar 4. 6 - Kode Program Tahapan Tokenizing.....	52
Gambar 4. 7 - Kode Program Tahap Stopword Removal.....	53
Gambar 4. 8 - Kode Program Tahap Stemming.....	54
Gambar 4. 9 – Kode Program Tahap SVM Tanpa Optimasi.....	56
Gambar 4. 10 - Confusion Matrix SVM Tanpa Optimasi.....	58
Gambar 4. 11 – Kode Program Tahap N-Gram TF-IDF.....	60
Gambar 4. 12 - Confusion Matrix N-Gram TF-IDF.....	62
Gambar 4. 13 – Kode Program Tahap Embedding GloVe.....	66
Gambar 4. 14 - Confusion Matrix Glove Linear.....	68
Gambar 4. 15 - Diagram Perbandingan Hasil Evaluasi Model.....	72

DAFTAR LAMPIRAN

Lampiran 1 - Ulasan Pengguna Score 1.....	80
Lampiran 2- Ulasan Pengguna Score 2.....	80
Lampiran 3 - Ulasan Pengguna Score 3.....	81
Lampiran 4 - Ulasan Pengguna Score 4.....	81
Lampiran 5 - Ulasan Pengguna Score 5.....	81



DAFTAR LAMBANG DAN SINGKATAN

N	Jumlah dokumen dalam korpus (TF-IDF)
$f(t, d)$	Frekuensi kemunculan suatu term t dalam dokumen d
$P(x)$	Probabilitas suatu kejadian x (model embedding)
$w(t, d)$	Bobot term dalam dokumen berdasarkan TF-IDF
X_{ij}	Matriks co-occurrence antar kata
V	Dimensi atau ukuran vektor embedding
C	Parameter regularisasi dalam Support Vector Machine
w	Vektor bobot dalam model SVM
b	Bias dalam model SVM
$K(x_i, x_j)$	Fungsi kernel untuk pemetaan data
$TF(t, d)$	Proporsi kemunculan term dalam dokumen
$IDF(t)$	Pentingnya term dalam korpus
$w(t, d)$	Bobot term dalam dokumen (TF-IDF)
X_{ij}	Matriks hubungan antar kata (GloVe)
$P(x)$	Probabilitas suatu kejadian x
\sum	Simbol penjumlahan data
$\text{Tanh}(x)$	Fungsi aktivasi sigmoid
$ w $	Panjang vektor w
e	Bilangan Euler
\log	Logaritma untuk perhitungan
IDF	Inverse Document Frequency
TF	Term Frequency
SVM	Support Vector Machines
EDA	Exploratory Data Analysis
NLP	Natural Language Processing
GloVe	Global Vectors for Word Representation
CSV	Comma Separated Values
KAI	Kereta Api Indonesia

DAFTAR ISTILAH

Access by KAI	Aplikasi digital PT KAI untuk layanan kereta
Accuracy	Persentase prediksi benar terhadap total data
Analisis Sentimen	Proses memahami opini dalam teks
CSV	Format file data berbasis teks
EDA	Eksplorasi awal untuk memahami dataset
Embedding	Representasi kata dalam bentuk vektor
F1-Score	Rata-rata harmonis dari precision dan recall
GloVe	Teknik untuk merepresentasikan kata dalam vektor
Hyperplane	Garis atau bidang pemisah dalam SVM
Kernel	Fungsi untuk memisahkan data non-linear
Klasifikasi	Proses membagi data ke dalam kategori
Machine Learning	Pemrograman komputer berbasis data
N-Gram	Urutan kata atau karakter dalam teks
Normalization	Penyesuaian teks ke bentuk seragam
Overfitting	Model terlalu menyesuaikan data pelatihan
Precision	Proporsi prediksi positif yang benar
Preprocessing	Pengolahan awal data mentah
Recall	Mengukur seberapa baik model menemukan data relevan
Stemming	Mengubah kata ke bentuk dasar
Stopword Removal	Menghapus kata umum yang tidak penting
Supervised Learning	Pembelajaran dengan data berlabel
Support Vector	Vektor penentu margin klasifikasi
SVM	Algoritma klasifikasi berbasis margin
Text Mining	Proses ekstraksi informasi teks
TF-IDF	Bobot kata berdasarkan kemunculannya
Tokenization	Memecah teks menjadi unit kecil seperti kata
Web Scraping	Pengambilan data otomatis dari web

INTISARI

Penelitian ini bertujuan untuk mengoptimalkan metode klasifikasi Support Vector Machine (SVM) dalam analisis sentimen terhadap aplikasi Access by KAI dengan memanfaatkan pendekatan N-Gram dan GloVe. Analisis sentimen sangat penting untuk memahami opini pengguna, mengidentifikasi keluhan, dan preferensi yang dapat digunakan untuk meningkatkan kualitas layanan. Dataset yang digunakan dalam penelitian ini terdiri dari 1718 ulasan pengguna yang telah dilabeli secara manual, dengan fokus pada sentimen positif dan negatif yang dikumpulkan antara tahun 2023 hingga 2024. Proses penelitian mencakup pengumpulan data, pelabelan, analisis eksploratori, preprocessing, pemodelan SVM, serta evaluasi hasil.

Hasil penelitian menunjukkan bahwa pendekatan N-Gram berbasis TF-IDF menghasilkan akurasi terbaik sebesar 84,88% dengan nilai precision negatif 0,88, recall negatif 0,90, dan F1-score negatif 0,89, menandakan kemampuannya dalam menangkap pola lokal dengan baik. Di sisi lain, GloVe embedding hanya mencapai akurasi 79,65%, dengan precision negatif 0,766, recall negatif 0,82, dan F1-score negatif 0,79. Sebagai perbandingan, pemodelan SVM tanpa optimasi mencapai akurasi 83,43%, dengan precision negatif 0,80, recall negatif 0,91, dan F1-score negatif 0,85. Perbandingan ini menegaskan bahwa pemilihan metode ekstraksi fitur yang tepat sangat berpengaruh terhadap kinerja klasifikasi dalam analisis sentimen.

Analisis sentimen ini mengungkapkan bahwa sentimen negatif mendominasi dengan persentase 51,7% dibandingkan dengan sentimen positif yang hanya 48,3%. Temuan ini mencerminkan tantangan yang dihadapi aplikasi dalam hal performa dan fitur, serta memberikan wawasan berharga untuk pengembangan layanan yang lebih responsif dan relevan bagi pengguna. Penelitian ini menekankan pentingnya optimasi metode klasifikasi dan ekstraksi fitur dalam meningkatkan akurasi dan efisiensi analisis sentimen, yang pada gilirannya dapat mendukung transformasi digital dalam layanan publik.

Kata kunci : Analisis Sentimen, SVM, N-Gram, Glove, Optimasi Klasifikasi.

ABSTRACT

This study aims to optimize the Support Vector Machine (SVM) classification method in sentiment analysis of the Access by KAI application by utilizing the N-Gram and GloVe approaches. Sentiment analysis is very important for understanding user opinions, identifying complaints, and preferences that can be used to improve service quality. The dataset used in this study consists of 1718 manually labeled user reviews, focusing on positive and negative sentiments collected between 2023 and 2024. The research process includes data collection, labeling, exploratory analysis, preprocessing, SVM modeling, and evaluation of results.

The results showed that the TF-IDF-based N-Gram approach produced the best accuracy of 84.88% with a negative precision value of 0.88, a negative recall of 0.90, and a negative F1-score of 0.89, indicating its ability to capture local patterns well. On the other hand, GloVe embedding only achieved an accuracy of 79.65%, with a negative precision of 0.766, a negative recall of 0.82, and a negative F1-score of 0.79. In comparison, SVM modeling without optimization achieved an accuracy of 83.43%, with a negative precision of 0.80, a negative recall of 0.91, and a negative F1-score of 0.85. This comparison confirms that choosing the right feature extraction method greatly affects the classification performance in sentiment analysis.

This sentiment analysis revealed that negative sentiment dominated with a percentage of 51.7% compared to positive sentiment which was only 48.3%. This finding reflects the challenges faced by the application in terms of performance and features, and provides valuable insights for the development of more responsive and relevant services for users. This study emphasizes the importance of optimizing classification and feature extraction methods in improving the accuracy and efficiency of sentiment analysis, which in turn can support digital transformation in public services.

Keywords : *Sentiment Analysis, SVM, N-Gram, Glove, Classification Optimization.*