

**OPTIMASI SUPPORT VECTOR MACHINE MENGGUNAKAN
SMOTE UNTUK MENANGANI KETIDAKSEIMBANGAN
DATA PADA KLASIFIKASI SENTIMEN
SKRIPSI**

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



disusun oleh

AJI NUGROHO ASSHIDIQ

19.11.3186

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**OPTIMASI SUPPORT VECTOR MACHINE MENGGUNAKAN
SMOTE UNTUK MENANGANI KETIDAKSEIMBANGAN
DATA PADA KLASIFIKASI SENTIMEN
SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AJI NUGROHO ASSHIDIQ

19.11.3186

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**HALAMAN PERSETUJUAN
SKRIPSI**

**OPTIMASI SUPPORT VECTOR MACHINE MENGGUNAKAN
SMOTE UNTUK MENANGANI KETIDAKSEIMBANGAN
DATA PADA KLASIFIKASI SENTIMENT**

yang disusun dan diajukan oleh

Aji Nugroho Asshidiq

19.11.3186

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Februari 2026

Dosen Pembimbing,



Anna Baita, S.Kom, M.Kom
NIK. 190302290

HALAMAN PENGESAHAN

SKRIPSI

**OPTIMASI SUPPORT VECTOR MACHINE MENGGUNAKAN
SMOTE UNTUK MENANGANI KETIDAKSEIMBANGAN
DATA PADA KLASIFIKASI SENTIMENT**

yang disusun dan diajukan oleh

Aji Nugroho Asshidiq

19.11.3186

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bayu Setiaji, M.Kom
NIK. 190302216

Muhammad Tofa Nurcholis, S.Kom, M.Kom
NIK. 190302281

Anna Baita, S.Kom, M.Kom
NIK. 190302290



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aji Nugroho Asshidiq
NIM : 19.11.3186

Menyatakan bahwa Skripsi dengan judul berikut:

Optimasi Support Vector Machine Menggunakan SMOTE Untuk Menangani Ketidakseimbangan Data Pada Klasifikasi Sentimen

Dosen Pembimbing: Anna Baita, S.Kom, 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, 26 Februari 2026

Yang Menyatakan,



Aji Nugroho Asshidiq

MOTTO

“Keberhasilan bukan milik orang pintar. Keberhasilan milik mereka yang terus berusaha”. – B. J. Habibie

"Orang yang meraih kesuksesan tidak selalu orang yang pintar. Orang yang selalu meraih kesuksesan adalah orang yang gigih dan pantang menyerah". – Susi Pudjiastuti

“Perang telah usai, aku bisa pulang
Kubaringkan panah dan berteriak Menang!”. – Nadin Amizah

HALAMAN PERSEMBAHAN

Puji dan Syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas anugerah dan nikmat yang tak terkira sehingga penulis dapat menyelesaikan karya tulis ini. Pada kesempatan ini penulis ingin menyampaikan ucapan terimakasih yang sebesar-besarnya kepada:

1. Orang tua saya yang selalu memberikan dukungan, doa, dan cinta tanpa henti
2. Keluarga besar penulis yang selalu menjadi sumber inspirasi dan semangat
3. Dosen dan staf pengajar di Program Studi Teknik Informatika Universitas Amikom Yogyakarta yang telah memberikan ilmu dan pengalaman berharga selama masa studi
4. Teman-teman seangkatan dan seluruh mahasiswa Universitas Amikom Yogyakarta yang selalu memberikan dukungan dan motivasi selama perkuliahan dan penyusunan skripsi ini. Penulis mengucapkan terima kasih yang sebesar-besarnya atas dukungan, doa, dan semangat yang telah diberikan. Semoga Allah SWT senantiasa memberikan keberkahan dan kesejahteraan kepada kita semua.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul "*Optimasi Support Vector Machine Menggunakan SMOTE untuk Menangani Ketidakseimbangan Data Pada Klasifikasi Sentimen*". Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan studi di Program Sarjana Universitas Amikom Yogyakarta.

Dengan selesainya skripsi ini, maka penulis tidak lupa mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusri, M.Kom, selaku Dekan Fakultas Ilmu Komputer.
3. Anna Baita, S.Kom, M.Kom, selaku dosen pembimbing yang telah banyak memberikan pengarahan bagi penulis dalam pembuatan skripsi
4. Bapak-Ibu Dosen Universitas Amikom Yogyakarta, atas ilmu dan pengetahuan yang telah diberikan kepada penulis selama duduk di bangku kuliah.
5. Kedua orang tua saya, yang tak pernah berhenti untuk selalu memberikan selamat serta doa.
6. Seluruh teman-teman kuliah saya, terutama untuk kelas IF 10, terimakasih atas kerjasamanya selama ini.
7. Teman-teman seperjuangan yang telah memberikan dukungan dan terimakasih atas momen momen yang tak pernah terlupakan.

Penulis tentunya menyadari bahwa pembuatan skripsi ini masih banyak sekali kekurangan-kekurangan dan kelemahan-kelemahannya. Oleh karena itu penulis berharap kepada semua pihak agar dapat menyampaikan kritik dan saran yang membangun untuk menambah kesempurnaan skripsi ini. Namun penulis tetap berharap skripsi ini akan bermanfaat bagi semua pihak yang membacanya.

Yogyakarta, 26 Februari 2026

Penulis

DAFTAR ISI

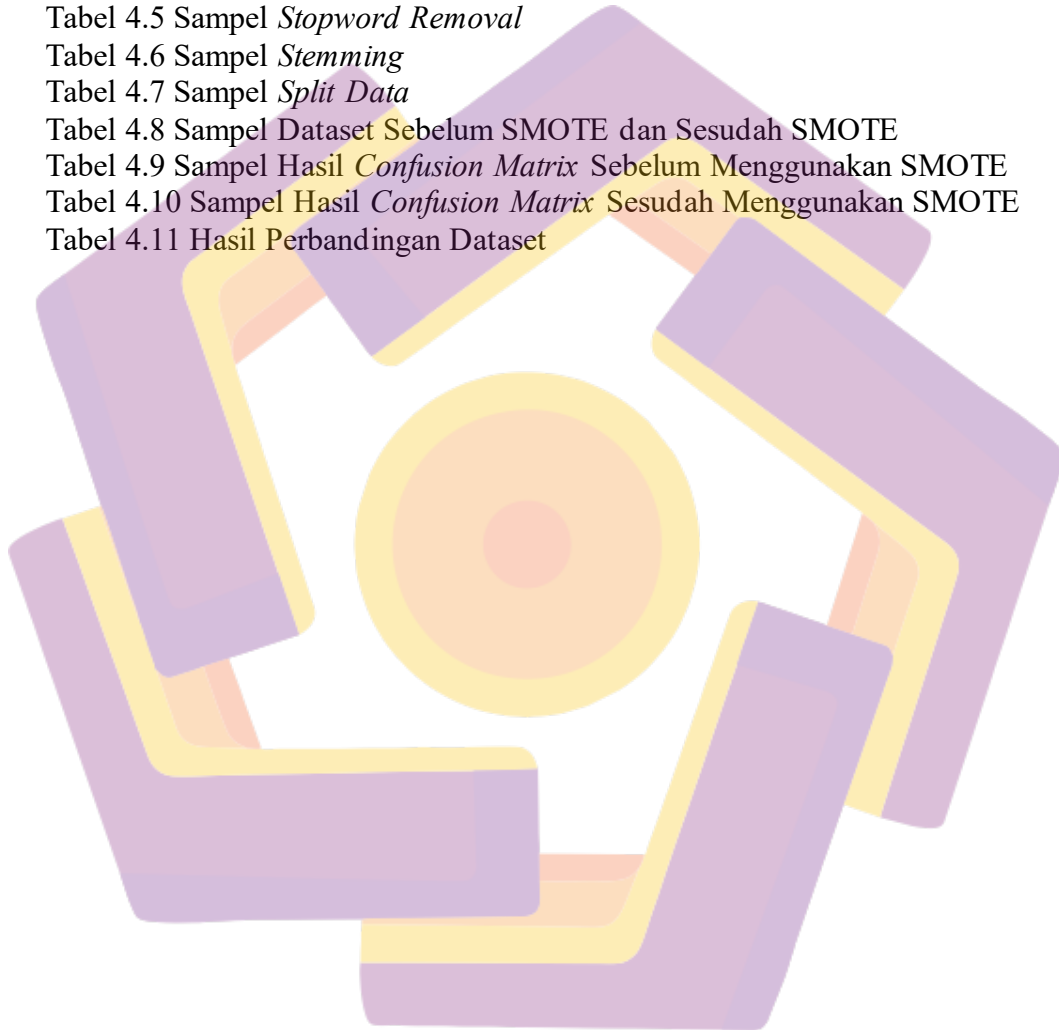
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	2
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	4
2.1 Studi Literatur	4
2.2 Dasar Teori.....	9

2.2.1 Sentimen Analisis	9
2.2.2 Dataset.....	9
2.2.3 Klasifikasi	9
2.2.4 Pre-processing	10
2.2.5 Pembobotan TF-IDF	11
2.2.6 Imbalance Ratio	12
2.2.7 SMOTE (Synthetic Minority Over-sampling Technique)	13
2.2.8 Text Mining.....	14
2.2.9 Support Vector Machine	15
2.2.10 Confusion Matrix	17
BAB III METODE PENELITIAN	20
3.1 Objek Penelitian	20
3.2 Alur Penelitian	20
3.2.1 Identifikasi Masalah	21
3.2.2 Pengumpulan Data	22
3.2.3 Praproses Data.....	22
3.2.4 Split Data.....	23
3.2.5 TF-IDF (Term Frequency – Inverse Document Frequency).....	24
3.2.6 SMOTE (Synthetic Minority Oversampling Technique).....	24
3.2.7 Klasifikasi Support Vector Machine	24
3.3 Pengujian Model.....	25
3.4 Alat dan Bahan Penelitian	25
3.4.1 Spesifikasi Perangkat Keras (Hardware)	26
3.4.2 Spesifikasi Perangkat Lunak (Software).....	26
3.5 Bahan Penelitian	27

BAB IV HASIL DAN PEMBAHASAN	28
4.1 Hasil Pengumpulan Data	28
4.1.1 Akuisisi Data Melalui Repositori GitHub.....	29
4.1.2 Integrasi Dataset Melalui API Kaggle	29
4.2 Preprocessing Data	30
4.2.1 Case Folding	30
4.2.2 Cleaning	31
4.2.3 Tokenizing	31
4.2.4 Stopword Removal.....	32
4.2.5 Stemming	33
4.3 Split Data.....	34
4.4 TF-IDF (Term Frequency-Inverse Document Frequency).....	35
4.5 SMOTE (Synthetic Minority Over-sampling Technique).....	38
4.6 Hasil Confusion Matrix	39
BAB V PENUTUP	43
5.1 Kesimpulan	43
5.2 Saran	43
REFERENSI	44

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian
Tabel 4.1 Pengumpulan Data
Tabel 4.2 Sampel <i>Case Folding</i>
Tabel 4.3 Sampel <i>Cleaning</i>
Tabel 4.4 Sampel <i>Tokenizing</i>
Tabel 4.5 Sampel <i>Stopword Removal</i>
Tabel 4.6 Sampel <i>Stemming</i>
Tabel 4.7 Sampel <i>Split Data</i>
Tabel 4.8 Sampel Dataset Sebelum SMOTE dan Sesudah SMOTE
Tabel 4.9 Sampel Hasil <i>Confusion Matrix</i> Sebelum Menggunakan SMOTE
Tabel 4.10 Sampel Hasil <i>Confusion Matrix</i> Sesudah Menggunakan SMOTE
Tabel 4.11 Hasil Perbandingan Dataset



DAFTAR GAMBAR

Gambar 2.1 Preprocessing Data

Gambar 2.2 Hyperlane Kelas Positif dan Kelas Negatif

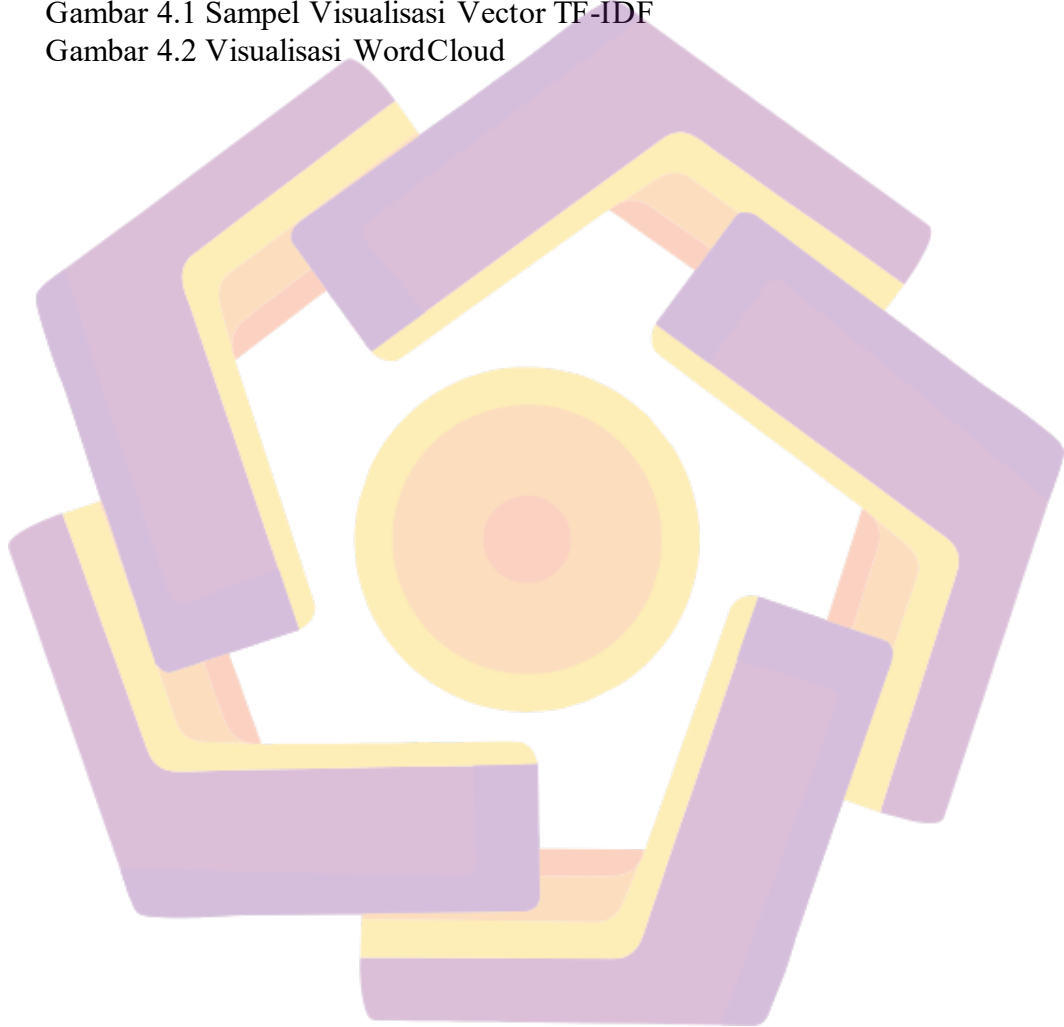
Gambar 2.3 Confusion Matrix

Gambar 3.1 Alur Penelitian

Gambar 3.2 Praproses Data

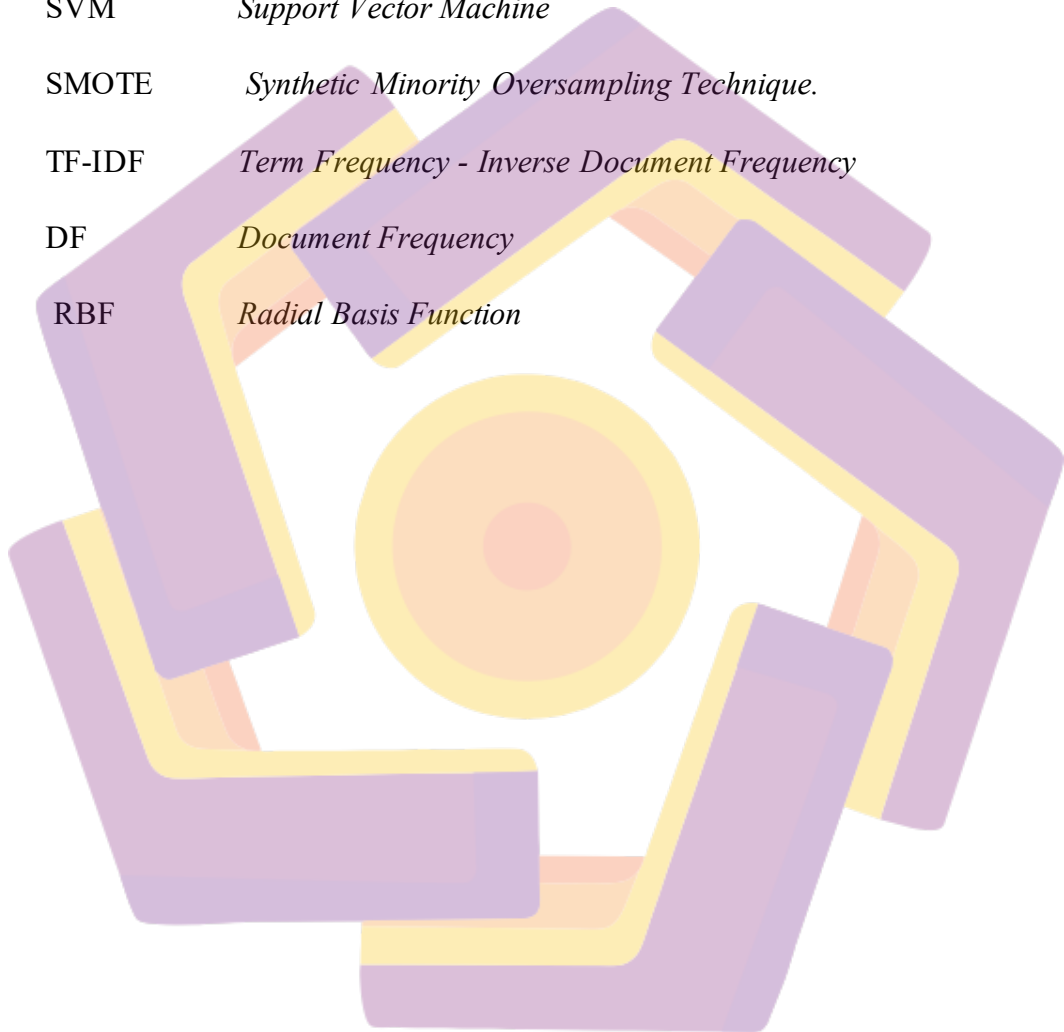
Gambar 4.1 Sampel Visualisasi Vector TF-IDF

Gambar 4.2 Visualisasi WordCloud



DAFTAR LAMBANG DAN SINGKATAN

SVM	<i>Support Vector Machine</i>
SMOTE	<i>Synthetic Minority Oversampling Technique.</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
DF	<i>Document Frequency</i>
RBF	<i>Radial Basis Function</i>



INTISARI

Analisis sentimen seringkali menghadapi masalah ketidakseimbangan kelas data (*imbalanced dataset*) yang menyebabkan model *Support Vector Machine* (SVM) cenderung bias dan kesulitan memprediksi kelas minoritas. Penelitian ini bertujuan untuk mengoptimalkan kinerja SVM dalam mengklasifikasikan data sentimen menggunakan teknik penyeimbangan data *Synthetic Minority Over-sampling Technique* (SMOTE) dan metode pembobotan teks *Term Frequency-Inverse Document Frequency* (TF-IDF). Pengujian dilakukan melalui tiga rasio pembagian data latih dan uji, yaitu 80:20 (Skenario 1), 70:30 (Skenario 2), dan 60:40 (Skenario 3). Peningkatan paling terlihat ketika menguji skenario 80:20 (Skenario 1), dimana kinerja model mencapai titik paling optimal dengan peningkatan rata-rata akurasi sebesar 82,83% dan *F1-Score* 82,42%. Kesimpulannya, penerapan SMOTE terbukti sangat efektif untuk mengatasi bias terhadap kelas mayoritas, sehingga menghasilkan sistem klasifikasi sentimen yang jauh lebih adil, stabil, dan dapat diandalkan.

Kata kunci: Analisis Sentimen, *Support Vector Machine* (SVM), *Imbalanced Dataset*, SMOTE, TF-IDF.

ABSTRACT

Sentiment analysis frequently encounters the problem of imbalanced datasets, which causes the *Support Vector Machine* (SVM) model to be biased and struggle to predict the minority class. This study aims to optimize the performance of SVM in classifying sentiment data by utilizing the Synthetic Minority Over-sampling Technique (SMOTE) for data balancing and the Term Frequency-Inverse Document Frequency (TF-IDF) method for text weighting. The evaluation was conducted through three train-test data split ratios, namely 80:20 (Scenario 1), 70:30 (Scenario 2), and 60:40 (Scenario 3). The most significant improvement was observed when testing the 80:20 split (Scenario 1), where the model's performance reached its optimal point, achieving an average accuracy of 82.83% and an *F1-Score* of 82.42%. In conclusion, the application of SMOTE is proven to be highly effective in mitigating bias toward the majority class, thereby producing a sentiment classification system that is significantly fairer, more stable, and reliable.

Keyword: *Sentiment Analysis, Support Vector Machine (SVM), Imbalanced Dataset, SMOTE, TF-IDF.*