

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI
MYSMARTFREN DI GOOGLE PLAYSTORE
MENGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN)**

SKRIPSI SARJANA SISTEM INFORMASI



**Oleh
Cikal Rifaldi
(217006516067)**

**PROGRAM STUDI SISTEM INFORMASI FAKULTAS
TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
2025**

HALAMAN PENGESAHAN
TUGAS AKHIR

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI MYSMARTFREN DI
GOOGLE PLAYSTORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN)**



Dosen Pembimbing

Ira Diana Bholihati, S.Si., MMSI.

HALAMAN PERNYATAAN ORISINALITAS

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI MYSMARTFREN DI
GOOGLE PLAYSTORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN)**

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

UNIVERSITAS NASIONAL

Jakarta, 2 September 2025


METERAI TEMPEL
F71ANX020627808
Cikal Rifaldi
217006516067

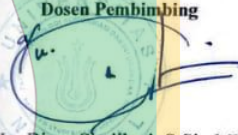
HALAMAN PENGESAHAN TUGAS SARJANA

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI MYSMARTFREN DI
GOOGLE PLAYSTORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN)**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Genap 2024-2025 pada tanggal 27 Agustus Tahun 2025

Dosen Pembimbing

Ira Diana Sholihati, S.Si., MMSI.
NIDN. 0328037304

Ketua Program Studi

Dr. Ir. Andrianingsih, S.Kom., MMSI.
NIDN. 0303097902

UNIVERSITAS NASIONAL

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Cikal Rifaldi

NPM : 217006516067

Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika

Program Studi : Sistem Informasi

Tanggal Sidang : 27 Agustus 2025

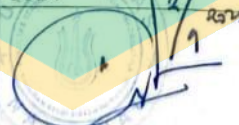
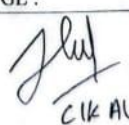
JUDUL DALAM BAHASA INDONESIA :

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI
MYSMARTFREN DI GOOGLE PLAYSTORE MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST
NEIGHBOR (KNN)**

JUDUL DALAM BAHASA INGGRIS :

**SENTIMENT CLASSIFICATION OF USER REVIEWS OF THE
MYSMARTFREN APPLICATION ON GOOGLE PLAYSTORE USING
SUPPORT VECTOR MACHINE (SVM) AND K-NEAREST NEIGHBOR (KNN)
ALGORITHMS**

TANDA TANGAN DAN TANGGAL

Dosen Pembimbing	Ka. Prodi	Mahasiswa
TGL : 27/8/2025	TGL : 27/8/2025	TGL :
		 C IK AL

KATA PENGANTAR

Segala puji dan Syukur penulis panjatkan kepada Allah SWT atas segala Rahmat, nikmat, berkah, dan kemudahan yang telah di berikan, sehingga penulis dapat menyelesaikan skripsi yang berjudul :KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI MYSMARTFREN DI GOOGLE PLAYSTORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN), skripsi ini merupakan salah satu syarat kelulusan untuk Program Studi Sistem Informasi di Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.

Penyusunan skripsi ini tidak akan terlaksana tanpa dukungan dan bantuan dari berbagai pihak yang telah mengorbankan waktu, tenaga, pemikiran, dan sumber daya. Oleh karena itu, penulis ingin menyampaikan ucapan terimakasih yang sebesar-besarnya kepada:

1. Allah SWT yang telah melimpahkan Rahmat dan nikmat -Nya sehingga skripsi ini dapat terselesaikan.
2. Kedua orang tua yang telah merawat, mendampingi, mendidik dan memberikan dukungan materi dan doa yang tiada henti kepada penulis.
3. Ibu Ira Diana Sholihati, S.Si., MMSI., dosen pembimbing di prodi Sistem Informasi Universitas Nasional.
4. Seluruh dosen pengajar Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, terimakasih atas segala pengetahuan dan wawasan yang telah diberikan selama masa perkuliahan.
5. Sinta, Nouval, Isal, Maldy, Haidar, Fatarur, Hilmy, Alfiansyah, Zahy, Wasis, Rheza, Zidan, Vivancita dan semua pihak yang tidak dapat disebutkan satu per satu. Selaku kerabat yang selalu memberikan semangat, berbagi ilmu, motivasi, serta membantu dalam proses penyusunan skripsi ini sehingga dapat berjalan dengan lancar.

6. Diri penulis sendiri, yang telah berjuang atas kerja keras, rasa lelah, kebingungan, dan keraguan, namun tetap berusaha konsisten hingga skripsi ini selesai.

Penulis menyadari bahwa skripsi ini masih terdapat kekurangan dan belum sepenuhnya sempurna. Oleh sebab itu, penulis sangat mengharapkan kritik, saran, dan masukan yang membangun untuk penyempurnaan di masa mendatang. Penulis juga berharap karya ini dapat memberikan manfaat bagi semua pihak yang berkepentingan.

Jakarta, 23 Agustus 2025

Cikal Rifaldi



ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap aplikasi MySmartfren menggunakan algoritma *Support Vector Machine* dan *K-Nearest Neighbor*, serta mengevaluasi kinerjanya berdasarkan akurasi, presisi, *recall*, dan *F1-score*. Data sebanyak 3000 ulasan yang diperoleh melalui proses *crawling* menggunakan Google Play Scraper, yang kemudian melalui tahapan *preprocessing* yang mencakup *remove duplicates*, *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, *stemming*, penerjemahan, pelabelan otomatis menggunakan TextBlob, dan pembobotan data menggunakan metode TF-IDF, sehingga model dapat mengklasifikasikan ulasan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Hasil pengujian menunjukkan bahwa algoritma SVM memberikan akurasi sebesar 79%, dengan presisi, *recall*, dan *F1-score* yang konsisten, sedangkan algoritma KNN menghasilkan akurasi sebesar 55%. Temuan ini menunjukkan bahwa SVM lebih unggul dibanding KNN dalam klasifikasi sentimen ulasan pengguna aplikasi MySmartfren. Penelitian ini diharapkan menjadi referensi bagi pengembang untuk meningkatkan kualitas aplikasi berdasarkan umpan balik pengguna yang dianalisis secara otomatis.

Kata kunci: analisis sentimen, MySmartfren, Support Vector Machine, K-Nearest Neighbor, Google Playstore.

ABSTRACT

This study aims to classify public sentiment towards the MySmartfren application using the Support Vector Machine and K-Nearest Neighbor algorithms, and evaluate their performance based on accuracy, precision, recall, and F1-score. Data of 3000 reviews were obtained through a crawling process using Google Play Scraper, which then went through preprocessing stages including removing duplicates, cleaning, case folding, normalization, tokenizing, stopword removal, stemming, translation, automatic labeling using TextBlob, and data weighting using the TF-IDF method, so that the model can classify reviews into three sentiment categories, namely positive, negative, and neutral. The test results showed that the SVM algorithm provided an accuracy of 79%, with consistent precision, recall, and F1-score, while the KNN algorithm produced an accuracy of 55%. These findings indicate that SVM is superior to KNN in classifying sentiments of MySmartfren application user reviews. This study is expected to be a reference for developers to improve the quality of applications based on automatically analyzed user feedback.

Keywords: sentiment analysis, MySmartfren, Support Vector Machine, K-Nearest Neighbor, Google Playstore.

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	ii
HALAMAN PENGESAHAN TUGAS SARJANA	iv
KATA PENGANTAR	vi
ABSTRAK.....	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Batasan Masalah.....	5
BAB II TINJAUAN	7
2.1 Studi Literatur.....	7
2.2 Landasan Teori	35
2.2.1 Klasifikasi	35
2.2.2 Ulasan Pengguna.....	35
2.2.3 MySmartfren	35
2.2.4 Google PlayStore	36

2.2.5	Text Mining.....	36
2.2.6	Phyton	37
2.2.7	Google Collab	37
2.2.8	Machine Learning	38
2.2.9	Support Vector Machine	38
2.2.10	K-Nearest Neighbor	39
BAB III METODOLOGI PENELITIAN		40
3.1	Lokasi Penelitian	40
3.2	Waktu Penelitian	40
3.3	Metode Penelitian.....	40
3.4	Teknik Pengumpulan Data	40
3.5	Fokus Penelitian	41
3.6	Tahapan Penelitian	42
3.6.1	Crawling.....	43
3.6.2	Preprocessing	43
3.6.3	Data Preparation.....	46
3.6.4	Klasifikasi Algoritma Support Vector Machine	55
3.6.5	Klasifikasi Algoritma K-Nearest Neighbor	59
3.6.6	Confusion Matrix	63
3.6.7	Visualisasi Data.....	65
BAB IV HASIL DAN PEMBAHASAN		66
4.1	Hasil Analisis	66
4.1.1	Hasil Crawling	66
4.1.2	Hasil Preprocessing.....	69
4.1.3	Hasil Labeling	80

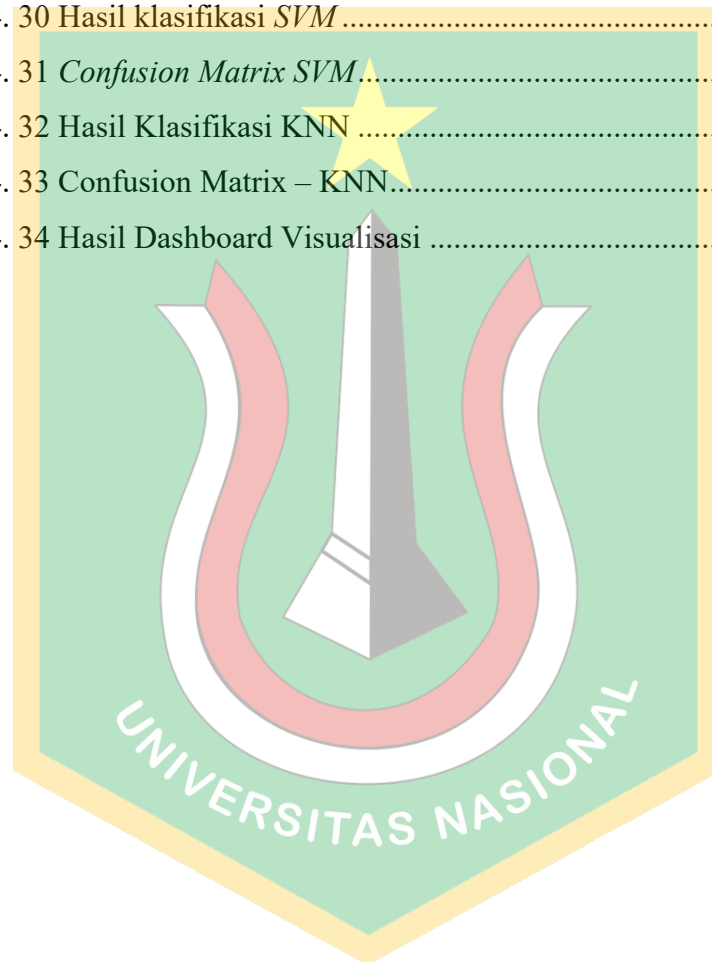
4.1.4	Hasil Pembobotan (Weighting).....	84
4.1.5	Hasil Klasifikasi Algoritma Support Vector Machine (SVM).....	85
4.1.6	Hasil Klasifikasi Algoritma K-Nearest Neighbor (K-NN)	87
4.2	Tampilan Visualisasi	89
4.3	Pembahasan	90
BAB V KESIMPULAN DAN SARAN.....		92
5.1	Kesimpulan.....	92
5.2	Saran.....	92
DAFTAR PUSTAKA		94
LAMPIRAN.....		98



DAFTAR GAMBAR

Gambar 2. 1 Logo MySmartfren.....	36
Gambar 2. 2 Logo Google PlayStore	36
Gambar 2. 3 Logo Phytton	37
Gambar 2. 4 Logo Google Collab	38
Gambar 3. 1 Tahapan Penelitian	42
Gambar 4. 1 <i>Package Name MySf</i>	66
Gambar 4. 2 <i>Scrapping data</i>	67
Gambar 4. 3 Hasil Web Scrapping.....	67
Gambar 4. 4 Diagram Hasil Crawling	68
Gambar 4. 5 <i>Source Code Remove Duplicates</i>	69
Gambar 4. 6 Hasil Dari <i>Remove Duplicates</i>	69
Gambar 4. 7 <i>Source Code Cleaning Data 1</i>	70
Gambar 4. 8 <i>Source Code Cleaning Data 2</i>	70
Gambar 4. 9 <i>Source Code Cleaning Data 3</i>	71
Gambar 4. 10 <i>Source Code Cleaning Data 4</i>	71
Gambar 4. 11 Hasil <i>Cleaning Data</i>	72
Gambar 4. 12 <i>Source Code Case Folding</i>	72
Gambar 4. 13 Hasil <i>Case Folding</i>	73
Gambar 4. 14 Source code Normalisasi	73
Gambar 4. 15 Hasil Normalisasi	74
Gambar 4. 16 <i>Source Code Tokenize</i>	74
Gambar 4. 17 Hasil Tokenizing	75
Gambar 4. 18 Source code Filter Stopword	76
Gambar 4. 19 Hasil Filter Stopword	76
Gambar 4. 20 <i>Source Code Stemming Data</i>	77
Gambar 4. 21 Hasil <i>Stemming Data</i>	78
Gambar 4. 22 <i>Source Code Translate Data</i>	79
Gambar 4. 23 Hasil <i>Translate Data</i>	79

Gambar 4. 24 Proses <i>Labeling</i>	80
Gambar 4. 25 Hasil <i>Labeling</i>	81
Gambar 4. 26 Diagram distribusi data perkelas	82
Gambar 4. 27 Diagram data latih dan data uji	82
Gambar 4. 28 <i>Wordcloud Dataset</i>	83
Gambar 4. 29 Hasil <i>Weighting</i>	84
Gambar 4. 30 Hasil klasifikasi <i>SVM</i>	85
Gambar 4. 31 <i>Confusion Matrix SVM</i>	86
Gambar 4. 32 Hasil Klasifikasi KNN	87
Gambar 4. 33 Confusion Matrix – KNN.....	88
Gambar 4. 34 Hasil Dashboard Visualisasi	89



DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	7
Tabel 3. 1 <i>Timeline</i> Penelitian.....	40
Tabel 3. 2 TF Dokumen 1	47
Tabel 3. 3 TF Dokumen 2	48
Tabel 3. 4 TF Dokumen 4	49
Tabel 3. 5 TF Dokumen 4	49
Tabel 3. 6 TF Dokumen 5	50
Tabel 3. 7 Sampel data latih	56
Tabel 3. 8 Sampel data uji.....	56
Tabel 3. 9 Gabungan Kata Relevan.....	56
Tabel 3. 10 Vektorisasi	56
Tabel 3. 11 Tabel Representasi Numerik Pada Ulasan Positif	57
Tabel 3. 12 Tabel Representasi Numerik Pada Ulasan Negatif.....	57
Tabel 3. 13 Tabel Representasi Numerik Pada Ulasan Netral	58
Tabel 3. 14 Sampel Data Latih.....	60
Tabel 3. 15 Tabel Representasi numerik Pada ulasan Positif	61
Tabel 3. 16 Tabel Representasi Numerik Pada Ulasan Negatif.....	61
Tabel 3. 17 Tabel Representasi Numerik Pada Ulasan Netral	61
Tabel 3. 18 Tabel klasifikasi ulasan baru.....	62