

**“PERBANDINGAN ANALISIS SENTIMEN REVIEW APLIKASI
STREAMING FILM NETFLIX PADA GOOGLE PLAY STORE
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN K
– NEAREST NEIGHBOR”**

SKRIPSI SARJANA SISTEM INFORMASI



Oleh :

Harits Ahmad Fauzan

217006516012

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

**PERBANDINGAN ANALISIS SENTIMEN REVIEW APLIKASI
STREAMING FILM NETFLIX PADA GOOGLE PLAY STORE
MENGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE DAN K – NEAREST NEIGHBOR**

SKRIPSI SARJANA SISTEM INFORMASI

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi dari
Fakultas Teknologi Komunikasi dan Informatika

Oleh

Harits Ahmad Fauzan

217006516024



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

HALAMAN PENGESAHAN

TUGAS AKHIR

PERBANDINGAN ANALISIS SENTIMEN REVIEW APLIKASI
STREAMING FILM NETFLIX PADA GOOGLE PLAY STORE
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN K -
NEAREST NEIGHBOR



Harits Ahmad Fauzan

217006516012

Dosen Pembimbing

(Novi Dian Nathasia, S.Kom., MMSI)

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

PERBANDINGAN ANALISIS SENTIMEN REVIEW APLIKASI STREAMING FILM NETFLIX PADA GOOGLE PLAY STORE MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN K – NEAREST NEIGHBOR

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.

Jakarta, 30 Agustus 2025


10000
MBI 10000
TEMPER
509C8AKX705911026
Harits Ahmad Fauzan
217006516012

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

PERBANDINGAN ANALISIS SENTIMEN REVIEW APLIKASI STREAMING FILM NETFLIX PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN K – NEAREST

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 Ganjil 2024-2025 pada tanggal 27 Agustus Tahun 2025

NEIGHBOR

Dosen Pembimbing


Novi Dian Nathasia, S.Kom.,MMSI
NIDN. 0714127701

Ketua Program Studi


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

UNIVERSITAS NASIONAL

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Harits Ahmad Fauzan
NPM : 217006516012
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : Selasa, 27 Agustus 2025

JUDUL DALAM BAHASA INDONESIA :

PERBANDINGAN ANALISIS SENTIMEN REVIEW APLIKASI STREAMING
FILM NETFLIX PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE DAN K - NEAREST NEIGHBOR

JUDUL DALAM BAHASA INGGRIS :

COMPARISON OF SENTIMENT ANALYSIS OF NETFLIX MOVIE STREAMING
APPLICATION REVIEWS ON GOOGLE PLAY STORE USING SUPPORT
VECTOR MACHINE AND K-NEAREST NEIGHBOR ALGORITHMS

TANDA TANGAN DAN TANGGAL		
Pembimbing	Ka. Prodi	Mahasiswa
TGL : 	TGL :  5/9/2025	TGL : 

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Perbandingan Analisis Sentimen Review Aplikasi Streaming Film Netflix pada Google Play Store Menggunakan Algoritma Support Vector Machine dan K-Nearest Neighbor”**. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan program Sarjana, Jurusan Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika.

Dalam proses penyusunan skripsi ini, penulis mendapatkan banyak bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya terutama kepada dosen pembimbing Tugas Akhir, Bapak Novi Dian Nathasia, S.Kom., MMSI atas waktu, tenaga, dan pikiran dalam membimbing penulis selama penelitian dan penulisan skripsi. Penulis turut menyampaikan rasa terima kasih kepada :

1. Allah SWT yang telah memberikan nikmat panjang umur, kesehatan, kemudahan dan kelancaran dalam menyelesaikan penelitian serta penulisan skripsi ini dengan sebaik mungkin.
2. Kedua orang tua penulis, Bapak Wasino dan Ibu Ida Farida yang telah memberikan dukungan dalam berbagai bentuk yang tak terhitung.
3. Bapak Novi Dian Nathasia, selaku Dosen Pembimbing Skripsi.
4. Seluruh dosen pengajar di Program Studi Sistem Informasi FTKI maupun dosen di Program Studi lain yang telah memberikan banyak ilmu.
5. Niken Sabrina, sebagai teman yang sudah memberikan bantuan serta dukungan.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun. Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi kontribusi yang berguna dalam bidang analisis data.

Jakarta, 14 Februari 2025

Harits Ahmad Fauzan

ABSTRAK

Perkembangan teknologi informasi telah mendorong pertumbuhan layanan streaming film seperti Netflix yang menjadi pilihan utama masyarakat dalam menikmati hiburan secara online. Ulasan pengguna di Google Play Store menjadi sumber informasi yang berharga untuk mengevaluasi kepuasan dan pengalaman pengguna. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Netflix berbahasa Indonesia. Data penelitian berjumlah 5.000 ulasan yang diperoleh melalui teknik web scraping, kemudian melalui tahap preprocessing berupa data cleansing, case folding, normalisasi, penghapusan stopword, tokenizing, dan stemming. Pelabelan sentimen dilakukan dengan pendekatan Lexicon-Based menjadi tiga kelas (positif, netral, negatif). Selanjutnya, data dibagi menjadi training set dan testing set dengan rasio 80:20 serta direpresentasikan menggunakan TF-IDF. Evaluasi kinerja model dilakukan menggunakan confusion matrix dan metode k-fold cross-validation dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan distribusi sentimen terdiri dari 36,11% netral, 34,29% positif, dan 29,59% negatif. Algoritma SVM (linear) memperoleh rata-rata akurasi sebesar 90,87%, precision sebesar 91,08%, recall sebesar 90,82%, dan F1-score sebesar 90,93%, dengan performa yang relatif stabil pada setiap fold. Sementara itu, algoritma KNN (k=5) menghasilkan akurasi rata-rata sebesar 60,86%, precision 72,96%, recall 59,24%, dan F1-score 59,23%, dengan variasi performa antar fold yang lebih besar. Perbandingan ini menunjukkan bahwa SVM memiliki performa lebih unggul dan konsisten dalam mengklasifikasikan sentimen dibandingkan KNN, sehingga dapat menjadi metode yang lebih efektif untuk analisis sentimen ulasan aplikasi berbasis pengguna.

Kata kunci: Analisis Sentimen, Support Vector Machine, K-Nearest Neighbor, Cross Validation, Netflix, Google Play Store, *Text Mining*.

ABSTRACT

The development of information technology has driven the growth of film streaming services such as Netflix, which has become one of the main choices for people to enjoy entertainment online. User reviews on the Google Play Store serve as a valuable source of information to evaluate user satisfaction and experience. This study aims to analyze and compare the performance of the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms in classifying Indonesian-language Netflix user reviews. A total of 5,000 reviews were collected through web scraping and then processed through several preprocessing steps, including data cleansing, case folding, normalization, stopword removal, tokenizing, and stemming. Sentiment labeling was performed using a Lexicon-Based approach into three classes (positive, neutral, negative). The dataset was then split into training and testing sets with an 80:20 ratio and represented using TF-IDF before being processed by both algorithms. Model performance was evaluated using a confusion matrix and k-fold cross-validation with accuracy, precision, recall, and F1-score as evaluation metrics. The results show that the sentiment distribution consists of 36.11% neutral, 34.29% positive, and 29.59% negative reviews. The SVM (linear) algorithm achieved an average accuracy of 90.87%, precision of 91.08%, recall of 90.82%, and F1-score of 90.93%, with relatively stable performance across folds. Meanwhile, the KNN algorithm ($k=5$) achieved an average accuracy of 60.86%, precision of 72.96%, recall of 59.24%, and F1-score of 59.23%, with greater performance variations between folds. These findings indicate that SVM outperforms KNN and demonstrates more consistent performance in sentiment classification, making it a more effective method for analyzing user-generated application reviews.

Keywords: *Sentiment Analysis, Support Vector Machine, K-Nearest Neighbor, Cross Validation, Netflix, Google Play Store, Text Mining.*

DAFTAR ISI

KATA PENGANTAR	1
ABSTRAK.....	2
ABSTRACT	3
DAFTAR ISI.....	4
DAFTAR GAMBAR.....	7
DAFTAR TABEL.....	9
BAB I	10
1.1 Latar Belakang	10
1.2 Identifikasi Masalah.....	12
1.3 Rumusan Masalah.....	13
1.4 Tujuan Penelitian	13
1.5 Batasan Masalah	13
1.6 Kontribusi	14
BAB II.....	15
2.1 Studi Literatur	15
2.2 Landasan Teori.....	18
2.2.1 Analisis Sentimen	18
2.2.2 Natural Language Processing	19
2.2.3 Web Scraping	19
2.2.4 Support Vector Machine.....	20
2.2.5 K – Nearest Neighbor	20
2.2.6 Visual Code Studio	21
2.2.7 Python	21
2.2.8 Streamlit.....	22
2.2.9 Google Play Store	22
2.2.10 Evaluasi Model Klasifikasi	23
BAB III	27
3.1 Lokasi Penelitian & Timeline Penelitian	27
3.1.1 Lokasi Penelitian.....	27
3.1.2 <i>Timeline</i> Penelitian.....	27

3.1.2	<i>Timeline</i> Penelitian.....	27
3.2	Fokus Penelitian.....	27
3.3	Populasi & Sampel Penelitian.....	28
3.4	Teknik Pengumpulan Data.....	28
3.4.1	Studi literatur	28
3.4.2	Jenis Data	28
3.4.3	Web Scraping	28
3.5	Desain Penelitian	28
3.6	Metode Evaluasi Hasil	32
3.6.1	Confusion Matrix	33
3.7	Perangkat Penelitian.....	34
BAB IV		35
4.1	Pengumpulan Data.....	35
4.2	Preprocessing Data.....	36
4.2.1	<i>Data Cleansing</i>	36
4.2.2	Case Folding	38
4.2.3	Normalisasi	39
4.2.4	Tokenization.....	40
4.2.5	Stopword Removal.....	41
4.2.6	Steaming	43
4.3	Data Preparation.....	44
4.3.1	Translating	44
4.3.2	Labeling	46
4.4	Implementasi Algoritma.....	48
4.4.1	<i>Splitting</i> Data dan TF – IDF.....	48
4.4.2	Perhitungan Manual Algoritma.....	49
4.4.3	Penerapan K – Nearest Neighbor.....	54
4.4.4	Penerapan Support Vector Machine.....	56
4.5	Analisis Hasil dan Evaluasi	58
4.5.1	Distribusi Sentimen.....	60
4.5.2	<i>WordCloud</i> Modus Kata.....	61
4.5.3	Analisis Hasil Sentimen Positif	62
4.5.4	Analisis Hasil Sentimen Netral.....	64
4.5.5	Analisis Hasil Sentimen Negatif.....	66
4.5.6	<i>Heatmap Confussion Matrix KNN & SVM</i>	68
4.6	Deployment Model Klasifikasi Sederhana	69

<u>BAB V</u>	73
5.1 <u>Kesimpulan</u>	73
5.2 <u>Saran</u>	74
<u>DAFTAR PUSTAKA</u>	75



DAFTAR GAMBAR

Gambar 3. 1 Desain Penelitian	29
Gambar 3. 2 Logika penelitian	32
Gambar 4 1 Instant Data Scraper	35
Gambar 4 2 proses crawling	35
Gambar 4 3 Hasil Scraping	36
Gambar 4 4 Data Cleansing	37
Gambar 4 5 Gambar 4. 5 Hasil Cleansing	38
Gambar 4 6 Case Folding	38
Gambar 4 7 Hasil Case Folding	39
Gambar 4 8 Normalisasi	39
Gambar 4 9 Hasil Normalisasi	40
Gambar 4 10 Tokenization	40
Gambar 4 11 Hasil Tokenization	41
Gambar 4 12 Stopword Removal	42
Gambar 4 13 Hasil Stopword Removal	42
Gambar 4 14 Steaming	43
Gambar 4 15 Hasil Steaming	44
Gambar 4 16 Translating	45
Gambar 4 17 Hasil Translating	45
Gambar 4 18 Hasil Vader Sentimen	46
Gambar 4 19 Labeling	46
Gambar 4 20 Hasil Labelling	47
Gambar 4 21 Hasil Bar Chart Lexicon-Based 3 Class	47
Gambar 4 22 Splitting Data & TfidfVectorizer	48
Gambar 4 23 Hasil Pembagian Data Latih & Uji	49
Gambar 4 24 Model K – Nearest Neighbor	54
Gambar 4 25 Classification K – Nearest Neighbor	55
Gambar 4 26 Classification matrix K – Nearest Neighbor	55
Gambar 4 27 Model Support Vector Machine	56
Gambar 4 28 Classification Support Vector Machine	57

Gambar 4 29 Classification matrix Support Vector Machine.....	57
Gambar 4 30 Cross Validation.....	59
Gambar 4 31 Hasil Cross Validation	59
Gambar 4 32 Distribusi Sentimen	60
Gambar 4 33 Wordcloud Modus Kata.....	61
Gambar 4 34 Wordcloud Modus Kata Sentimen Positif.	62
Gambar 4 35 Bar Chart Modus Kata Sentimen Positif.	62
Gambar 4 36 Network Chart Hubungan Kata Sentimen Positif.	62
Gambar 4 37 Wordcloud Modus Kata Sentimen Netral.....	64
Gambar 4 38 Bar Chart Modus Kata Sentimen Netral.....	64
Gambar 4 39 Network Chart Hubungan Kata Sentimen Netral.	64
Gambar 4 40 Wordcloud Modus Kata Sentimen Negatif.....	66
Gambar 4 41 Bar Chart Modus Kata Sentimen Negatif.....	66
Gambar 4 42 Network Chart Hubungan Kata Sentimen Negatif.....	66
Gambar 4 43 Confussion K – Nearest Neighbor.....	68
Gambar 4 44 Confussion Support Vector Machine.....	69
Gambar 4 45 Hasil Prediksi Sentimen Model KNN & SVM.	70



DAFTAR TABEL

Tabel 2. 1 Studi Literatur	15
Tabel 2. 2 Confusion Matrix	23
Tabel 2. 3 Contoh Confusion Matrix	25
 Tabel 3. 1 Timeline Penelitian	 27
Tabel 3. 2 Confussion Matrix.....	33
Tabel 3. 3 Perangkat Penelitian.....	34
 Tabel 4. 1 Sampel Data Latih K – Nearest Neighbor	 49
Tabel 4. 2 Sampel Data Uji K – Nearest Neighbor	49
Tabel 4. 3 Hasil Preprocessing K – Nearest Neighbor.....	50
Tabel 4. 4 Kata Unik K – Nearest Neighbor.....	50
Tabel 4. 5 Representasi Vektor K – Nearest Neighbor	50
Tabel 4. 6 Jarak Eucliden K – Nearest Neighbor.....	51
Tabel 4. 7 Jarak Eucliden K – Nearest Neighbor.....	51
Tabel 4. 8 Sampel Data SVM	52
Tabel 4. 9 Bag of Words.....	52
Tabel 4. 10 Contoh Vektor Fitur	52
Tabel 4. 11 Contoh Vektor Biner Ulasan Baru.....	53
Tabel 4. 12 Hasil Uji Prediksi Kalimat Baru	71



