

**PERBANDINGAN MODEL KLASIFIKASI DALAM ANALISIS
SENTIMEN PENGGUNA TWITTER TERHADAP FILM VINA:
SEBELUM 7 HARI**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

FARHAN YAHYA

21.12.2168

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**PERBANDINGAN MODEL KLASIFIKASI DALAM ANALISIS
SENTIMEN PENGGUNA TWITTER TERHADAP FILM VINA:
SEBELUM 7 HARI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

FARHAN YAHYA

21.12.2168

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN MODEL KLASIFIKASI DALAM ANALISIS
SENTIMEN PENGGUNA TWITTER TERHADAP FILM VINA:
SEBELUM 7 HARI**

yang disusun dan diajukan oleh

Farhan Yahya

21.12.2168

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 Februari 2025

Dosen Pembimbing,



Yoga Pristyanto, S.Kom, M.Eng
NIK. 190302412

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN MODEL KLASIFIKASI DALAM ANALISIS
SENTIMEN PENGGUNA TWITTER TERHADAP FILM VINA:
SEBELUM 7 HARI

yang disusun dan diajukan oleh

Farhan Yahya

21.12.2168

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Acihmah Sidauruk, S.Kom., M.Kom.
NIK. 190302238



Ikma, S.Kom., M.Kom.
NIK. 190302282



Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Februari 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 : Farhan Yahya
NIM : 21.12.2168

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN MODEL KLASIFIKASI DALAM ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP FILM VINA: SEBELUM 7 HARI

Dosen Pembimbing : Yoga Pristyanto, S.Kom, M.Eng

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, 18 Februari 2025

Yang Menyatakan,



METERAI
TEMPEL
10000
080D6AMX189713840

Farhan Yahya

HALAMAN PERSEMBAHAN

Keluarga

Almarhum bapak, ibu serta kakak-kakak saya yang selalu memberikan semangat melalui uang serta doa dan dukungan.

Sahabat

Yang saling bertukar cerita beratnya mengerjakan skripsi ini, bertukar cerita malasnya mengerjakan skripsi ini, bertukar cerita pusingnya mengerjakan skripsi ini, bertukar cerita capenya mengerjakan skripsi ini. Sehingga saya tidak merasa sendirian dan akhirnya kita saling memberi dukungan, doa, serta saling menguatkan. Selamat dan sukses yang terbaik untuk orang yang baik.

Musisi

Yang telah menciptakan lagu-lagu untuk menemani mengerjakan skripsi ini.

Streamer

Yang telah menghibur untuk menemani mengerjakan skripsi ini.

Diri sendiri

Yang telah berjuang dengan segala keterbatasan fasilitas, wawasan dan pengetahuan untuk menyelesaikan skripsi ini, kamu hebat han.

KATA PENGANTAR

Puji dan syukur saya panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi yang berjudul "Perbandingan Model Klasifikasi dalam Analisis Sentimen Pengguna Twitter terhadap Film *Vina: Sebelum 7 Hari*" ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Universitas Amikom.

Dalam penyusunan skripsi ini, saya menyadari bahwa keberhasilannya tidak terlepas dari dukungan, bimbingan, serta doa dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, saya ingin mengucapkan terima kasih kepada:

1. Bapak Prof, Dr. M. Suyanto, MM selaku rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D selaku dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Anggit Dwi Hartanto, M.Kom selaku kepala program studi Sistem Informasi Universitas Amikom Yogyakarta.
4. Bapak Yoga Pristyanto, S.Kom., M.Eng selaku dosen pembimbing.
5. Kedua orang tua dan keluarga tercinta, yang selalu memberikan doa, dukungan moral, dan semangat yang tak terhingga.
6. Teman-teman seperjuangan dan seluruh pihak yang telah membantu baik secara langsung maupun tidak langsung dalam proses penyelesaian skripsi ini.

Penulis menyadari bahwa dengan keterbatasan wawasan serta pengalaman penulis, membuat Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun. Semoga skripsi ini dapat memberikan manfaat. Akhir kata, semoga kita semua mendapatkan rahmat dan selalu ada dalam lindungan Allah SWT.

Yogyakarta, 18 Februari 2025

Farhan Yahya

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	vii
DAFTAR GAMBAR.....	xii
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	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
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.....	11
2.2.1	Film	11
2.2.2	Twitter.....	11
2.2.3	Analisis Sentimen	13
2.2.4	Preprocessing Text Mining.....	13
2.2.5	Fitur Selection.....	13
2.2.6	Pembobotan TF-IDF	14
2.2.7	Naive Bayes	15
2.2.8	Random Forest.....	15
2.2.9	Support Vector Machine	16
2.2.10	Evaluasi.....	17
BAB III METODE PENELITIAN		19
3.1	Objek Penelitian.....	19
3.2	Alur Penelitian	19
3.2.1	Data Crawling	21
3.2.2	Preprocessing Data.....	21
3.2.2.1	Case Folding	21
3.2.2.2	Data Cleaning.....	21
3.2.2.3	Normalisasi	21
3.2.2.4	Stopword Removal.....	22
3.2.2.5	Stemming	22
3.2.2.6	Tokenizing	22
3.2.3	Labeling	22
3.2.4	Pembobotan TF-IDF`	23
3.2.5	Splitting Data	23

3.2.6	Implementasi SMOTE	23
3.2.7	Implementasi Naïve Bayes.....	23
3.2.8	Implementasi Random Forest	23
3.2.9	Implementasi Support Vector Machine	24
3.2.10	Evaluasi.....	24
3.2.11	Deployment.....	24
3.3	Alat dan Bahan.....	25
BAB IV HASIL DAN PEMBAHASAN		26
4.1	Deskripsi Implementasi	26
4.2	Data Crawling	26
4.3	Implementasi Preprocessing Data.....	27
4.3.1	Data Cleaning.....	28
4.3.2	Case Folding	29
4.3.3	Normalisasi	30
4.3.4	Stopword Removal.....	31
4.3.5	Stemming	32
4.3.6	Tokenizing	33
4.4	Labeling	34
4.5	Visualisasi Data	35
4.5.1	Visualisasi Pie Chart.....	36
4.5.2	Visualisasi Word Cloud	37
4.6	Implementasi Splitting Data	41
4.7	Pembobotan TF-IDF	41
4.8	Implementasi SMOTE	43
4.9	Pengujian Model Algoritma.....	44

4.9.1	Pengujian Naive Bayes	44
4.9.1.1	Hasil Pengujian Naïve Bayes.....	45
4.9.1.2	Evaluasi Pengujian Naive Bayes	46
4.9.2	Pengujian Random Forest.....	47
4.9.2.1	Hasil Pengujian Random Forest.....	48
4.9.2.2	Evaluasi Pengujian Random Forest	49
4.9.3	Pengujian Support Vector Machine	50
4.9.3.1	Hasil Pengujian Support Vector Machine.....	51
4.9.3.2	Evaluasi Pengujian Support Vector Machine	52
4.9.4	Interprestasi Hasil	53
4.9.5	Deployment.....	54
BAB V PENUTUP		59
5.1	Kesimpulan	59
5.2	Saran	59
REFERENSI		60
LAMPIRAN.....		63

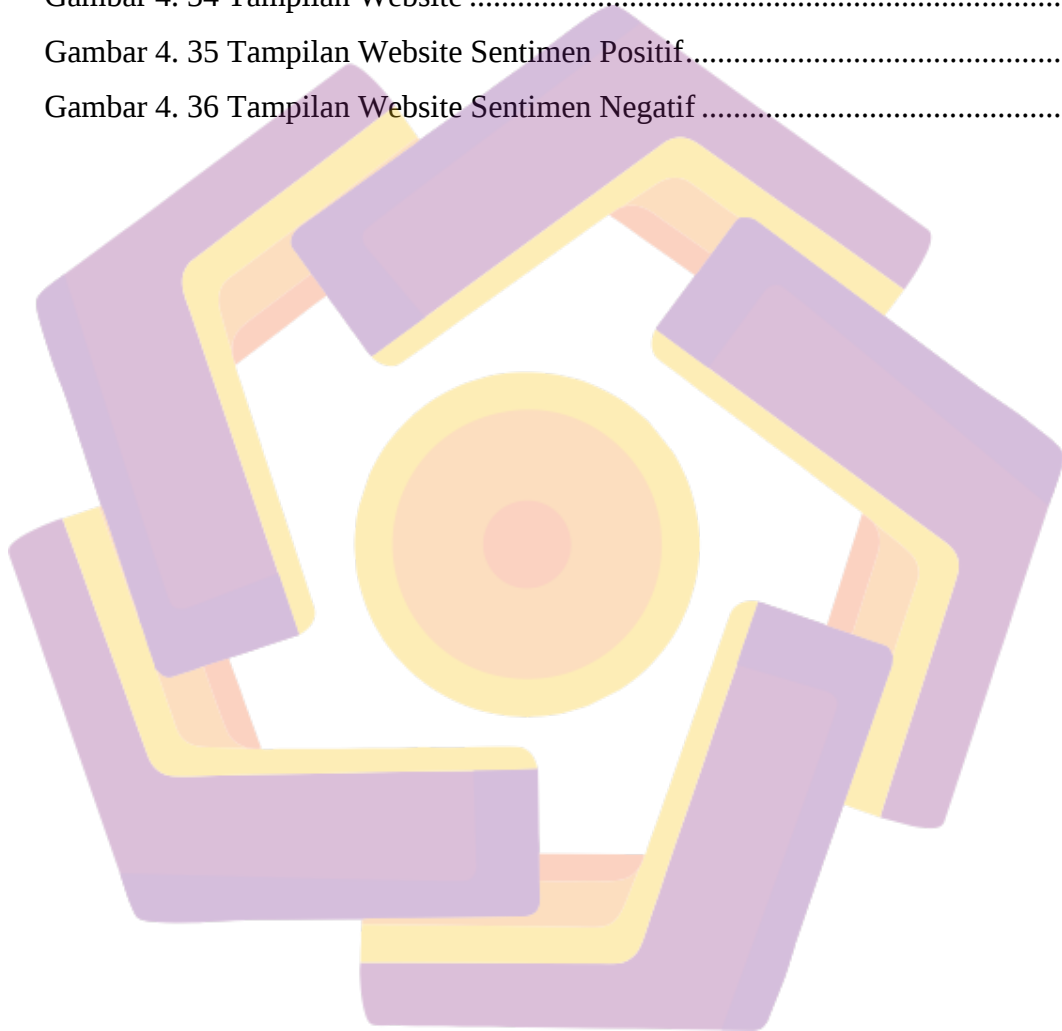
DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 2. 2 Perhitungan Confusion Matrix.....	18
Tabel 4. 1 Implementasi Data Cleaning.....	28
Tabel 4. 2 Implementasi Case Folding	29
Tabel 4. 3 Implementasi Normalisasi	31
Tabel 4. 4 Implementasi Stopword Removal.....	32
Tabel 4. 5 Implementasi Stemming	33
Tabel 4. 6 Implementasi Tokenizing.....	34
Tabel 4. 7 Implementasi Labeling.....	35
Tabel 4. 8 Perhitungan TF-IDF	42
Tabel 4. 9 Perbandingan Model Naive Bayes, Random Forest dan SVM.....	53
Tabel 4. 10 Hasil Perbandingan Penelitian Sebelumnya	54

DAFTAR GAMBAR

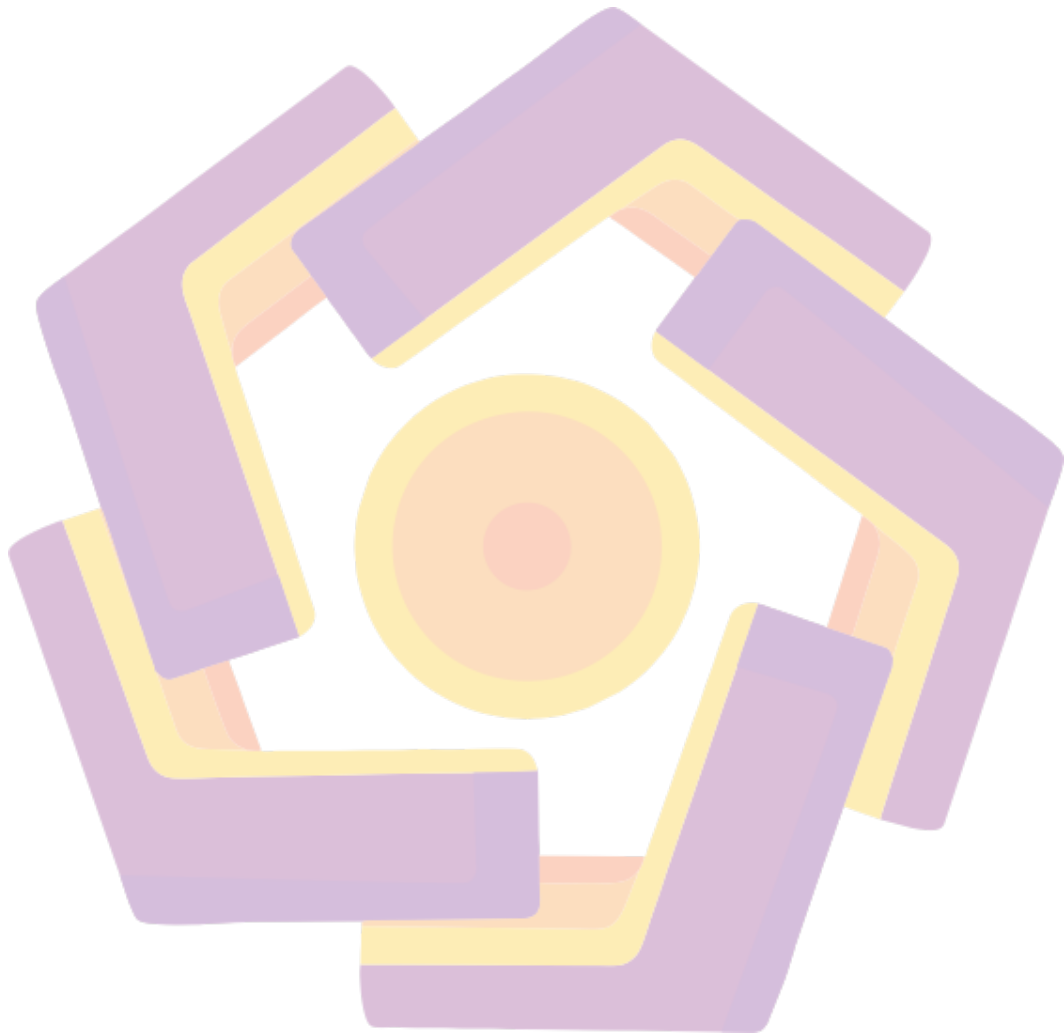
Gambar 2. 1 Hyperplane Support Vector Machine	17
Gambar 4. 1 Source Code Crawling Data.....	27
Gambar 4. 2 Hasil Crawling Data.....	27
Gambar 4. 3 Source Code Data Cleaning	28
Gambar 4. 4 Source Code Case Folding.....	29
Gambar 4. 5 Source Code Normalisasi.....	30
Gambar 4. 6 Source Code Stopword Removal	31
Gambar 4. 7 Source Code Stemming.....	33
Gambar 4. 8 Source Code Tokenizing.....	33
Gambar 4. 9 Source Code Labeling.....	35
Gambar 4. 10 Source Code Pie Chart	36
Gambar 4. 11 Output Pie Chart.....	37
Gambar 4. 12 Source Code Wordcloud	38
Gambar 4. 13 Output Wordcloud Sentimen Positif	39
Gambar 4. 14 Output Wordcloud Sentimen Negatif.....	40
Gambar 4. 15 Splitting Data 10% Data Uji 90% Data Latih	41
Gambar 4. 16 Source Code TF-IDF.....	42
Gambar 4. 17 Data sebelum SMOTE	43
Gambar 4. 18 Data setelah SMOTE	44
Gambar 4. 19 Source Code Naive Bayes.....	44
Gambar 4. 20 Source Code Classification Report Naïve Bayes	45
Gambar 4. 21 Hasil Naive Bayes	45
Gambar 4. 22 Hasil Confusion Matrix Naive Bayes	46
Gambar 4. 23 Source Code Random Forest.....	47
Gambar 4. 24 Source Code Classification Report Random Forest.....	48
Gambar 4. 25 Hasil Random Forest.....	48
Gambar 4. 26 Hasil Confusion Matrix Random Forest	49
Gambar 4. 27 Source Code Support Vector Machine.....	50
Gambar 4. 28 Source Code Classification Report SVM.....	51

Gambar 4. 29 Hasil Support Vector Machine.....	51
Gambar 4. 30 Hasil Confusion Matrix Support Vector Machine	52
Gambar 4. 31 Source Code Pickle	54
Gambar 4. 32 Source Code app.py	55
Gambar 4. 33 Source Code index.html	56
Gambar 4. 34 Tampilan Website	56
Gambar 4. 35 Tampilan Website Sentimen Positif.....	57
Gambar 4. 36 Tampilan Website Sentimen Negatif	58



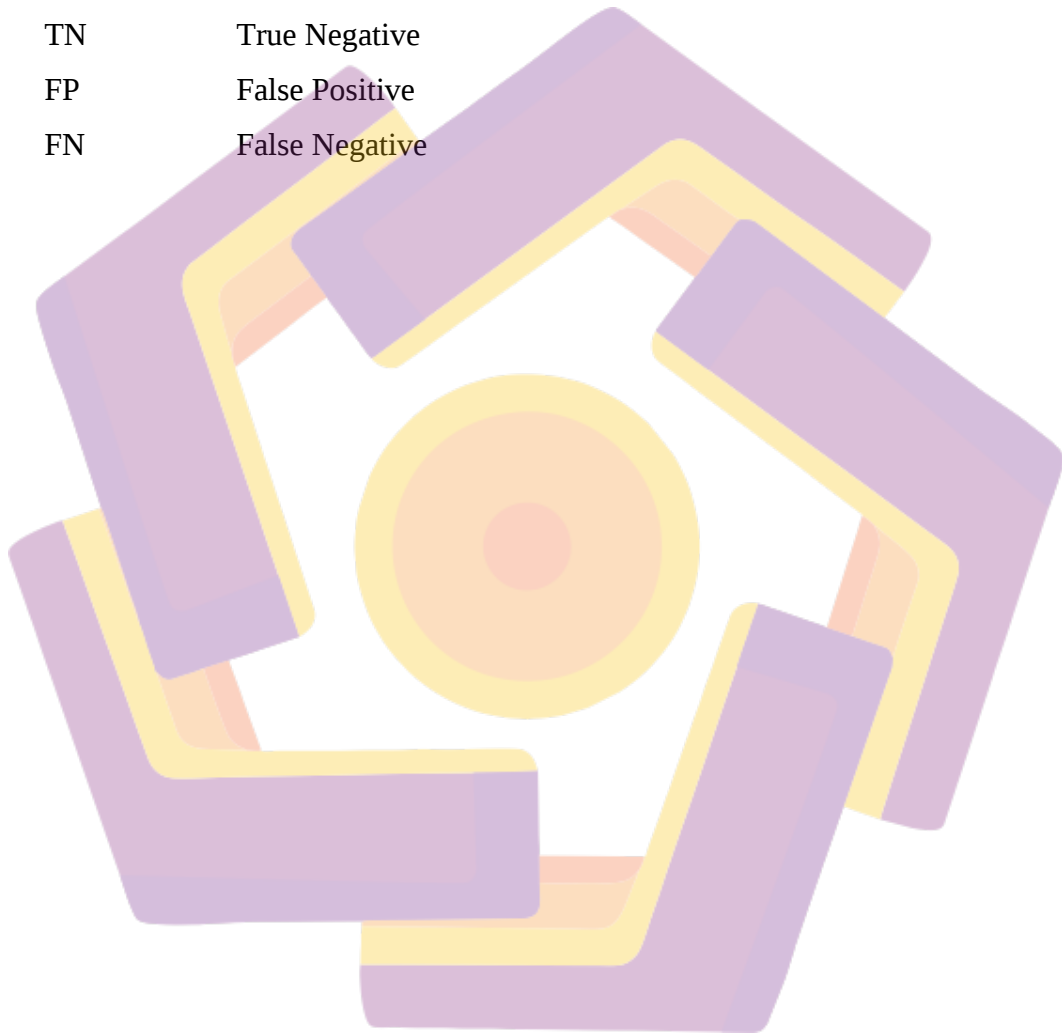
DAFTAR LAMPIRAN

Lampiran 1 Bukti ACC Pendadaran	63
---------------------------------------	----



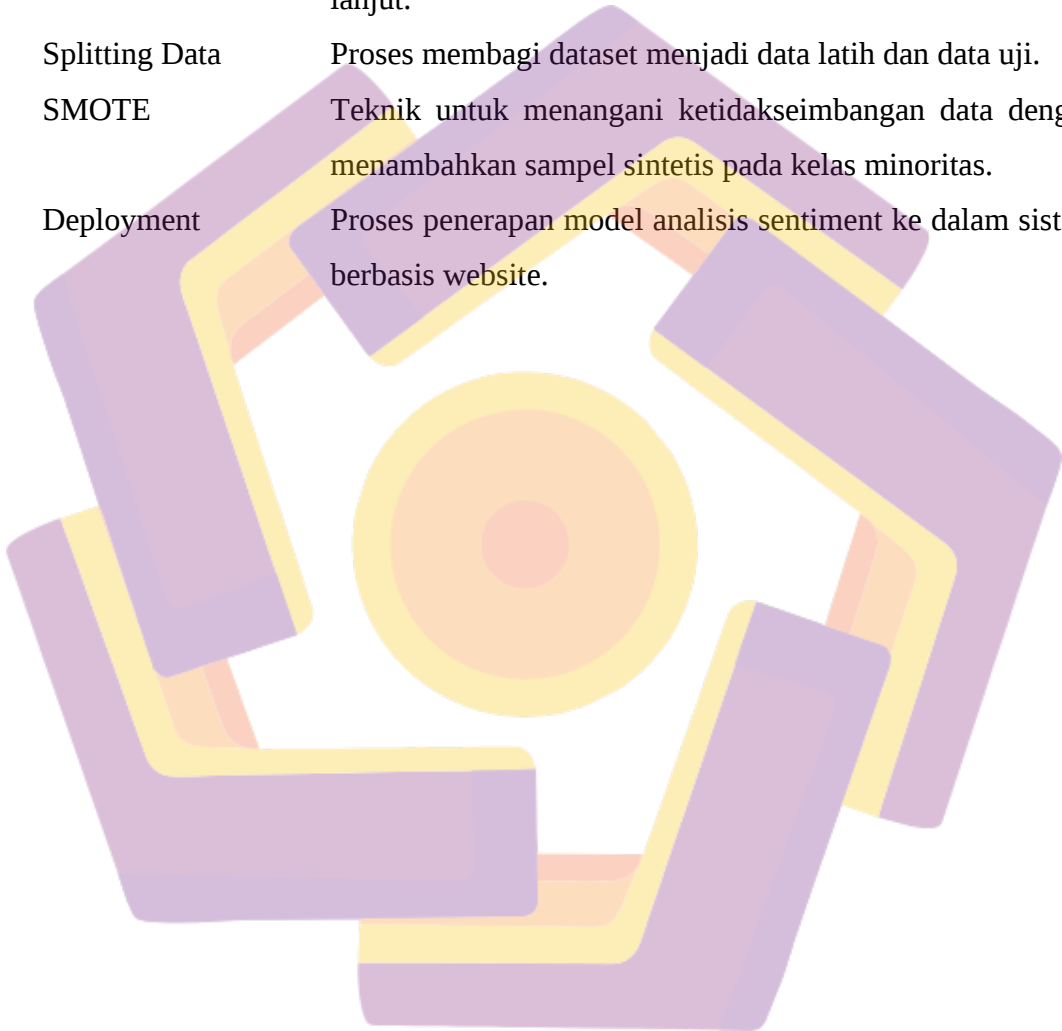
DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
TF-IDF	Term Frequency - Inverse Document Frequency
SMOTE	Synthetic Minority Over-sampling Technique
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative



DAFTAR ISTILAH

Crawling Data	Proses otomatis dalam mengumpulkan data.
Preprocessing Data	Tahapan awal dalam analisis teks yang mencakup pembersihan dan standarisasi data sebelum dianalisis lebih lanjut.
Splitting Data	Proses membagi dataset menjadi data latih dan data uji.
SMOTE	Teknik untuk menangani ketidakseimbangan data dengan menambahkan sampel sintetis pada kelas minoritas.
Deployment	Proses penerapan model analisis sentiment ke dalam sistem berbasis website.



INTISARI

Twitter merupakan media sosial yang berkembang pesat dan sering digunakan sebagai sumber data untuk menganalisis sentimen masyarakat terhadap suatu topik, termasuk kasus kriminal seperti film *Vina: Sebelum 7 Hari* yang mengangkat kisah nyata dan memicu pro dan kontra. Film ini banyak dikritik karena dianggap tidak etis dan mengeksploitasi peristiwa tragis. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap film tersebut dengan menggunakan tiga model algoritma, yaitu Naïve Bayes, Random Forest, dan Support Vector Machine (SVM). Hasil analisis menunjukkan bahwa sentimen negatif mendominasi dengan 677 data, sementara sentimen positif hanya 309 data. Dalam pengujian performa, SVM mencapai akurasi tertinggi sebesar 88%, diikuti oleh Random Forest sebesar 83% dan Naïve Bayes sebesar 76%. Berdasarkan hasil tersebut, model algoritma SVM terbukti sebagai model yang paling efektif dalam menganalisis sentimen pengguna Twitter terhadap film *Vina: Sebelum 7 Hari*.

Kata kunci: Analisis Sentimen, Twitter, Naïve Bayes, Random Forest, Support Vector Machine (SVM)

ABSTRACT

Twitter is a rapidly growing social media platform that is frequently used as a data source for analyzing public sentiment on various topics, including criminal cases such as the film *Vina: Sebelum 7 Hari*, which is based on a true story and has sparked controversy. The film has received criticism for being considered unethical and exploiting a tragic event. Therefore, this study aims to analyze Twitter users' sentiment toward the film using three algorithm models: Naïve Bayes, Random Forest, and Support Vector Machine (SVM). The analysis results indicate that negative sentiment dominates with 677 data points, while positive sentiment only accounts for 309 data points. In performance testing, SVM achieved the highest accuracy of 88%, followed by Random Forest with 83% and Naïve Bayes with 76%. Based on these results, the SVM algorithm model proves to be the most effective in analyzing Twitter users' sentiment toward the film *Vina: Sebelum 7 Hari*.

Keywords: Sentiment Analysis, Twitter, Naïve Bayes, Random Forest, Support Vector Machine (SVM)

