

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
CAPCUT DI PLAY STORE MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi *SI Teknik Komputer*



disusun oleh

Gibran Hait Sami

22.83.0831

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
CAPCUT DI PLAY STORE MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *SI Teknik Komputer*



disusun oleh

Gibran Hait Sami

22.83.0831

Kepada

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

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
CAPCUT DI PLAY STORE MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE (SVM)**

yang disusun dan diajukan oleh

Gibran Hait Sami

22.83.0831

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Desember 2025

Dosen Pembimbing,



Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
CAPCUT DI PLAY STORE MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE (SVM)**

yang disusun dan diajukan oleh

Gibran Hait Sami

22.83.0831

Telah dipertahankan di depan Dewan Penguji
pada tanggal 30 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Muhammad Kopravi, S.Kom., M.Eng.
NIK. 190302454

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Gibran Hait Sami
NIM : 22.83.0831

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI CAPCUT DI PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Dosen Pembimbing: Robert Marco, S.T., M.T., Ph.D

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.

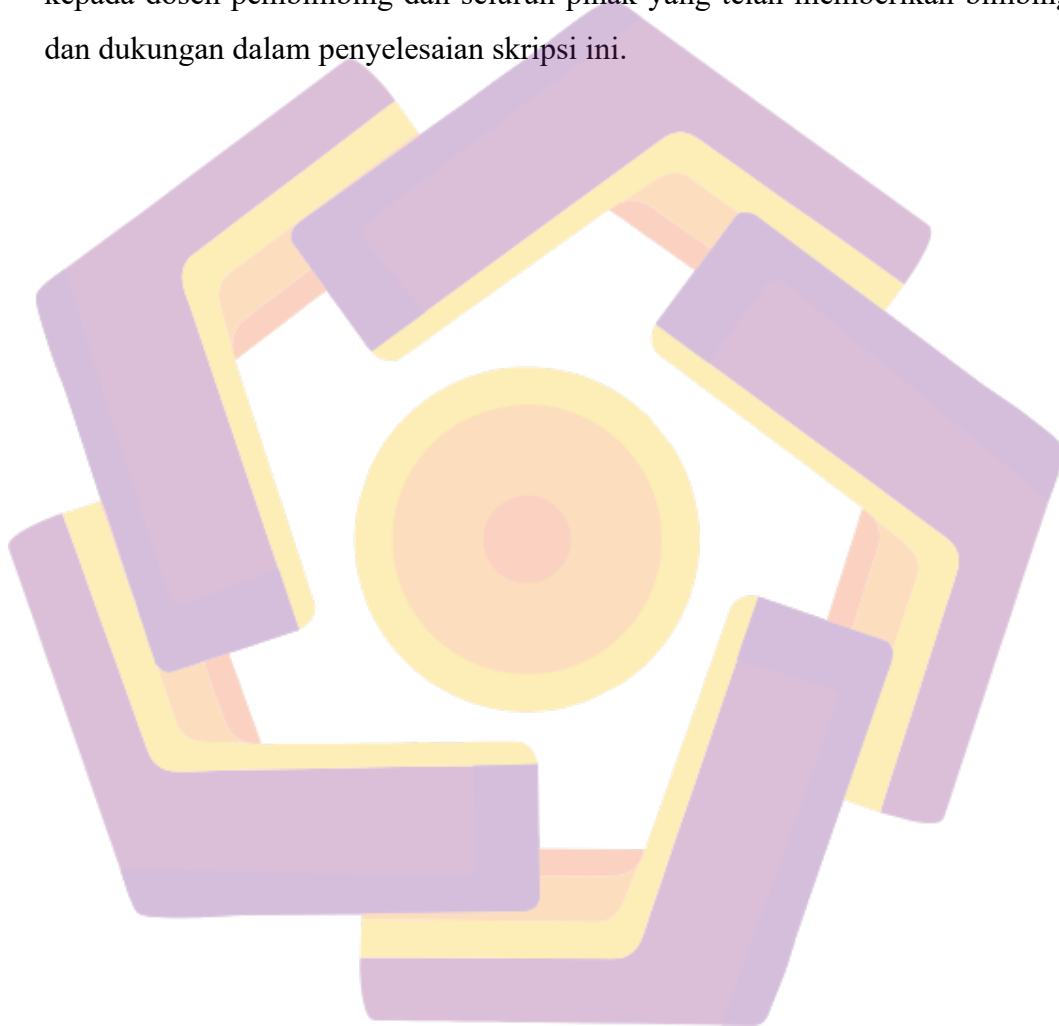
Yogyakarta, 30 Januari 2026

Yang Menyatakan,


Gibran Hait Sami

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya. Skripsi ini juga penulis persembahkan kepada kedua orang tua tercinta atas doa, dukungan, dan pengorbanan yang tiada henti, serta kepada dosen pembimbing dan seluruh pihak yang telah memberikan bimbingan dan dukungan dalam penyelesaian skripsi ini.



KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala limpahan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul Analisis Sentimen Ulasan Aplikasi CapCut di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM). Penyusunan skripsi ini dimaksudkan untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Teknik Komputer di Amikom Yogyakarta.

pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang setulus-tulusnya kepada:

- 1) Bapak selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran dengan penuh kesabaran untuk membimbing,
- 2) Tim Dosen Penguji yang telah memberikan masukan, kritik, dan saran yang membangun.
- 3) Teristimewa untuk kedua orang tua tercinta, Bapak dan Ibu . Terima kasih yang tak terhingga atas segala pengorbanan, kasih sayang, dukungan moral maupun materiil, serta doa yang tiada henti dipanjatkan demi kesuksesan dan kelancaran studi penulis. Kalian adalah motivasi terbesar dalam hidup penulis.
- 4) Teman-teman seperjuangan angkatan serta sahabat terdekat terima kasih atas kebersamaan yang saling menguatkan di kala susah maupun senang selama masa perkuliahan ini.

Penulis menyadari bahwa dalam proses penyusunan skripsi ini banyak mengalami hambatan dan kesulitan. Oleh karena itu kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang.

Yogyakarta, 22 Desember 2025

Penulis

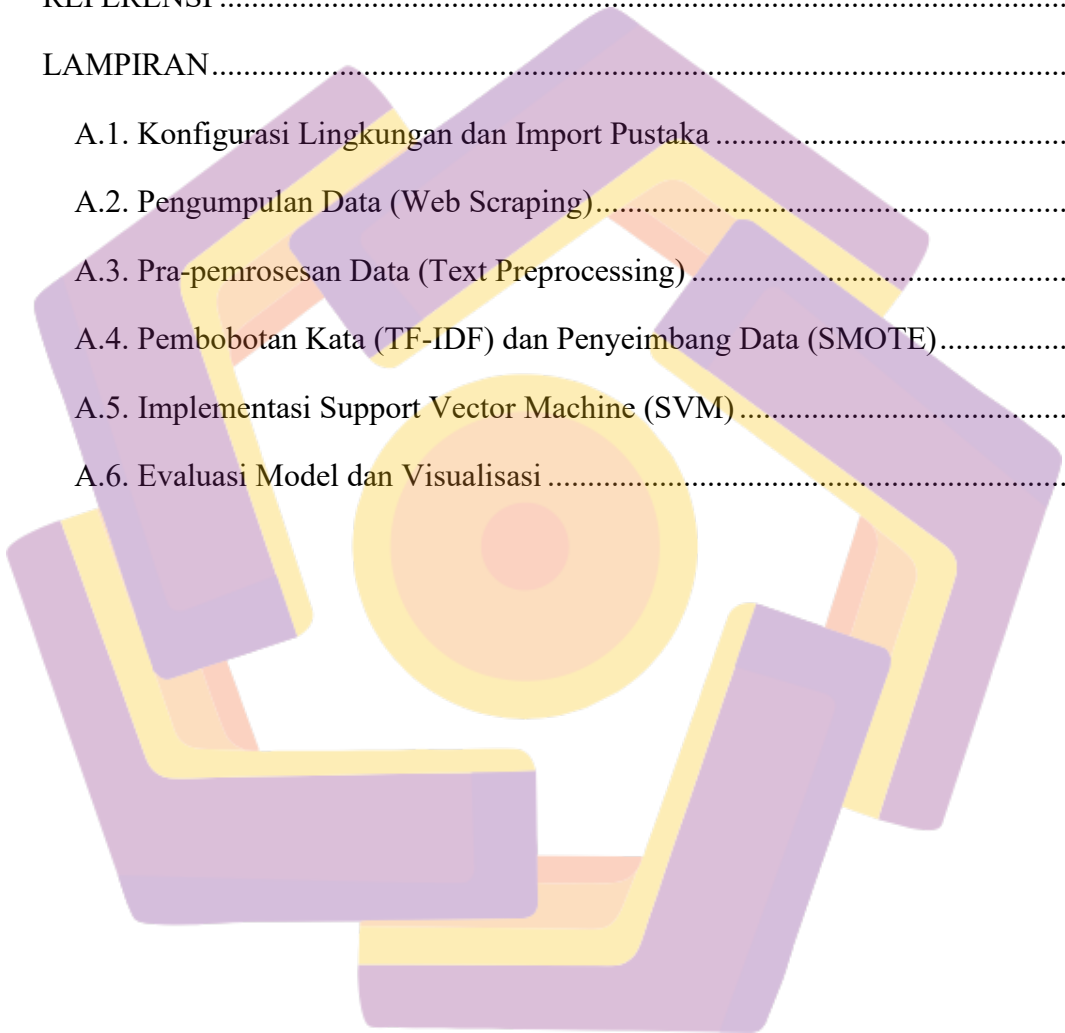
DAFTAR ISI

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.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH.....	xvi
INTISARI	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	6
1.6.1 BAB I PENDAHULUAN.....	6
1.6.2 BAB II TINJAUAN PUSTAKA	6

1.6.3	BAB III METODE PENELITIAN	7
1.6.4	BAB IV HASIL DAN PEMBAHASAN	7
1.6.5	BAB V PENUTUP	7
BAB II TINJAUAN PUSTAKA		8
2.1	Studi Literatur	8
2.2	Dasar Teori.....	15
2.2.1	Natural Language Processing (NLP)	15
2.2.2	Analisis Sentimen	16
2.2.3	Pelabelan Data (Labeling).....	16
2.2.4	Text Preprocessing	17
2.2.5	Pembobotan Kata (TF-IDF).....	17
2.2.6	Split Data	18
2.2.7	Penanganan Ketidakseimbangan Data (SMOTE).....	18
2.2.8	Support Vector Machine (SVM).....	19
2.2.9	Python	20
2.3	Metode Pengujian	20
2.3.1	Akurasi	20
2.3.2	Confusion Matrix	21
BAB III METODE PENELITIAN		23
3.1	Objek Penelitian.....	23
3.2	Alur Penelitian	24
3.2.1	Crawling Data	25
3.2.2	Pelabelan Data	25
3.2.3	Pra-Pemrosesan Data (Preprocessing)	26
3.2.4	Pembobotan Kata	27

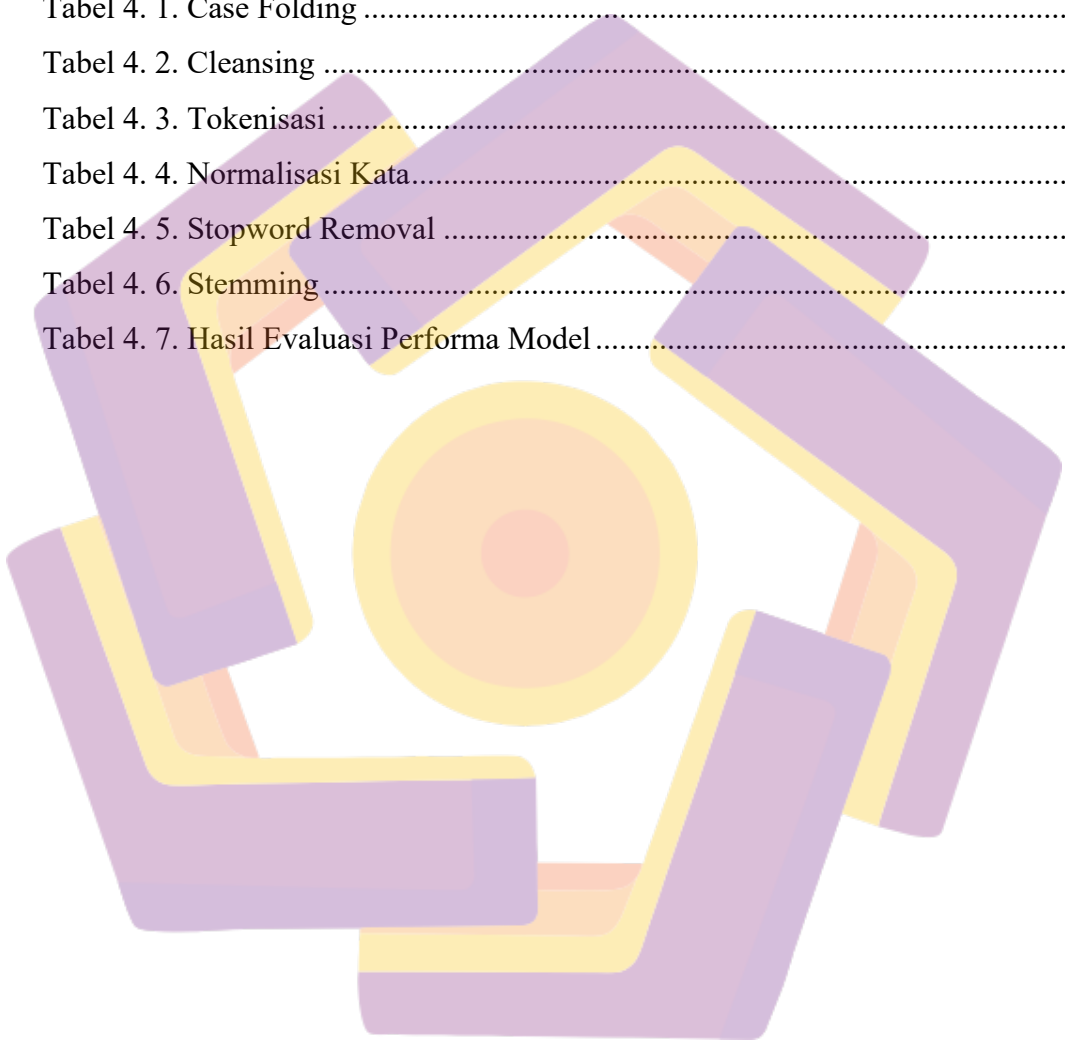
3.2.5	Penanganan Ketidakseimbangan Data (SMOTE)	27
3.2.6	Pembagian Data (Split Data)	28
3.2.7	Klasifikasi Support Vector Machine (SVM)	28
3.2.8	Evaluasi Model	28
3.3	Alat dan Bahan	29
3.3.1	Perangkat Keras	29
3.3.2	Perangkat Lunak	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Persiapan Data	30
4.2	Pelabelan	30
4.3	Pra-pemrosesan (Preprocessing)	31
4.3.1	Case Folding	31
4.3.2	Cleansing	32
4.3.3	Tokenisasi	33
4.3.4	Normalisasi Kata	33
4.3.5	Stoword Removal	34
4.3.6	Stemming	34
4.3.7	Visualisasi WordCloud	35
4.4	Pembobotan TF-IDF	37
4.5	SMOTE	38
4.6	Pembagian Data	39
4.7	Hasil Klasifikasi SVM	40
4.8	Hasil Evaluasi Model	40
4.8.1	Analisis Confusion Matrix	41
4.8.2	Hasil akurasi dan Metrik Evaluasi	42

4.8.3 Pembahasan Hasil	44
BAB V PENUTUP	46
5.1 Kesimpulan	46
5.2 Saran	47
REFERENSI	48
LAMPIRAN.....	53
A.1. Konfigurasi Lingkungan dan Import Pustaka	53
A.2. Pengumpulan Data (Web Scraping).....	54
A.3. Pra-pemrosesan Data (Text Preprocessing)	54
A.4. Pembobotan Kata (TF-IDF) dan Penyeimbang Data (SMOTE).....	55
A.5. Implementasi Support Vector Machine (SVM).....	56
A.6. Evaluasi Model dan Visualisasi	56



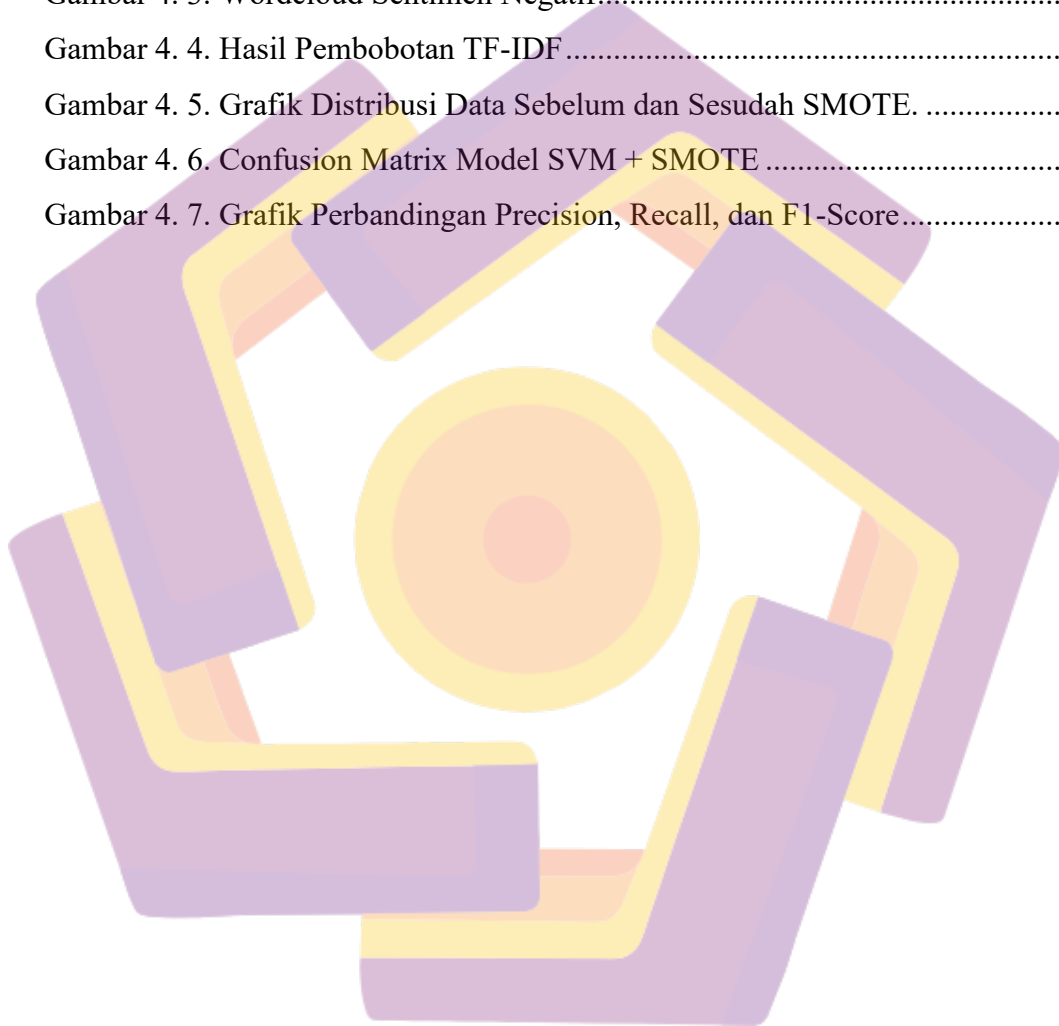
DAFTAR TABEL

Tabel 2. 1. Keaslian Penelitian	10
Tabel 2. 2. confusion matrix kelas 2	21
Tabel 3. 1. Split Data	28
Tabel 4. 1. Case Folding	31
Tabel 4. 2. Cleansing	32
Tabel 4. 3. Tokenisasi	33
Tabel 4. 4. Normalisasi Kata.....	33
Tabel 4. 5. Stopword Removal	34
Tabel 4. 6. Stemming	35
Tabel 4. 7. Hasil Evaluasi Performa Model.....	42



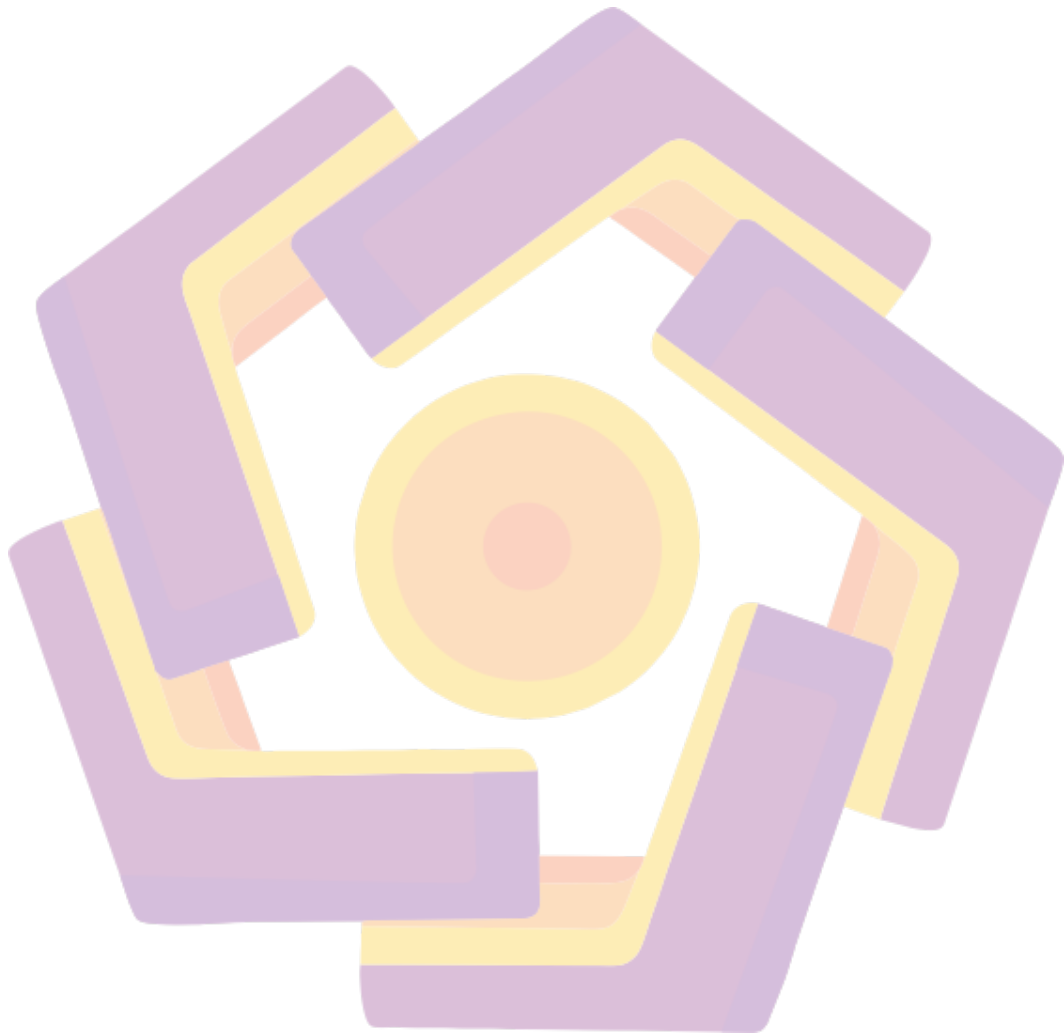
DAFTAR GAMBAR

Gambar 3. 1. Alur Penelitian	24
Gambar 4. 1. Grafik Distribusi Sentimen Sebelum SMOTE.....	31
Gambar 4. 2. Wordcloud Sentimen Positif.....	36
Gambar 4. 3. Wordcloud Sentimen Negatif.....	37
Gambar 4. 4. Hasil Pembobotan TF-IDF	38
Gambar 4. 5. Grafik Distribusi Data Sebelum dan Sesudah SMOTE.	39
Gambar 4. 6. Confusion Matrix Model SVM + SMOTE	41
Gambar 4. 7. Grafik Perbandingan Precision, Recall, dan F1-Score.....	44

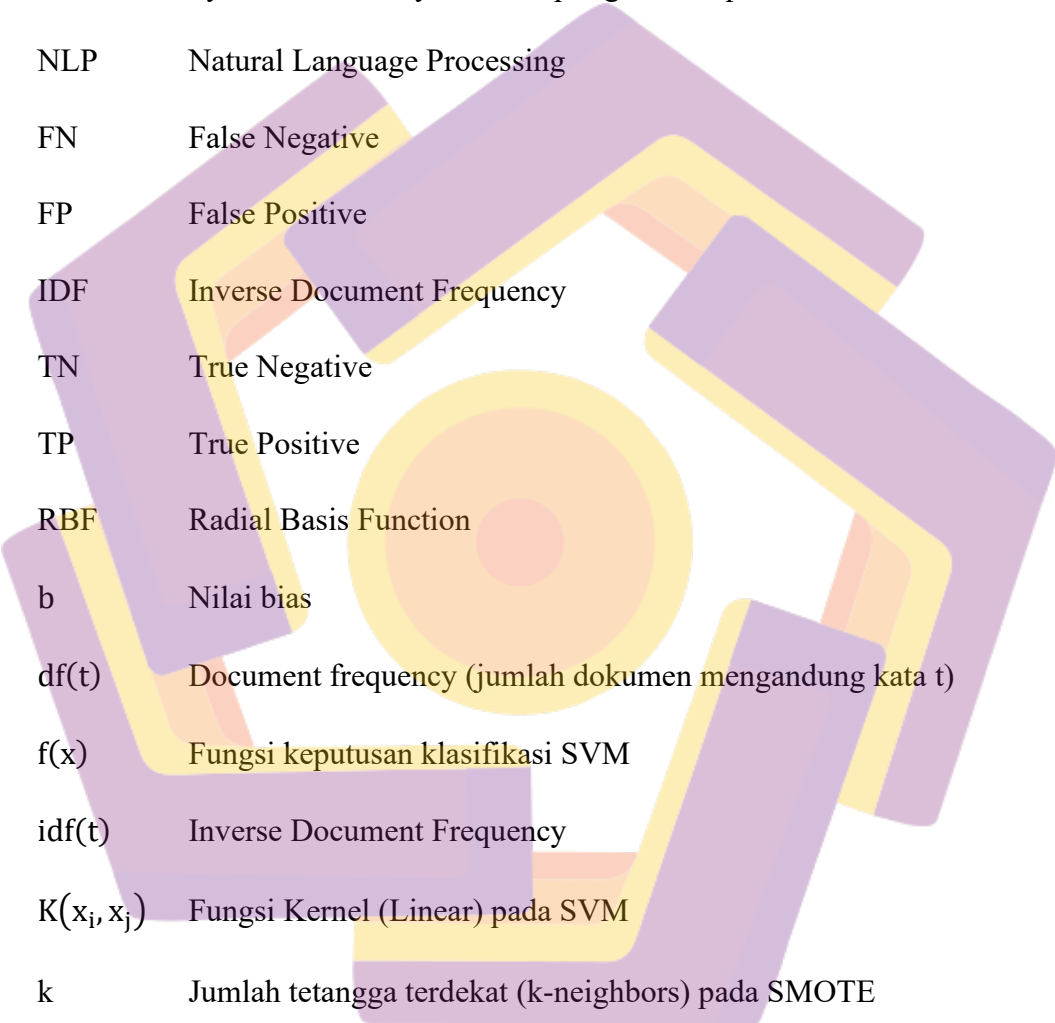


DAFTAR LAMPIRAN

Lampiran A Source Code Penelitian.....	53
--	----

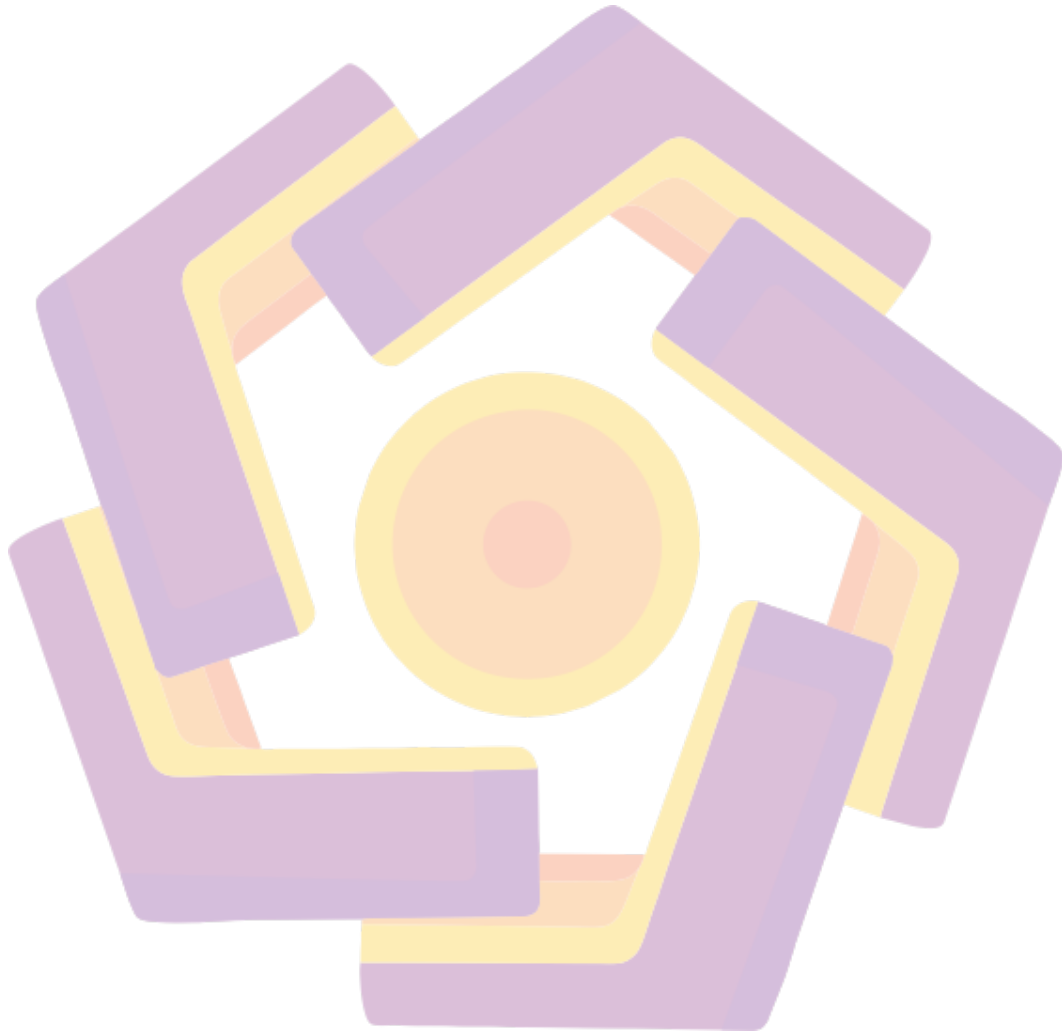


DAFTAR LAMBANG DAN SINGKATAN



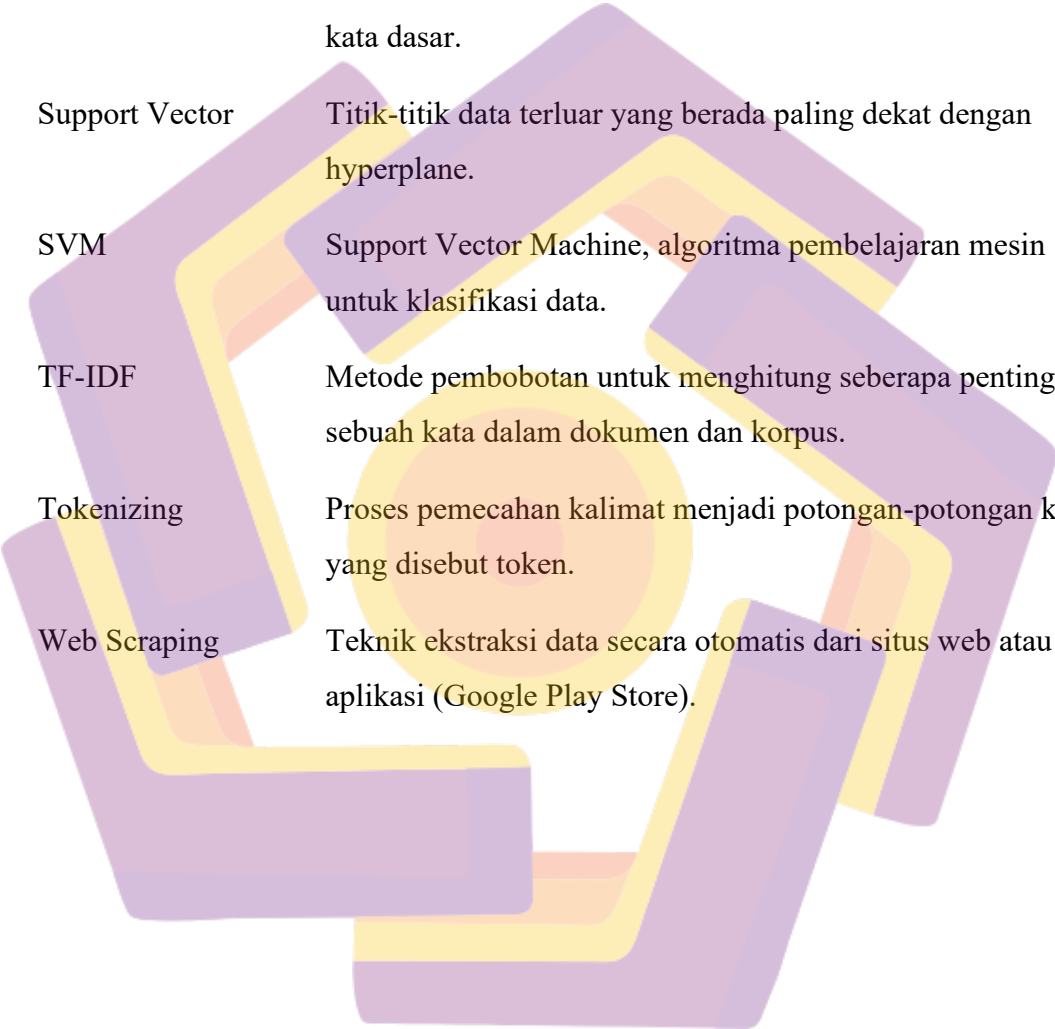
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
SMOTE	Synthetic Minority Oversampling Technique
NLP	Natural Language Processing
FN	False Negative
FP	False Positive
IDF	Inverse Document Frequency
TN	True Negative
TP	True Positive
RBF	Radial Basis Function
b	Nilai bias
$df(t)$	Document frequency (jumlah dokumen mengandung kata t)
$f(x)$	Fungsi keputusan klasifikasi SVM
$idf(t)$	Inverse Document Frequency
$K(x_i, x_j)$	Fungsi Kernel (Linear) pada SVM
k	Jumlah tetangga terdekat (k-neighbors) pada SMOTE
N	Jumlah total dokumen dalam korpus dataset
$tf(t, d)$	Term Frequency (frekuensi kata dalam dokumen)
W	Nilai Bobot (Weight) TF-IDF
w	Vektor bobot (weight vector)
x	Vektor input atau fitur data

x_{new}	Sampel data sintetis baru hasil proses SMOTE
y	Label kelas target (Positif/Negatif)
δ	Simbol Delta (bilangan acak 0-1 pada SMOTE)
Σ	Simbol Sigma (notasi penjumlahan)



DAFTAR ISTILAH

Accuracy	Tingkat kedekatan antara nilai prediksi model dengan nilai aktual (sebenarnya).
Confusion Matrix	Tabel evaluasi yang membandingkan hasil klasifikasi sistem dengan data sebenarnya.
Corpus	Kumpulan seluruh dokumen atau data teks yang digunakan dalam penelitian.
Data Training	Data latih yang digunakan untuk mengajarkan algoritma mengenali pola sentimen.
Data Testing	Data uji yang digunakan untuk mengukur performa model yang telah dilatih.
Hyperplane	Garis atau bidang pemisah optimal yang membagi dua kelas dalam algoritma SVM.
Imbalanced Data	Kondisi di mana jumlah data antar kelas (positif dan negatif) tidak seimbang.
Kernel Linear	Fungsi matematis pada SVM untuk memisahkan data secara garis lurus (linear).
Margin	Jarak antara hyperplane dengan titik data terdekat (support vectors) dari masing-masing kelas.
NLP	Natural Language Processing, cabang AI yang fokus pada pengolahan bahasa manusia.
Precision	Rasio ketepatan antara data yang diprediksi positif dengan keseluruhan data yang benar-benar positif.
Preprocessing	Tahapan awal untuk mempersiapkan, membersihkan, dan mengubah format data mentah.



Recall	Rasio keberhasilan model dalam menemukan kembali informasi yang relevan.
SMOTE	Synthetic Minority Over-sampling Technique, teknik menyeimbangkan data dengan membuat data sintetis baru.
Stemming	Proses mengubah kata yang memiliki imbuhan menjadi kata dasar.
Support Vector	Titik-titik data terluar yang berada paling dekat dengan hyperplane.
SVM	Support Vector Machine, algoritma pembelajaran mesin untuk klasifikasi data.
TF-IDF	Metode pembobotan untuk menghitung seberapa penting sebuah kata dalam dokumen dan korpus.
Tokenizing	Proses pemecahan kalimat menjadi potongan-potongan kata yang disebut token.
Web Scraping	Teknik ekstraksi data secara otomatis dari situs web atau aplikasi (Google Play Store).

INTISARI

Volume ulasan aplikasi CapCut di Google Play Store yang sangat besar dan tidak terstruktur, meliputi komponen seperti emoji serta ragam bahasa informal, menyulitkan pengembang dalam mengekstraksi wawasan secara manual. Selain itu, ketimpangan jumlah data yang signifikan antara ulasan positif dan negatif menimbulkan masalah ketidakseimbangan data (imbalanced data) yang berdampak pada biasanya prediksi model. Penelitian ini bertujuan mengklasifikasikan sentimen ulasan menggunakan algoritma Support Vector Machine (SVM) yang dikombinasikan dengan Synthetic Minority Over-sampling Technique (SMOTE) untuk mengatasi tantangan tersebut. Metode penelitian meliputi pengumpulan data, filtering rating bintang 3, pra-pemrosesan teks menggunakan library Sastrawi untuk menangani pola bahasa, pembobotan fitur TF-IDF, serta penerapan SMOTE untuk menyeimbangkan dataset menjadi 4.800 data sebelum diklasifikasikan dengan SVM Kernel Linear. Hasil akhir pada penelitian menunjukkan bahwa penerapan SMOTE mampu meningkatkan performa model secara signifikan dengan perolehan akurasi sebesar 94% dan recall kelas negatif mencapai 0.98. Hasil penelitian ini dapat dimanfaatkan oleh tim pengembang CapCut untuk melakukan evaluasi perbaikan fitur secara otomatis dan efisien, serta menjadi referensi bagi peneliti lain dalam penanganan data teks tidak seimbang.

Kata kunci: Analisis Sentimen, CapCut, SMOTE, SVM

ABSTRACT

The substantial and unstructured volume of CapCut application reviews on the Google Play Store, encompassing components such as emojis and informal language varieties, challenges developers in manually extracting actionable insights. Additionally, the significant disparity between positive and negative reviews creates an imbalanced data problem, leading to prediction bias in the model. This study aims to classify review sentiment using the Support Vector Machine (SVM) algorithm combined with the Synthetic Minority Over-sampling Technique (SMOTE) to address these challenges. The research methodology involves data collection, filtering out 3-star ratings, text preprocessing using the Sastrawi library to handle language patterns, TF-IDF feature weighting, and applying SMOTE to balance the dataset to 4,800 data points before classification using Linear Kernel SVM. The final results indicate that applying SMOTE significantly improved model performance, achieving an accuracy of 94% and a negative class recall of 0.98. These findings can be utilized by the CapCut development team to conduct automatic and efficient feature improvement evaluations, and serve as a reference for other researchers in handling imbalanced text data.

Keyword: *Sentiment Analysis, CapCut, SMOTE, SVM*