

SKRIPSI

**PENINGKATAN KINERJA *SUPPORT VECTOR MACHINE* (SVM)
UNTUK ANALISIS SENTIMEN PADA APLIKASI *CHATGPT*
MENGUNAKAN *TF-IDF* DAN *SMOTE***



Oleh:

Muhamad Reza Salsabila Azhar

42419031

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PERADABAN**

2025

PERSETUJUAN SKRIPSI

JUDUL : PENINGKATAN KINERJA *SUPPORT VECTOR MACHINE*
(SVM) UNTUK ANALISIS SENTIMEN PADA APLIKASI
CHATGPT MENGGUNAKAN *TF-IDF* DAN *SMOTE*

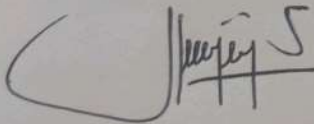
NAMA : Muhamad Reza Salsabila Azhar

NIM : 42419031

Skripsi ini telah diperiksa dan disetujui untuk
diujikan dihadapan Dewan Penguji
Bumiayu, 11 Oktober 2025

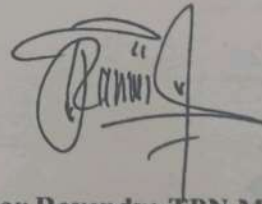
Menyetujui,

Pembimbing I,



Fathulloh, S.T., M.Kom
NIDN. 0623048201

Pembimbing II,



Tezhar Rayendra TPN, M.Kom
NIDN. 0619019201

Mengetahui,

Ketua Program Studi,



Nurul Mega Saraswati, M.Kom
NIDN. 0606069102

PENGESAHAN SKRIPSI

JUDUL : PENINGKATAN KINERJA *SUPPORT VECTOR MACHINE*
(SVM) UNTUK ANALISIS SENTIMEN PADA APLIKASI
CHATGPT MENGGUNAKAN *TF-IDF* DAN *SMOTE*

NAMA : Muhamad Reza Salsabila Azhar

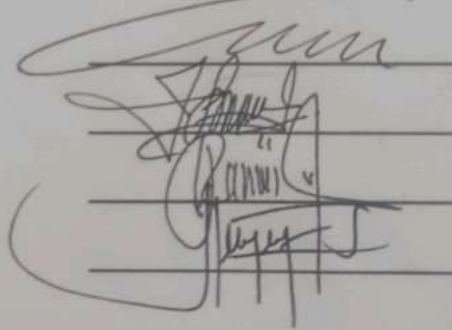
NIM : 42419031

Skripsi ini telah diujikan dan dipertahankan didepan Dewan Penguji Skripsi.
Pada tanggal 26 Oktober 2025. Menurut pandangan kami, Skripsi ini memadai
dari segi kualitas untuk tujuan mencapai gelar Sarjana Komputer (S.Kom)
Bumiayu, 26 Oktober 2025

Nama Penguji

1. Khurotul Aeni, M.Kom
2. Asep Saeful Millah, M.Kom
3. Tezhar Rayendra TPN, M.Kom
4. Fathulloh, S.T., M.Kom

Tanda Tangan



Dekan Fakultas Sains dan Teknologi

Universitas Peradaban,

Bizki Noor Prasetyono, M.Pd.
NIDN. 0611099101

Ketua Program Studi

Informatika,

Nurul Mega Saraswati, M.Kom
NIDN. 0606069102

PERNYATAAN PENULIS

JUDUL : PENINGKATAN KINERJA *SUPPORT VECTOR MACHINE*
(SVM) UNTUK ANALISIS SENTIMEN PADA APLIKASI
CHATGPT MENGGUNAKAN *TF-IDF* DAN *SMOTE*

NAMA : Muhamad Reza Salsabila Azhar

NIM : 42419031

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”

Bumiayu, 26 Oktober 2025

Penulis



Muhamad Reza Salsabila Azhar

ABSTRACT

ChatGPT is an artificial intelligence-based application developed by OpenAI and available on the Google Play Store to assist users in various conversational and information-seeking needs. The large number of user reviews, both positive and negative, makes sentiment analysis essential to understand users' perceptions of the application's performance. In this study, review data were collected using a scraping technique with the Google Play Scraper library, resulting in 10,000 Indonesian language reviews with the application ID "com.openai.chatgpt" during the period of July to August 2025. The collected data then underwent a preprocessing stage to clean irrelevant characters before applying word weighting using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The class distribution was further balanced using the Synthetic Minority Over-sampling Technique (SMOTE) to prevent the model from being biased toward the majority class. The classification process employed the Support Vector Machine (SVM) algorithm, which is known for its effectiveness in handling high-dimensional data. The experimental results show that the combination of TF-IDF and SMOTE significantly improves model performance, achieving an accuracy of 97.81%, precision of 99.12%, recall of 98.50%, F1-score of 98.81%, and AUC of 99.85%. Therefore, this approach is proven to be accurate, stable, and effective for conducting sentiment analysis on ChatGPT user reviews.

Keywords: Sentiment Analysis, ChatGPT, TF-IDF, SMOTE, Support Vector Machine

ABSTRAK

ChatGPT merupakan aplikasi berbasis kecerdasan buatan yang dikembangkan oleh *OpenAI* dan tersedia di *Google Play Store* untuk membantu pengguna dalam berbagai kebutuhan percakapan serta pencarian informasi. Banyaknya ulasan pengguna, baik positif maupun negatif, menjadikan analisis sentimen penting dilakukan untuk mengetahui persepsi pengguna terhadap kinerja aplikasi tersebut. Pada penelitian ini, data ulasan diperoleh melalui teknik *scraping* menggunakan pustaka *Google Play Scraper* sebanyak 10.000 ulasan berbahasa Indonesia dengan *application ID* “com.openai.chatgpt” pada periode Juli hingga Agustus 2025. Data yang terkumpul kemudian diproses melalui tahapan *preprocessing* guna membersihkan teks dari karakter yang tidak relevan sebelum dilakukan pembobotan kata menggunakan metode *Term Frequency Inverse Document Frequency* (TF-IDF). Selanjutnya, dilakukan penyeimbangan distribusi kelas menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) agar model tidak bias terhadap kelas mayoritas. Proses klasifikasi dilakukan dengan algoritma *Support Vector Machine* (SVM) yang dikenal efektif dalam memisahkan data berdimensi tinggi. Hasil pengujian menunjukkan bahwa kombinasi *TF-IDF* dan *SMOTE* mampu meningkatkan kinerja model secara signifikan, ditunjukkan oleh nilai akurasi 97,81%, *precision* 99,12%, *recall* 98,50%, *F1-score* 98,81%, serta *AUC* 99,85%. Dengan demikian, pendekatan ini terbukti akurat, stabil, dan efektif untuk analisis sentimen pengguna aplikasi *ChatGPT*.

Kata kunci: Analisis Sentimen, *ChatGPT*, *TF-IDF*, *SMOTE*, *Support Vector Machine*.

KATA PENGANTAR

Alhamdulillahirobbil'aalamiin, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Peningkatan Kinerja *Support Vector Machine* (SVM) untuk Analisis Sentimen pada Aplikasi *ChatGPT* Menggunakan *TF-IDF* dan *SMOTE*” dengan baik dan lancar.

Skripsi ini merupakan salah satu syarat agar dapat memperoleh gelar Sarjana (S1) di Jurusan Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban Bumiayu. Penyelesaian skripsi ini tidak lepas dari segala bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Dr. M. Kadarisman, S.H, M.Si, selaku Rektor Universitas Peradaban.
2. Rizki Noor Prasetyono, M.Pd. selaku Dekan Fakultas Sains dan Teknologi Universitas Peradaban.
3. Nurul Mega Saraswati, M.Kom. selaku Ketua Jurusan Program Studi Informatika yang telah memberikan kesempatan kepada penulis untuk melaksanakan dan menyelesaikan sarjana di Program Studi Informatika.
4. Fathulloh, S.T., M.Kom. selaku pembimbing pertama yang sudah memberikan bimbingan dan arahan dari awal penulisan proposal sampai terselesaikannya skripsi ini.
5. Tezhar Rayendra TPN, M.Kom. selaku pembimbing kedua yang sudah memberikan bimbingan dan arahan dari awal proposal sampai terselesaikannya skripsi ini.
6. Khurotul Aeni, M.Kom. dan Asep Saful Millah, M.Kom. selaku penguji yang sudah memberikan arahan pada penulisan proposal sampai terselesaikannya skripsi ini.
7. Seluruh Dosen Universitas Peradaban Program Studi Informatika yang sudah memberikan ilmu pengetahuan sehingga penulis mempunyai kemampuan untuk menyelesaikan skripsi ini.
8. Kedua orang tua saya yang selalu mendoakan, memberikan dukungan moral dan material, serta menjadi sumber motivasi terbesar bagi saya.

9. Teman – teman yang telah mendukung saya dan menemani saya selama belajar di Universitas Peradaban.
10. Semua pihak yang membantu saya baik secara moril maupun materiil.

Bumiayu, 26 Oktober 2025

Penulis

Muhamad Reza Salsabila Azhar
42419031

DAFTAR ISI

SKRIPSI.....	i
PERSETUJUAN SKRIPSI.....	ii
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN PENULIS	iv
ABSTRACT.....	v
ABSTRAK.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1. Penelitian Terkait.....	6
2.2. Landasan Teori	10
2.2.1 Analisis Sentimen	10
2.2.2 <i>Text Mining</i>	10
2.2.3 <i>ChatGPT</i>	11

2.2.4	Dataset	11
2.2.5	<i>Algoritma Support Vector Machine (SVM)</i>	11
2.2.6	<i>Term Frequency–Inverse Document Frequency (TF-IDF)</i>	12
2.2.7	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	13
2.2.8	<i>Confusion Matrix</i>	13
2.2.9	<i>Kurva ROC dan AUC</i>	15
2.2.10	<i>Python</i>	15
2.3.	Kerangka Pemikiran.....	16
BAB III METODOLOGI PENELITIAN		20
3.1.	Tahapan Penelitian	20
3.1.1	Identifikasi Masalah.....	21
3.1.2	Studi Literatur	21
3.1.3	Pengumpulan Data	21
3.1.4	Pengolahan Data	21
3.1.5	<i>Split Dataset</i>	23
3.1.6	<i>SMOTE Pada Data Training</i>	23
3.1.7	Ekstraksi Fitur <i>Term Frequency–Inverse Document (TF-IDF)</i>	23
3.1.8	Model Yang Diusulkan (<i>Support Vector Machine (SVM)</i>).....	23
3.1.9	Evaluasi.....	28
3.1.10	Hasil	28
3.2.	Alat dan Bahan	28
BAB IV		30
HASIL DAN PEMBAHASAN.....		30
4.1	Pengumpulan Data	30
4.2	Pengolahan Data.....	30

4.2.1	<i>Preprocessing Data</i>	30
4.2.2	Labeling Sentimen	33
4.2.3	Pembobotan Kata dengan <i>TF-IDF</i>	34
4.2.4	<i>Split Dataset</i>	36
4.2.5	Penerapan <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	37
4.3	Klasifikasi dengan <i>Support Vector Machine (SVM)</i>	38
4.4	Evaluasi Model.....	39
4.4.1	<i>Heatmap Confusion Matrix</i>	40
4.4.2	Perbandingan Kinerja Model Sebelum Dan Sesudah Di <i>SMOTE</i>	41
4.4.3	Perhitungan Performa <i>Best Model SVM</i>	41
4.4.4	<i>ROC Curve</i> dan <i>AUC</i>	42
4.5	<i>Graphical User Interface (GUI)</i>	43
4.6	Hasil Pengujian <i>Graphical User Interface (GUI)</i>	45
BAB V		47
KESIMPULAN DAN SARAN		47
5.1	Kesimpulan	47
5.2	Saran.....	47
DAFTAR PUSTAKA.....		48
LAMPIRAN.....		52

DAFTAR GAMBAR

Gambar 2. 1 Diagram Kerangka Pemikiran.....	18
Gambar 3. 1 Tahapan Penelitian	20
Gambar 3. 2 Ilustrasi <i>SVM Synthetic: Hyperplane Dan Margin</i>	27
Gambar 3. 3 Visualisasi <i>SVM</i> pada <i>TruncatedSVD</i> (10.000 Ulasan <i>TF-IDF</i>)	27
Gambar 4. 1 Distribusi Ulasan Sebelum dan Sesudah <i>Preprocessing</i>	33
Gambar 4. 2 Visualisasi <i>confusion matrix</i>	40
Gambar 4. 3 <i>ROC Curve SVM</i>	43
Gambar 4. 4 Tampilan antarmuka <i>GUI</i>	44

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	8
Tabel 4. 1 Contoh Data Hasil <i>Scraping</i>	30
Tabel 4. 2 Hasil pembersihan teks	31
Tabel 4. 3 Hasil Normalisasi Kata + <i>Fuzzy Matching</i>	31
Tabel 4. 4 Hasil <i>Stopword Removal</i>	31
Tabel 4. 5 Hasil tahap pembersihan lanjutan	32
Tabel 4. 6 Hasil <i>Stemming</i>	32
Tabel 4. 7 Contoh Ulasan yang Terhapus Selama Tahap <i>Preprocessing</i>	32
Tabel 4. 8 Contoh Hasil Labeling	34
Tabel 4. 9 Contoh Hasil pembobotan <i>TF-IDF</i>	35
Tabel 4. 10 Jumlah Data Setelah <i>Split Dataset</i>	36
Tabel 4. 11 Hasil <i>SMOTE</i>	37
Tabel 4. 12 <i>Confusion Matrix</i> Model <i>SVM</i>	40
Tabel 4. 13 perbandingan kinerja model sebelum dan sesudah <i>SMOTE</i>	41
Tabel 4. 14 Perhitungan Performa <i>Best Model SVM</i>	42

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital yang pesat telah mendorong lahirnya berbagai inovasi berbasis kecerdasan buatan *Artificial Intelligence* (AI), khususnya dalam bidang pemrosesan bahasa alami *Natural Language Processing* (NLP). Salah satu inovasi yang menonjol adalah *ChatGPT*, sebuah model bahasa canggih yang dikembangkan oleh *OpenAI*. *ChatGPT* mampu melakukan interaksi percakapan berbasis teks secara cerdas dan responsif, sehingga dimanfaatkan secara luas dalam berbagai bidang seperti penulisan, pencarian informasi, pembelajaran, hingga pembuatan konten digital[1].

Sebagai salah satu aplikasi *NLP* yang populer, *ChatGPT* digunakan oleh berbagai kalangan, mulai dari pelajar, pekerja profesional, hingga masyarakat umum. Popularitas tersebut tercermin dari banyaknya ulasan pengguna di *Google PlayStore* yang tidak hanya berupa *rating* bintang, tetapi juga opini dalam bentuk teks. Ulasan ini mencerminkan pengalaman, kepuasan, maupun keluhan pengguna terhadap aplikasi[2]. Sayangnya, informasi berharga dalam ulasan teks belum dimanfaatkan secara optimal. Padahal, menurut Y. Perwira Bima Sakti dkk[3], analisis mendalam terhadap opini pengguna tidak hanya mampu mengungkap kecenderungan sentimen, tetapi juga dapat mengidentifikasi isu-isu spesifik yang sering muncul, memahami pola persepsi masyarakat, serta memberikan dasar yang lebih kuat bagi pengambilan keputusan.

Analisis sentimen hadir sebagai pendekatan yang relevan untuk mengolah ulasan tersebut secara sistematis. Dengan analisis sentimen, opini pengguna dapat diklasifikasikan ke dalam kategori positif maupun negatif sehingga menghasilkan informasi yang lebih bermakna. Namun, penggunaan algoritma *SVM* dalam analisis sentimen memiliki keterbatasan, khususnya ketika menghadapi distribusi data yang tidak seimbang. Kondisi ini dapat menurunkan performa model karena *SVM* cenderung bias terhadap kelas mayoritas[4]. Tantangan ini semakin kompleks jika dikaitkan dengan volume ulasan yang besar dan variasi bahasa.

Untuk mengatasi permasalahan tersebut, diperlukan teknik pendukung seperti *Term Frequency–Inverse Document Frequency* (TF-IDF) yang menghasilkan representasi fitur lebih informatif serta *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan distribusi kelas[5]. Kombinasi keduanya diharapkan mampu meningkatkan performa klasifikasi sentimen.

Untuk memperkuat landasan penelitian ini, dilakukan kajian terhadap beberapa penelitian terdahulu yang membahas analisis sentimen pada ulasan aplikasi menggunakan algoritma *Support Vector Machine* (SVM). Kajian tersebut membantu dalam memahami karakteristik data, pendekatan klasifikasi, serta performa algoritma *SVM* dalam konteks serupa.

Peneliti akan mengumpulkan artikel-artikel ilmiah atau jurnal dengan tema yang sama, yaitu penelitian dengan judul “Pendekatan *TF-IDF*, *SMOTE*, dan *SVM* dalam Klasifikasi Sentimen Masyarakat terhadap Pemblokiran Judi Online.” Dalam penelitian ini, dilakukan analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) dengan pembobotan *TF-IDF* dan penyeimbangan data menggunakan *SMOTE* terhadap komentar masyarakat di YouTube. Hasil evaluasi menunjukkan bahwa model *SVM* mencapai akurasi sebesar 61,84% dengan nilai *F1-score* 0,7590. Pada kelas positif, diperoleh *precision* 62%, *recall* 98%, dan *F1-score* 76%, sedangkan pada kelas negatif hanya mencapai *precision* 50%, *recall* 2%, dan *F1-score* 4%[6].

Sedangkan pada penelitian dengan judul “Klasifikasi Sentimen X (Twitter) Perihal Pemindehan Ibu Kota Indonesia Menggunakan *Ekstraksi Fitur TF-IDF* dan *Metode Support Vector Machine* (SVM).” Dalam penelitian ini, dilakukan analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) dengan pembobotan *TF-IDF* terhadap 1.027 data *tweet* yang membahas pemindehan Ibu Kota Negara ke Kalimantan Timur. Hasil evaluasi menunjukkan bahwa model *SVM* mencapai akurasi 73%. Pada kelas negatif diperoleh *precision* 71%, *recall* 70%, dan *F1-score* 70%; pada kelas netral diperoleh *precision* 67%, *recall* 64%, dan *F1-score* 65%; sedangkan pada kelas positif diperoleh *precision* 77%, *recall* 83%, dan *F1-score* 80%[7].

Serta pada penelitian yang berjudul “Analisis Sentimen Masyarakat terhadap Film Horor Indonesia Menggunakan Metode *SVM* dan *TF-IDF*” dilakukan analisis sentimen terhadap ulasan masyarakat di Twitter mengenai film horor Indonesia dengan menggunakan pembobotan kata *TF-IDF* dan algoritma *Support Vector Machine* (*SVM*). Hasil evaluasi menunjukkan bahwa model menghasilkan akurasi sebesar 91%. Pada kelas positif diperoleh *precision* 94%, *recall* 87%, dan *F1-score* 90%, sedangkan pada kelas negatif dicapai *precision* 89%, *recall* 95%, dan *F1-score* 92%. Penelitian ini memperlihatkan bahwa metode yang digunakan mampu memberikan performa yang seimbang pada kedua kelas dengan tingkat akurasi yang tinggi[8].

Berdasarkan kajian tersebut, penelitian ini difokuskan pada upaya peningkatan kinerja *SVM* melalui penerapan *TF-IDF* dan *SMOTE* dalam analisis sentimen terhadap ulasan pengguna aplikasi *ChatGPT* di *Google PlayStore*. *TF-IDF* digunakan sebagai metode pembobotan kata untuk merepresentasikan teks ke dalam bentuk numerik sehingga fitur yang lebih informatif dapat diperoleh, sedangkan *SMOTE* berfungsi menyeimbangkan distribusi data dengan cara menghasilkan sampel sintetis pada kelas minoritas. Dengan kombinasi ini, diharapkan performa klasifikasi sentimen dapat meningkat secara signifikan.

Penelitian ini bertujuan untuk menggambarkan persepsi masyarakat terhadap aplikasi *ChatGPT* sekaligus menguji sejauh mana penerapan *TF-IDF* dan *SMOTE* dapat meningkatkan akurasi klasifikasi menggunakan *SVM*. Hingga tahun 2025, *ChatGPT* telah diunduh lebih dari 500 juta kali di *Google PlayStore* dengan estimasi sekitar 70 juta pengguna di Indonesia, sehingga analisis sentimen terhadap ulasan pengguna menjadi semakin relevan untuk dilakukan.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah “Bagaimana akurasi algoritma *Support Vector Machine* (*SVM*) dengan menerapkan *Term Frequency–Inverse Document Frequency* (*TF-IDF*) dan *Synthetic Minority Over-sampling Technique* (*SMOTE*) untuk klasifikasi sentimen pada ulasan aplikasi *ChatGPT*?”

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengetahui akurasi algoritma *Support Vector Machine* (SVM) dengan menerapkan *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *Synthetic Minority Over-sampling Technique* (SMOTE)

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Bagi peneliti:
 - a. Memberikan pengalaman langsung dalam menerapkan algoritma *SVM* dengan dukungan metode *TF-IDF* dan *SMOTE*, pada analisis data teks.
 - b. Memberikan pemahaman yang lebih mendalam mengenai cara menganalisis *ChatGPT* berdasarkan ulasan pengguna, sehingga dapat menggambarkan persepsi masyarakat terhadap teknologi tersebut.
2. Bagi pengguna aplikasi *ChatGPT*:
 - a. Memberikan informasi mengenai persepsi atau sentimen pengguna terhadap aplikasi *ChatGPT* berdasarkan ulasan yang mereka tinggalkan di *Google PlayStore*.
 - b. Membantu calon pengguna baru dalam mempertimbangkan kelebihan dan kekurangan aplikasi sebelum mengunduh atau menggunakannya.

Hasil dari penelitian ini dapat dijadikan referensi dalam pengembangan studi lanjutan yang berkaitan dengan *Sentiment Analysis*, serta memberikan kontribusi dalam pengembangan model atau sistem evaluasi opini publik terhadap aplikasi berbasis kecerdasan buatan seperti *ChatGPT*.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan ulasan pengguna aplikasi *ChatGPT* yang diambil dari *Google PlayStore*. Total dataset 10.000, dari bulan juli 2025 sampai agustus 2025.
2. Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dengan penerapan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk pembobotan fitur dan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk penyeimbangan data.