

**PENGARUH TEKNIK RESAMPLING (OVERSAMPLING DAN
HYBRID SAMPLING) TERHADAP AKURASI MODEL SVM
DALAM ANALISIS SENTIMEN ULASAN GRAB DI
GOOGLE PLAY STORE
SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
EGIK TRI HATMOKO
20.11.3655

Kepada

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

**PENGARUH TEKNIK RESAMPLING (OVERSAMPLING DAN
HYBRID SAMPLING) TERHADAP AKURASI MODEL SVM
DALAM ANALISIS SENTIMEN ULASAN GRAB DI
GOOGLE PLAY STORE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

EGIK TRI HATMOKO

20.11.3655

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PENGARUH TEKNIK RESAMPLING (OVERSAMPLING DAN
HYBRID SAMPLING) TERHADAP AKURASI MODEL SVM
DALAM ANALISIS SENTIMEN ULASAN GRAB DI
GOOGLE PLAY STORE**

yang disusun dan diajukan oleh

EGIK TRI HATMOKO

20.11.3655

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Agustus 2025

Dosen Pembimbing,


Mulia Sulistyono, S.Kom. M.Kom.
NIK. 190302248

HALAMAN PENGESAHAN

SKRIPSI

**PENGARUH TEKNIK RESAMPLING (OVERSAMPLING DAN
HYBRID SAMPLING) TERHADAP AKURASI MODEL SVM
DALAM ANALISIS SENTIMEN ULASAN GRAB DI
GOOGLE PLAY STORE**

yang disusun dan diajukan oleh

EGIK TRI HATMOKO

20.11.3655

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Anna Baita, S.Kom., M.Kom.
NIK. 190302290

Rumini, S.Kom., M.Kom.
NIK. 190302246

Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
25 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Egik Tri Hatmoko
NIM : 20.11.3655

Menyatakan bahwa Skripsi dengan judul berikut:

Pengaruh teknik resampling (oversampling dan hybrid sampling) terhadap akurasi model svm dalam analisis sentimen ulasan grab di google play store

Dosen Pembimbing : Mulia Sulistiyono, S.Kom. M.Kom.

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, 25 Agustus 2025

Yang Menyatakan,



Egik Tri Hatmoko

HALAMAN PERSEMBAHAN

Pertama-tama, saya panjatkan puji syukur kepada Allah SWT yang telah memberikan rahmat, taufik, dan hidayah-Nya sehingga saya dapat menyelesaikan skripsi ini. Dengan segenap ketulusan dan kerendahan hati, penulis dapat sampai pada tahap ini yang merupakan sebuah pencapaian yang tentunya tidak diraih dengan mudah. Berbagai tantangan telah dilalui, namun berkat niat yang tulus, bimbingan, dan doa-doa yang teriring, serta dukungan dari orang-orang tercinta, skripsi ini dapat terselesaikan dengan baik. Skripsi ini penulis persembahkan dengan penuh cinta dan rasa hormat kepada:

1. Ibunda Saya yang telah meninggal dunia Orang tua adalah sumber kekuatan penulis untuk terus maju dan menyelesaikan setiap tantangan.
2. Ayahanda dan kakak saya tercinta yang selalu mendoakan dan memberi semangat. Kehadiran dan doa dari kalian memberi penulis kekuatan dalam menyelesaikan perjalanan ini. Segala pencapaian ini tak akan terwujud tanpa kehadiran dan cinta kalian.
3. Dosen pembimbing yang telah dengan sabar, bijaksana, dan penuh dedikasi membimbing saya selama proses penyusunan skripsi ini. Terima kasih atas setiap waktu, tenaga, dan perhatian yang Ibu berikan.
4. Teman – teman seperjuangan yang telah berbagi pengetahuan, pengalaman satu sama lain selama perjalanan akademik kami di perguruan tinggi.
5. Almamater tercinta, semoga karya sederhana ini menjadi sumbangsih kecil yang bermanfaat, bagi perkembangan ilmu pengetahuan dan kemajuan almamater tercinta, dan dapat memberi manfaat yang lebih luas.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, petunjuk, dan karunia-Nya yang telah membantu penulis menyelesaikan skripsi dengan judul **“Pengaruh teknik resampling (oversampling dan hybrid sampling) terhadap akurasi model svm dalam analisis sentimen ulasan grab di google play store.”** Skripsi ini disusun untuk memenuhi persyaratan kelulusan

Program Studi S1 Informatika pada Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari doa, bimbingan, dukungan, dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Keluarga tersayang, yang senantiasa memberikan doa terbaik, dukungan moral, materi, serta semangat dalam perjalanan penyusunan skripsi ini.
2. Prof. Dr. M. Suyanto., M.M. selaku Rektor Universitas Amikom Yogyakarta
3. Mulia Sulistiyono, S.Kom., M.Kom selaku dosen pembimbing yang telah memberikan bimbingan dan berbagai pengalaman kepada penulis.
4. Segenap Dosen Program Studi Informatika yang telah mendidik serta memberikan ilmu selama kuliah dan seluruh staff yang selalu sabar melayani segala administrasi.
5. Sahabat dan teman, yang telah berbagi ilmu, pengalaman, dan inspirasi sepanjang perjalanan akademik di Universitas Amikom Yogyakarta.
6. Semua individu yang turut menyumbangkan gagasan, pandangan, dan sokongan dalam berbagai bentuk, yang telah membantu kelancaran perjalanan penyelesaian skripsi ini.

Yogyakarta, 25 Agustus 2025

Egik Tri Hatmoko

DAFTAR ISI

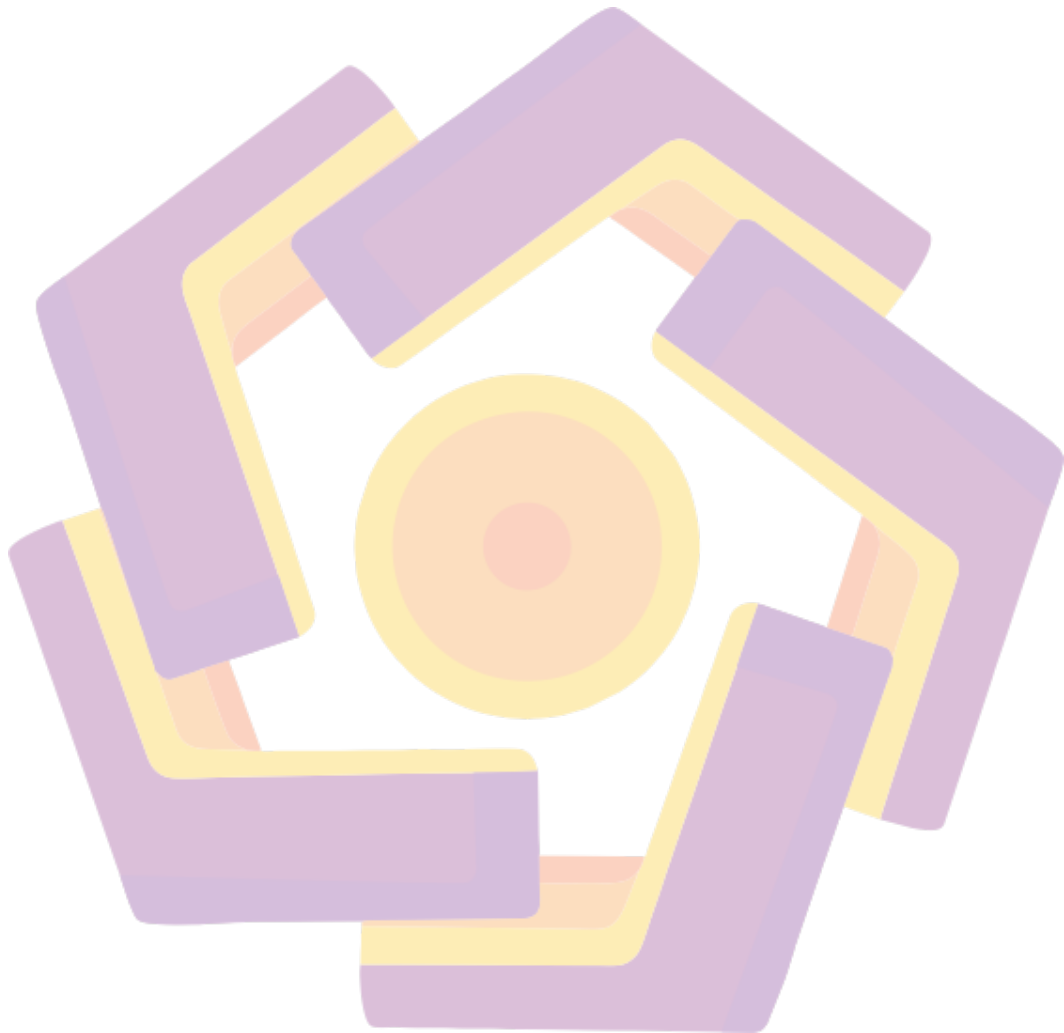
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH.....	xvi
INTISARI	xix
<i>ABSTRACT</i>	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	7

2.1	Studi Literatur	7
2.2	Dasar Teori.....	24
2.2.1	Analisis Sentimen	24
2.2.2	Google Play Store	24
2.2.3	Web Scrapping.....	24
2.2.4	Text Pre-processing	25
2.2.4.1	Cleaning	25
2.2.4.2	Case Folding	25
2.2.4.3	Tokenization	25
2.2.4.4	Stop Removal.....	26
2.2.4.5	Stemming	26
2.2.5	Labeling Data.....	26
2.2.6	Handling Imbalanced Data (Resampling).....	27
2.2.6.1	SMOTE	27
2.2.6.2	ADASYN	27
2.2.6.3	Random Oversampling	28
2.2.6.4	SMOTEENN	28
2.2.7	TF-IDF	29
2.2.8	Support Vector Machine	29
2.2.9	Cross Validation.....	31
2.2.9.1	K-Fold Cross-Validation Standar	31
2.2.9.2	Leave-One-Out Cross-Validation (LOOCV).....	32

2.2.9.3 Stratified Cross-Validation	32
2.2.10 Evaluasi Kinerja Model	32
2.2.10.1 Akurasi	33
2.2.10.2 Precision (Presisi)	33
2.2.10.3 Recall	33
2.2.10.4 F1-Score	33
BAB III METODE PENELITIAN	34
3.1 Alur Penelitian	34
3.1.1 Studi Literatur	35
3.1.2 Pengumpulan Data	35
3.1.3 Data Preprocessing	36
3.1.3.1 Data Cleaning	36
3.1.3.2 Remove Stopword	37
3.1.3.3 Tokenization	37
3.1.3.4 Stemming	38
3.1.4 Labeling	39
3.1.5 Visualisasi	39
3.1.6 Pembagian Data	40
3.1.7 Feature Extraction	41
3.1.8 Resampling Data	43
3.1.9 Modeling dan Evaluasi	43
3.2 Alat dan Bahan	45

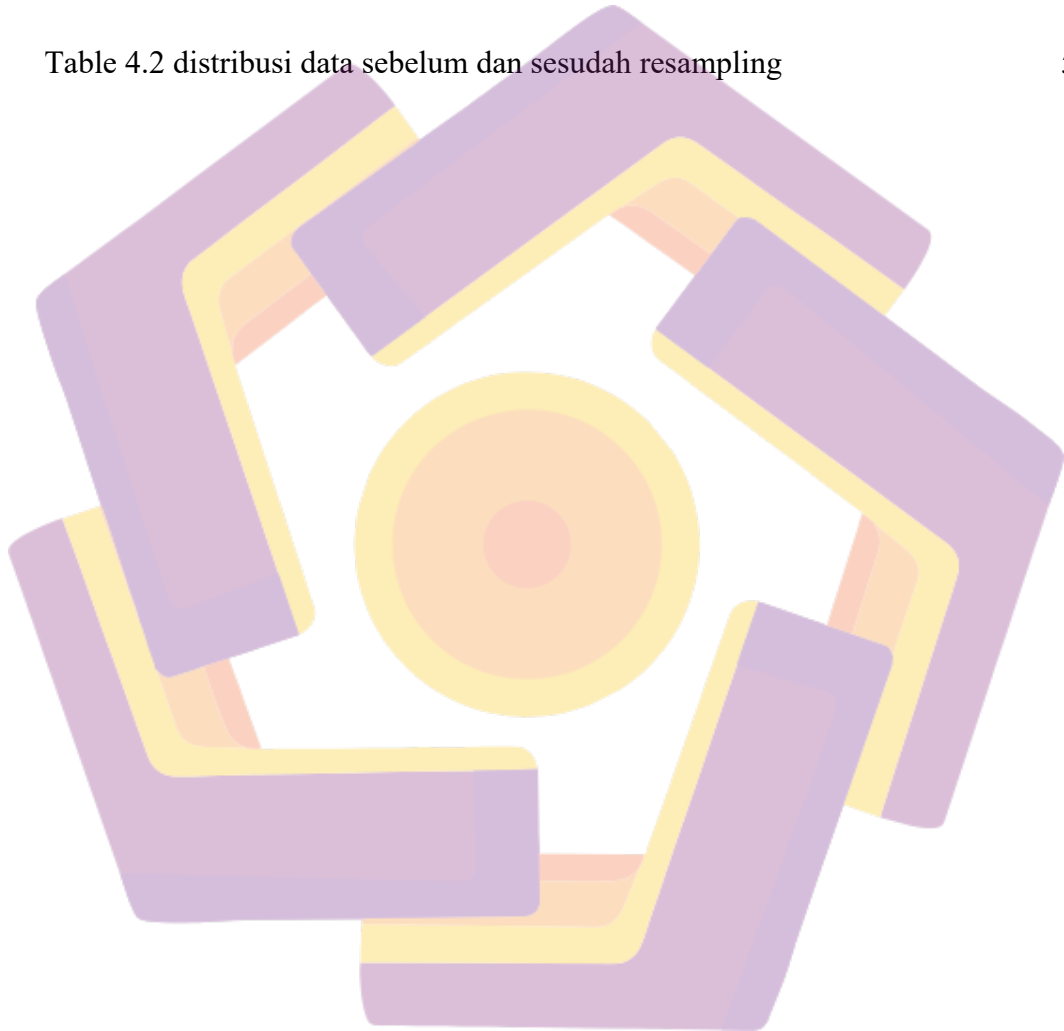
3.2.1	Data	45
3.2.2	Perangkat Keras	45
3.2.3	Perangkat Lunak dan Library	46
BAB IV HASIL DAN PEMBAHASAN		47
4.1	Hasil Web Scraping	47
4.2	Pre Processing Data	49
4.2.1	Data Cleaning	49
4.2.2	Stopword Removal	50
4.2.3	Tokenizing	51
4.2.4	Stemming	52
4.3	Labeling	53
4.4	Visualisasi	55
4.4.1	Visualisasi Sentiment Positif	55
4.4.2	Visualisasi Sentiment Negatif	57
4.5	Pembagian Data	59
4.6	Feature Extraction	59
4.7	Teknik Resampling	60
4.7.1	Random Oversampling	61
4.7.2	Smote	62
4.7.3	Adasyn	62
4.7.4	Smoteenn	63
4.8	Evaluasi	64

BAB V PENUTUP	67
5.1 Kesimpulan	67
5.2 Saran	68
REFERENSI	69



DAFTAR TABEL

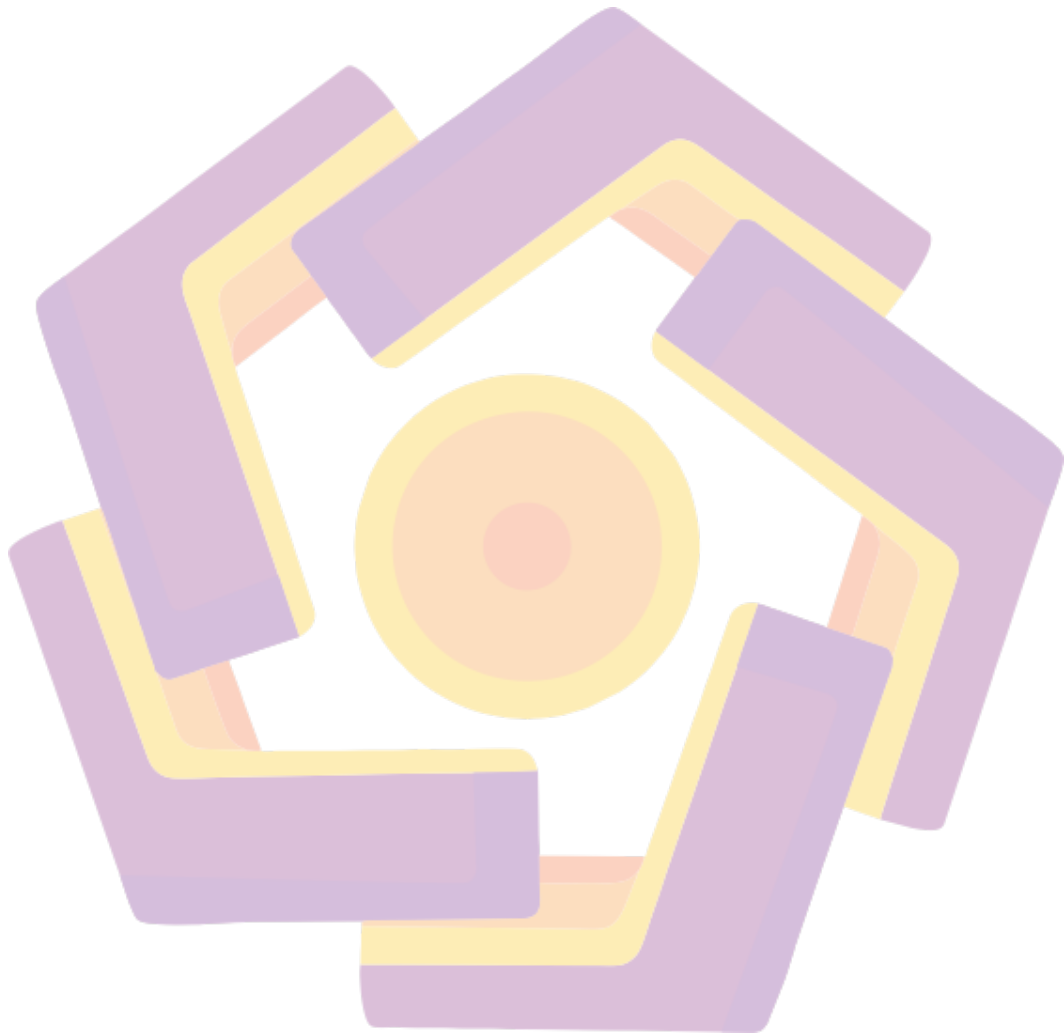
Tabel 2.1 Keaslian Penelitian	14
Tabel 2.2 <i>Confusion Matrix</i>	32
Tabel 4.1 Hasil Evaluasi <i>K-Fold10 Cross Validation</i>	64
Table 4.2 distribusi data sebelum dan sesudah resampling	59



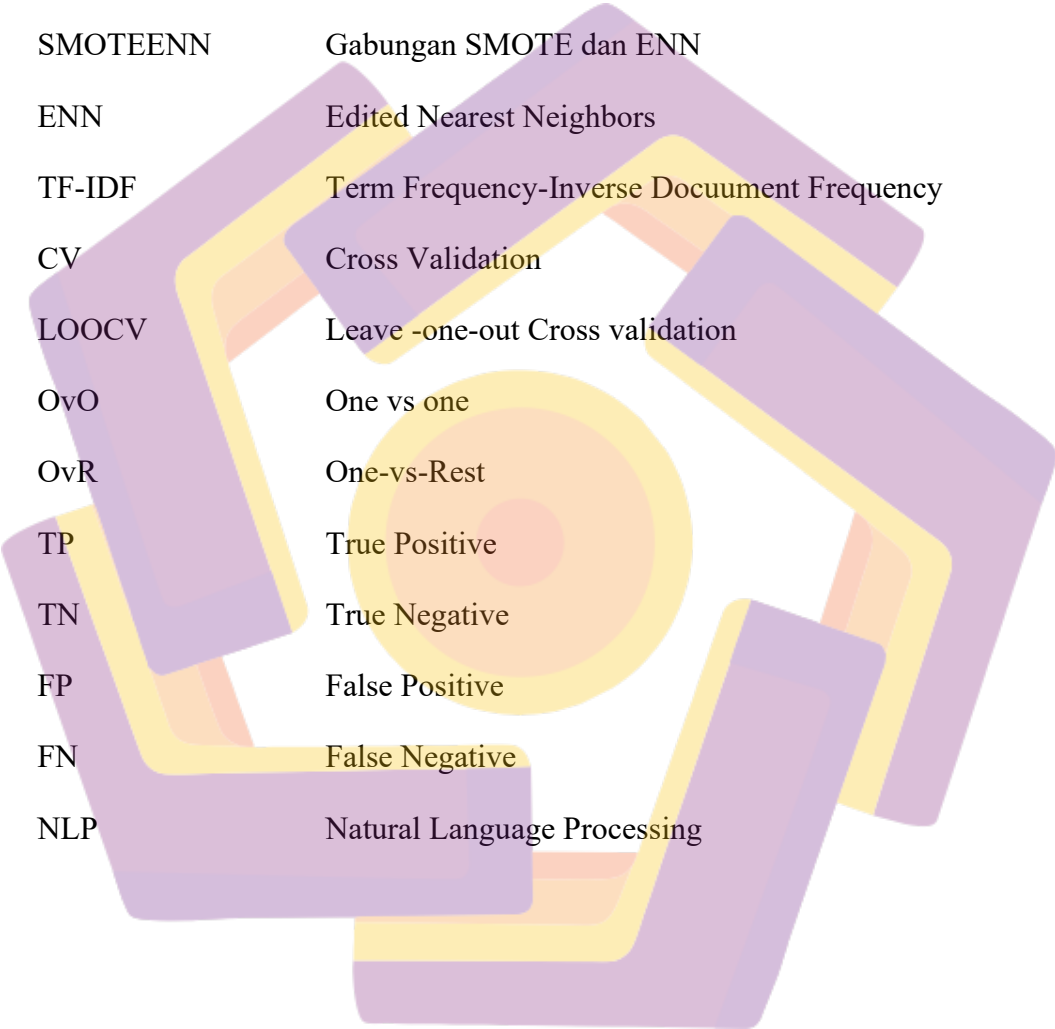
DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Hyperlane pada Algoritma SVM	27
Gambar 3.1 Alur Penelitian	34
Gambar 4.1 Hasil <i>Scrapping</i>	47
Gambar 4.2 Hasil Grafik Ulasan	48
Gambar 4.3 Hasil Grafik Skor Rating	49
Gambar 4.4 Hasil <i>Data Cleaning</i>	49
Gambar 4.5 Hasil <i>Stopword Removal</i>	50
Gambar 4.6 Hasil <i>Tokenizing</i>	51
Gambar 4.7 Hasil <i>Stemming</i>	52
Gambar 4.8 Hasil <i>Labeling</i>	53
Gambar 4.9 Hasil Distribusi Rating	54
Gambar 4.10 Grafik Perbandingan Sentimen	54
Gambar 4.11 Hasil Sentimen Positif	55
Gambar 4.12 Hasil 15 Kata Sentimen Positif	56
Gambar 4.13 Hasil Sentimen Negatif	57
Gambar 4.14 Hasil 15 Kata Sentimen Negatif	58
Gambar 4.15 Hasil Pembagian Data	59
Gambar 4.16 Hasil TF-IDF	59
Gambar 4.17 Evaluasi Model Sebelum <i>Resampling</i>	60
Gambar 4.18 <i>Resampling Random Oversampling</i>	61
Gambar 4.19 <i>Resampling Smote</i>	62

Gambar 4.20 <i>Resampling ADASYN</i>	62
Gambar 4.21 <i>Resampling Smoteenn</i>	63
Gambar 4.22 <i>Grafik Resampling</i>	65



DAFTAR LAMBANG DAN SINGKATAN



SVM	Support Vector Machines
SMOTE	Synthetic Minority Over-sampling Technique
ADASYN	Adaptive Synthetic Sampling
SMOTEENN	Gabungan SMOTE dan ENN
ENN	Edited Nearest Neighbors
TF-IDF	Term Frequency-Inverse Document Frequency
CV	Cross Validation
LOOCV	Leave-one-out Cross validation
OvO	One vs one
OvR	One-vs-Rest
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
NLP	Natural Language Processing

DAFTAR ISTILAH



Analisis sentimen	Besaran yang mempunyai arah
Web Scrapping	Pengumpulan data otomatis dari web
Text Preprocessing	Proses persiapan data sebelum dianalisis, termasuk cleaning, tokenizing, dan stopwords removal.
Dataset	Kumpulan data untuk analisis atau pelatihan.
Resampling Data	Penyesuaian data untuk mengatasi ketidakseimbangan kelas.
Case Folding	Proses mengubah seluruh huruf dalam teks menjadi huruf kecil dan menghilangkan karakter yang tidak relevan.
Stopword Removal	Penghapusan kata-kata umum yang tidak memiliki makna signifikan dalam analisis data.
Stemming	Proses mengubah kata-kata dalam teks menjadi bentuk dasar atau akar katanya.
Splitting	Proses membagi dataset menjadi data pelatihan dan data pengujian untuk memastikan model tidak overfitting.
Cross Validation	Teknik statistik untuk mengevaluasi kinerja model dengan membagi dataset menjadi data pelatihan dan pengujian secara terstruktur.
Hybrid Sampling	Kombinasi teknik oversampling dan undersampling untuk menangani ketidakseimbangan kelas pada dataset.
Oversampling	Teknik resampling untuk menambah jumlah data pada kelas minoritas.
Visualisasi	Penyajian data dalam bentuk grafis.
Tokenizing	Proses memecah teks menjadi unit-unit kecil, seperti kata atau kalimat.
Wordcloud	Representasi visual kata berdasarkan frekuensi.



TF-IDF	Metode untuk memberikan bobot pada kata berdasarkan frekuensi kemunculan dalam dokumen dan seberapa jarang kata tersebut muncul dalam korpus.
Sastrawi	Library Python yang digunakan untuk proses stemming dan analisis teks dalam bahasa Indonesia
Pandas	Library Python yang digunakan untuk manipulasi dan analisis data.
Akurasi	Ukuran yang menggambarkan seberapa banyak prediksi yang benar dari keseluruhan prediksi yang dilakukan oleh model.
Precision	Ukuran yang menunjukkan seberapa banyak prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model.
Recall	Ukuran yang menunjukkan seberapa banyak data positif yang berhasil diidentifikasi oleh model dari semua data positif
F1-Score	Ukuran performa model yang merupakan rata-rata harmonis dari precision dan recall.
Overfitting	Kondisi di mana model pembelajaran mesin terlalu menyesuaikan diri dengan data pelatihan, sehingga memiliki performa yang buruk pada data pengujian.
Labeling	Proses pemberian label atau kategori pada data untuk mengklasifikasikan sentimen atau jenis data lainnya.
Random oversampling	Teknik resampling dengan menggandakan sampel kelas minoritas secara acak hingga jumlahnya mendekati kelas mayoritas.
SMOTE	Teknik resampling yang membuat data baru di kelas minoritas dengan interpolasi antara sampel minoritas yang ada.
ADASYN	Teknik resampling yang dengan membuat data sintetis pada kelas minoritas secara adaptif, fokus pada data yang sulit diklasifikasikan.

SMOTEENN	Teknik resampling yang dilakukan dengan cara menggabungkan SMOTE untuk membuat data sintetis pada kelas minoritas dan Edited Nearest Neighbors untuk menghapus data yang dianggap noise.
Edited nearest neighbors	adalah teknik undersampling dengan menghapus data yang labelnya berbeda dengan mayoritas tetangga terdekatnya, sehingga data yang dianggap noise atau dihilangkan.
K-fold cross-validation standar	Metode validasi model dengan membagi data menjadi K bagian (<i>fold</i>) yang berukuran sama.
Leave-one-out-cross-validation	Metode validasi model yang menggunakan satu data sebagai data uji dan seluruh data lainnya sebagai data latih.
Strafied cross validation	Metode validasi model yang membagi data menjadi beberapa <i>fold</i> dengan mempertahankan proporsi kelas yang sama
Radial basis function	fungsi kernel yang paling populer dalam algoritma Support Vector Machine (SVM) untuk memecahkan masalah klasifikasi non-linier.
Logistic Regression	algoritma klasifikasi linear yang menghitung probabilitas suatu kelas menggunakan fungsi sigmoid, sederhana, cepat, dan sering dipakai sebagai baseline.
XGBoost	model probabilistik berbasis Teorema Bayes dengan asumsi independensi antar fitur, ringan, cepat, dan efektif khususnya untuk teks atau analisis sentimen.
Naive Bayes	metode ensemble yang menggabungkan banyak decision tree melalui voting, memberikan hasil stabil, akurat, dan lebih tahan terhadap overfitting.
Random Forest	algoritma boosting berbasis pohon keputusan yang mengoptimalkan performa dengan pembelajaran bertahap, sangat kuat dan sering unggul di kompetisi machine learning.

INTISARI

Grab merupakan salah satu platform transportasi daring yang banyak digunakan oleh masyarakat karena menawarkan berbagai layanan, seperti transportasi, pengantaran makanan, serta pembayaran digital. Sebagai aplikasi yang memiliki jutaan pengguna, Grab menerima berbagai ulasan dari penggunanya di Google Play Store. Data ini sangat berguna bagi pengembang aplikasi karena dapat memberikan informasi dalam meningkatkan kualitas dari layanan aplikasi. Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna terhadap aplikasi Grab dengan menerapkan metode Support Vector Machine (SVM). Salah satu kendala dalam penelitian ini adalah dominasi ulasan positif yang lebih tinggi dibandingkan ulasan negatif, sehingga berpotensi membuat model lebih cenderung mengenali ulasan positif dibandingkan negatif. Untuk mengatasi masalah tentang data yang akan di proses ini tidak seimbang atau tidak rata, maka penelitian ini menerapkan beberapa teknik resampling, seperti random oversampling, SMOTE, ADASYN dan SMOTEENN hal ini diterapkan agar data yang digunakan dalam proses analisis menjadi lebih seimbang, sehingga model dapat memahami dan mengenali pola sentimen dengan lebih akurat. Studi ini menghasilkan kebanyakan pengguna menyampaikan pendapat positif mengenai aplikasi. Model SVM yang telah melalui proses resampling terbukti memiliki akurasi lebih tinggi dibandingkan model yang tidak menerapkan teknik penyeimbangan data. Dengan demikian, penelitian ini memberikan pemahaman yang lebih mendalam mengenai bagaimana pengguna menilai layanan Grab berdasarkan ulasan yang mereka berikan. Teknik SMOTEENN memberikan akurasi tertinggi sebesar 97%, diikuti oleh random oversampling, ADASYN, SMOTE. Selain itu, hasil penelitian ini juga dapat menjadi referensi bagi pengembang dalam mengoptimalkan kualitas layanan berdasarkan umpan balik yang diberikan oleh pengguna.

Kata kunci: Google Play Store, Machine Learning, Support Vector Machine, Resampling, Sentimen Analisis.

ABSTRACT

Grab is one of the most widely used online transportation platforms, offering various services such as ride-hailing, food delivery, and digital payments. As an application with millions of users, Grab receives numerous reviews on the Google Play Store. These reviews are valuable for app developers as they provide insights into improving service quality. This study aims to classify user sentiment towards the Grab application using the Support Vector Machine (SVM) method. One challenge in this research is the dominance of positive reviews, which may cause the model to be biased toward recognizing positive sentiments over negative ones. To address the issue of imbalanced data, several resampling techniques were applied, including random oversampling, SMOTE, ADASYN and SMOTEENN. These techniques help balance the data used in the analysis, allowing the model to better recognize sentiment patterns. The study found that most users expressed positive opinions about the application. The SVM model with resampling techniques achieved higher accuracy compared to models without data balancing. Among the techniques, SMOTEENN produced the highest accuracy at 98%, followed by Random oversampling, ADASYN, SMOTE. Furthermore, the results of this study can serve as a reference for developers to optimize service quality based on user feedback.

Keyword: Google Play Store, Machine Learning, Support Vector Machine, Resampling, Sentiment Analysis.