

**OPTIMASI AKURASI KLASIFIKASI SENTIMEN ULASAN
GAME ROBLOX DENGAN KOMBINASI METODE NAIVE
BAYES DAN SMOTE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh
ADITIYA WAHYU ALEX S
22.11.4937

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

**OPTIMASI AKURASI KLASIFIKASI SENTIMEN ULASAN
GAME ROBLOX DENGAN KOMBINASI METODE NAIVE
BAYES DAN SMOTE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

ADITIYA WAHYU ALEX S

22.11.4937

Kepada

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

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMASI AKURASI KLASIFIKASI SENTIMEN ULASAN
GAME ROBLOX DENGAN KOMBINASI METODE NAIVE
BAYES DAN SMOTE**

yang disusun dan diajukan oleh

Aditiya Wahyu Alex S

22.11.4937

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Februari 2026

Dosen Pembimbing,



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

HALAMAN PENGESAHAN
SKRIPSI
OPTIMASI AKURASI KLASIFIKASI SENTIMEN ULASAN GAME
ROBLOX DENGAN KOMBINASI METODE NAIVE BAYES DAN
SMOTE

yang disusun dan diajukan oleh

Aditiya Wahyu Alex S

22.11.4937

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ferian Fauzi Abdullah, S.Kom., M.Kom.
NIK. 190302276

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

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



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aditiya Wahyu Alex S
NIM : 22.11.4937

Menyatakan bahwa Skripsi dengan judul berikut:

OPTIMASI AKURASI KLASIFIKASI SENTIMEN ULASAN GAME ROBLOX DENGAN KOMBINASI METODE NAIVE BAYES DAN SMOTE

Dosen Pembimbing : Rumini, 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 Februari 2025

Yang Menyatakan,



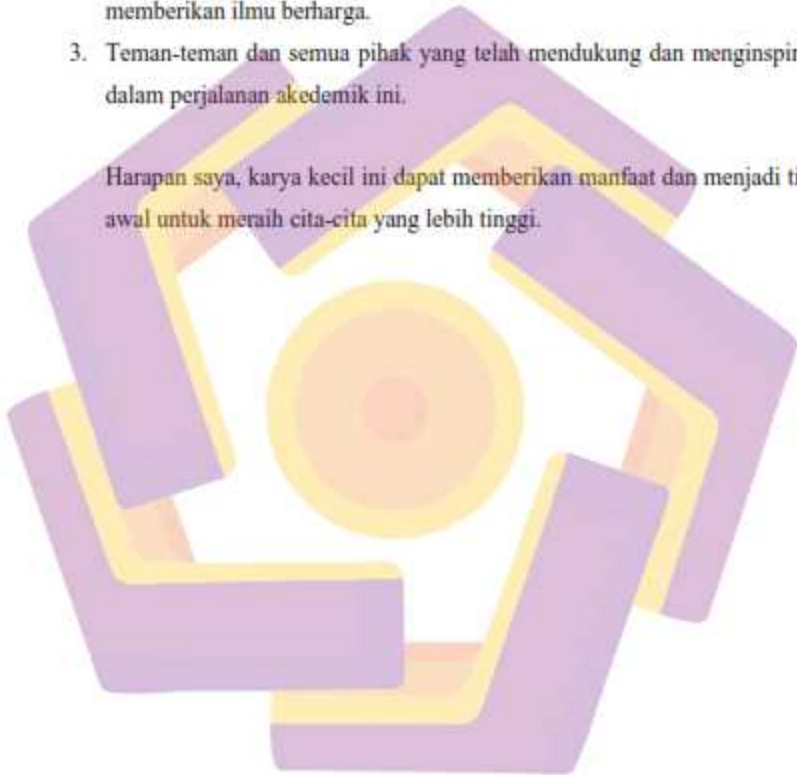
Aditiya Wahyu Alex S

HALAMAN PERSEMBAHAN

Dengan penuh rasa hormat, saya persembahkan skripsi ini kepada:

1. Orang tua dan keluarga tercinta yang senantiasa memberikan dukungan dan doa tanpa henti.
2. Dosen pembimbing Ibu Rumini yang dengan sabar membimbing dan memberikan ilmu berharga.
3. Teman-teman dan semua pihak yang telah mendukung dan menginspirasi dalam perjalanan akademik ini.

Harapan saya, karya kecil ini dapat memberikan manfaat dan menjadi titik awal untuk meraih cita-cita yang lebih tinggi.



KATA PENGANTAR

Puji dan syukur saya haturkan kepada Allah SWT karena berkat anugerah dan kasih-Nya, penulis berhasil menyelesaikan skripsi ini sebagai syarat untuk meraih gelar Sarjana Komputer dalam Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Dalam proses penulisan skripsi ini, tidak lepas dari berbagai dukungan, masukan, kritik, serta dukungan.

Pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusri, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom. selaku Ketua Program Studi Informatika dan Bapak Arifiyanto Hadinegoro, M.T. selaku Sekretaris Program Studi Informatika.
4. Ibu Rumini, S.Kom., M.Kom. selaku Dosen Pembimbing yang telah memberikan petunjuk, semangat, serta masukan, dan telah rela meluangkan waktu untuk mendampingi penulis sampai selesai penulisan skripsi ini dengan sebaik mungkin.
5. Seluruh civitan akademik, terutama para dosen dan staf karyawan di Fakultas Ilmu Komputer, telah memberikan dukungan dan pengetahuan selama penulis mengikuti proses pembelajaran.
6. Kedua orang tua tercinta dan sanak keluarga yang selalu mengirimkan doa, memberikan dukungan, serta memberikan semangat kepada penulis selama proses penulisan skripsi.
7. Teman-teman Informatika Angkatan 2022 yang sudah berusaha bersama untuk menyelesaikan skripsi.

Yogyakarta, 25 Februari 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	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT	xv
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	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	10
2.2.1 Analisis Sentimen	10
2.2.2 Google Play Store	10
2.2.3 Game	10
2.2.4 Roblox	11
2.2.5 Algoritma Naïve Bayes	11
2.2.6 Model SMOTE	12
2.2.7 Text Preprocessing	13
2.2.8 TF-IDF (<i>Term Frequency - Inverse Document Frequency</i>)	14
2.2.9 Confusion Matrix	15

BAB III METODE PENELITIAN	17
3.1 Objek Penelitian.....	17
3.2 Alur Penelitian	17
3.3 Alat dan Bahan.....	23
BAB IV HASIL DAN PEMBAHASAN	24
4.1 Pengumpulan Data	24
4.2 Preprocessing	24
4.3 Data Labeling.....	28
4.4 Wordcloud.....	28
4.5 TF-IDF (<i>Term Frequency - Inverse Document Frequency</i>)	30
4.6 Pembagian Data	31
4.7 SMOTE (<i>Synthetic Minority Over-sampling Technique</i>).....	32
4.8 Implementasi Model	32
4.9 Evaluasi Kinerja Algoritma	33
4.9.1 Confusion Matrix	33
4.9.2 Classification Report.....	36
4.9.3 ROC-AUC.....	37
4.9.4 Eksperimen Variasi Data Splitting dan Uji Stabilitas Model	38
4.10 Deploy.....	39
BAB V PENUTUP	43
5.1 Kesimpulan	43
5.2 Saran	43
REFERENSI	44
LAMPIRAN.....	50

DAFTAR TABEL

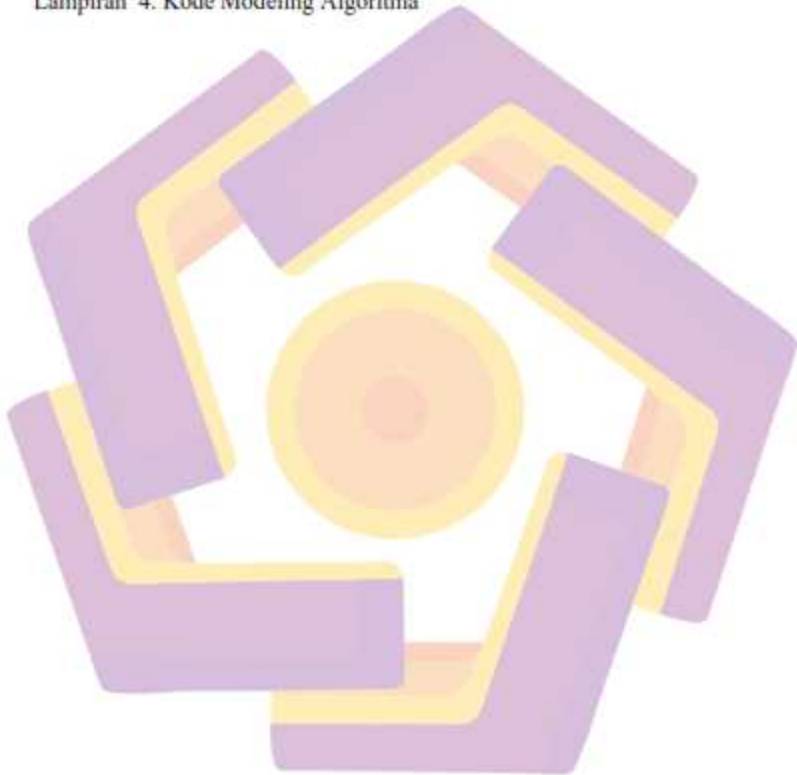
Tabel 2. 1 Confusion matrix	15
Tabel 4. 1 Scraping Data	24
Tabel 4. 2 Data Cleansing dan Case Folding	25
Tabel 4. 3 Normalisasi Kata	25
Tabel 4. 4 Tokenize	26
Tabel 4. 5 Stopword Removal	27
Tabel 4. 6 Stemming	27
Tabel 4. 7 Data Labeling	28
Tabel 4. 8 Hasil Bobot TF-IDF	31
Tabel 4. 9 Classification Report Naïve Bayes Baseline	36
Tabel 4. 10 Classification Report Naïve Bayes dengan SMOTE	36
Tabel 4. 11 Perbandingan Hasil Eksperimen Berdasarkan Skema Splitting	38
Tabel 4. 12 Rata-Rata Hasil Eksperimen Tiga Skema Splitting	38
Tabel 4. 13 Hasil 5-Fold Stratified Cross Validation	39

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	17
Gambar 4. 1 Wordcloud Before Preprocessing	29
Gambar 4. 2 Wordcloud After Preprocessing	30
Gambar 4. 3 Perbandingan Distribusi kelas sebelum dan sesudah SMOTE	32
Gambar 4. 4 Cofusion Matrix NB Baseline	33
Gambar 4. 5 Cofusion Matrix NB dengan SMOTE	34
Gambar 4. 6 Kurva ROC Perbandingan Model Multinomial Naive Bayes	37
Gambar 4. 7 Halaman Prediksi Sentimen	39
Gambar 4. 8 Halaman Dataset	40
Gambar 4. 9 Halaman Visualisasi	40
Gambar 4. 10 Halaman Metodologi	41
Gambar 4. 11 Pengujian Sentimen Positif	41
Gambar 4. 12 Pengujian Sentimen Negatif	42

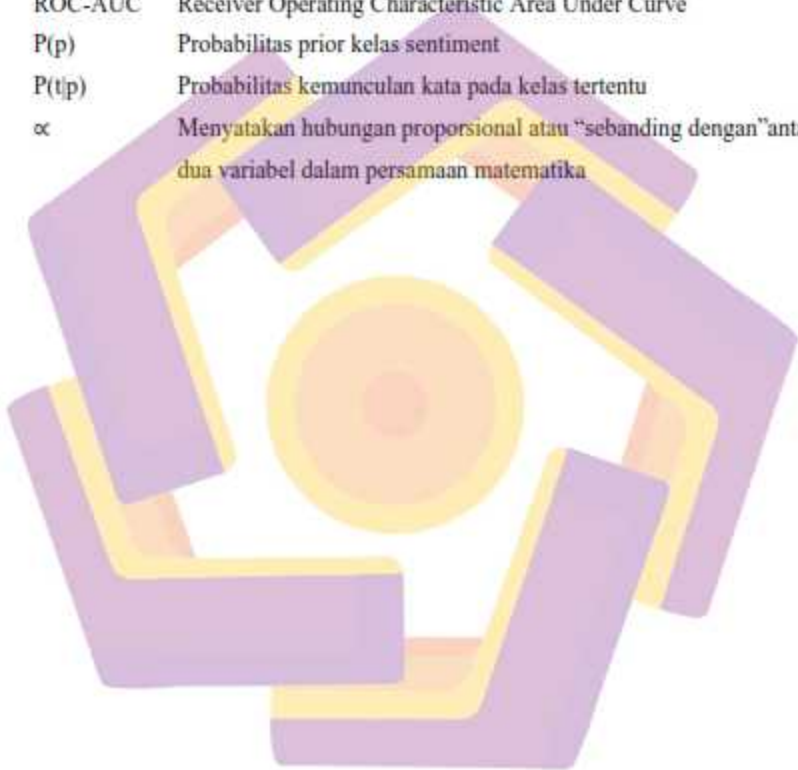
DAFTAR LAMPIRAN

Lampiran 1. Kode Pengumpulan Data ulasan Roblox	50
Lampiran 2. Kode Pelabelan Data	57
Lampiran 3. Kode Balancing Data dan TF-IDF	60
Lampiran 4. Kode Modeling Algoritma	62



DAFTAR LAMBANG DAN SINGKATAN

SMOTE	Teknik Oversampling Minoritas Sintetis.
TF-IDF	Term Frequency Inverse Document Frequency.
NLP	Natural Language Processing.
NB	Naïve Bayes
ROC-AUC	Receiver Operating Characteristic Area Under Curve
$P(p)$	Probabilitas prior kelas sentiment
$P(t p)$	Probabilitas kemunculan kata pada kelas tertentu
$\propto$	Menyatakan hubungan proporsional atau "sebanding dengan" antara dua variabel dalam persamaan matematika



DAFTAR ISTILAH

Analisis Sentimen	Proses mengklasifikasikan opini atau emosi dalam teks ke dalam kategori positif atau negatif
<i>Class Imbalance</i>	Kondisi ketika jumlah data pada satu kelas jauh lebih banyak dibanding kelas lainnya.
<i>Data Scraping</i>	Proses pengambilan data secara otomatis dari sebuah situs web.
<i>Oversampling</i>	Teknik untuk menyeimbangkan data dengan meningkatkan jumlah sampel di kelas minoritas.
<i>Lexicon-Based</i>	Metode pelabelan sentimen berdasarkan kamus kata positif dan negatif.
<i>Preprocessing</i>	Tahapan pembersihan dan normalisasi teks sebelum dianalisis
Multinomial NB	Varian Naive Bayes yang umum digunakan pada klasifikasi teks
<i>Data Training</i>	Data yang digunakan untuk melatih model.
<i>Data Testing</i>	Data yang digunakan untuk menguji akurasi model setelah dilatih.
Probabilitas Prior	Prior awal suatu kelas sebelum diperoleh data tambahan.
Likelihood	Probabilitas suatu fitur muncul dalam kelas tertentu.
<i>Library</i>	Kumpulan fungsi dan model yang digunakan untuk mempermudah pemrograman.
<i>Confusion Matrix</i>	Tabel evaluasi performa model klasifikasi
<i>Classification Report</i>	Ringkasan metrik evaluasi berupa precision, recall, F1-score, dan akurasi

INTISARI

Perkembangan sektor game digital menghasilkan jumlah ulasan pengguna yang sangat besar sehingga diperlukan metode otomatis untuk melakukan analisis sentimen secara efektif. Roblox sebagai salah satu platform game online terbesar memiliki ulasan berbahasa Indonesia dengan distribusi sentimen positif dan negatif yang tidak seimbang, yang dapat memengaruhi kinerja model klasifikasi. Penelitian ini bertujuan untuk meningkatkan efektivitas klasifikasi sentimen ulasan game Roblox dengan menggabungkan algoritma Multinomial Naive Bayes dan *Synthetic Minority Over-sampling Technique (SMOTE)* menggunakan 10.000 ulasan yang diambil dari Google Play Store pada Desember 2025. Tahapan penelitian meliputi pengumpulan data, *preprocessing teks*, pelabelan sentimen berbasis leksikon, ekstraksi fitur menggunakan TF-IDF, variasi pembagian data latih dan uji (70:30, 80:20, dan 90:10), penerapan SMOTE pada data latih, serta evaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan ROC AUC. Berdasarkan rata-rata tiga skema *splitting*, model dengan SMOTE memperoleh *accuracy* sebesar 83,54%, *precision* sebesar 82,64%, *recall* sebesar 82,21%, dan *F1-score* sebesar 82,40%, dengan nilai ROC AUC sebesar 0,9151 yang menunjukkan kemampuan diskriminatif sangat baik. Hasil tersebut menunjukkan bahwa penerapan SMOTE mampu meningkatkan keseimbangan performa antar kelas serta menghasilkan model klasifikasi yang lebih stabil dan representatif pada dataset yang tidak seimbang.

Kata kunci: Analisis sentimen, Roblox, Naive Bayes, SMOTE, TF-IDF

ABSTRACT

The growth of the digital gaming sector has resulted in a huge number of user reviews, requiring automated methods to perform effective sentiment analysis. Roblox, as one of the largest online gaming platforms, has Indonesian-language reviews with an uneven distribution of positive and negative sentiments, which can affect the performance of classification models. This study aims to improve the effectiveness of Roblox game review sentiment classification by combining the Multinomial Naive Bayes algorithm and the Synthetic Minority Over-sampling Technique (SMOTE) using 10,000 reviews taken from the Google Play Store in December 2025. The research stages include data collection, text preprocessing, lexicon-based sentiment labeling, feature extraction using TF-IDF, variations in training and testing data division (70:30, 80:20, and 90:10), application of SMOTE to the training data, and evaluation of model performance using accuracy, precision, recall, F1-score, and ROC AUC metrics. Based on the average of the three splitting schemes, the model with SMOTE obtained an accuracy of 83.54%, precision of 82.64%, recall of 82.21%, and F1-score of 82.40%, with an ROC AUC value of 0.9151, indicating excellent discriminatory ability. These results show that the application of SMOTE can improve the balance of performance between classes and produce a more stable and representative classification model in unbalanced datasets.

Keyword: Sentiment Analysis, Roblox, Naive Bayes, SMOTE, TF-IDF.