

**ANALISIS SENTIMEN ULASAN APLIKASI GOOGLE
GEMINI PADA GOOGLE PLAY STORE MENGGUNAKAN
ALGORITMA NAIVE BAYES**

SKRIPSI

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



disusun oleh
ITSNAINI QIWAM ROBBAYANI
21.11.4141

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

**ANALISIS SENTIMEN ULASAN APLIKASI GOOGLE
GEMINI PADA GOOGLE PLAY STORE MENGGUNAKAN
ALGORITMA NAIVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ITSNAINI QIWAM ROBBAYANI

21.11.4141

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN APLIKASI GOOGLE GEMINI PADA
GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE**

BAYES

yang disusun dan diajukan oleh

Itsaini Qiwan Robbayani

21.11.4141

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Februari 2025

Dosen Pembimbing,



Ike Verawati, M.Kom

NIK. 190302237

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN APLIKASI GOOGLE GEMINI PADA
GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE**

BAYES

yang disusun dan diajukan oleh

Itsnaini Qiwam Robbayani

21.11.4141

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dwi Nurani, M.Kom
NIK. 190302236

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393

Ike Verawati, M.Kom
NIK. 190302237



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 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 : Itsnaini Qiwam Robbayani
NIM : 21.11.4141

Menyatakan bahwa Skripsi dengan judul berikut :

Analisis Sentimen Ulasan Aplikasi Google Gemini Pada Google Play Store Menggunakan Algoritma Naive Bayes

Dosen Pembimbing : Ike Verawati, 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, 19 Februari 2025

Yang Menyatakan,



Itsnaini Qiwam Robbayani

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Ulasan Aplikasi Google Gemini pada Google Play Store Menggunakan Algoritma Naive Bayes” ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Fakultas Ilmu Komputer Program Studi Informatika Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, penulis mendapatkan banyak bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom, selaku Ketua Program Studi SI Informatika Universitas Amikom Yogyakarta.
4. Ibu Ike Verawati, M.Kom, selaku dosen pembimbing yang telah memberikan bimbingan, masukan, serta motivasi kepada penulis dalam menyelesaikan skripsi ini.
5. Keluarga tercinta yang selalu memberikan doa, dukungan, dan semangat yang tiada henti kepada penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa yang akan datang. Semoga skripsi ini bermanfaat dan dapat menjadi referensi bagi penelitian selanjutnya.

Yogyakarta, 25 Februari 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
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	13
2.2.1 Analisis Sentimen	13
2.2.2 Google Play Store	13
2.2.3 Google Gemini	13
2.2.4 Algoritma Naive Bayes	14
2.2.5 Preprocessing	15
2.2.6 Pembobotan TF-IDF	15
2.2.7 Oversampling	16
2.2.8 Undersampling	16
2.2.9 Confusion Matrix	16

BAB III METODE PENELITIAN	19
3.1 Objek Penelitian	19
3.2 Alur Penelitian	19
3.3 Alat dan Bahan	21
BAB IV HASIL DAN PEMBAHASAN	23
4.1 Pengumpulan Data	23
4.2 Pelabelan Data	25
4.3 Text Preprocessing	26
4.3.1 Case Folding	26
4.3.2 Cleansing	27
4.3.3 Stopwords Removal	28
4.3.4 Tokenizing	30
4.3.5 Stemming	31
4.4 Splitting Data	33
4.5 Pembobotan TF-IDF	34
4.6 SMOTE (<i>Synthetic Minority Oversampling Technique</i>)	36
4.7 ADASYN (<i>Adaptive Synthetic Sampling</i>)	37
4.8 ROS (<i>Random Oversampling</i>)	39
4.9 Tomek Links	40
4.10 ENN (<i>Edited Nearest Neighbor</i>)	41
4.11 RUS (<i>Random Undersampling</i>)	43
4.12 Kombinasi SMOTE - Tomek Links	44
4.13 Kombinasi SMOTE - ENN	46
4.14 Kombinasi ADASYN - RUS	48
4.15 Klasifikasi Model	49
4.16 Evaluasi	51
4.17 Perbandingan Hasil Evaluasi	66
BAB V PENUTUP	68
5.1 Kesimpulan	68
5.2 Saran	68
REFERENSI	69

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2 Confusion Matrix	17
Tabel 3.1 Sampel Data	22
Tabel 4.1 Hasil Scraping Data	24
Tabel 4.2 Hasil Filtering	24
Tabel 4.3 Hasil Pelabelan Data	25
Tabel 4.4 Hasil Case Folding	26
Tabel 4.5 Hasil Cleansing	28
Tabel 4.6 Hasil Stopwords Removal	29
Tabel 4.7 Hasil Tokenizing	31
Tabel 4.8 Hasil Stemming	32
Tabel 4.9 Hasil Pembobotan TF-IDF	35
Tabel 4.10 Confusion Matrix	52
Tabel 4.11 Confusion Matrix Setelah SMOTE	52
Tabel 4.12 Confusion Matrix Setelah ADASYN	53
Tabel 4.13 Confusion Matrix Setelah ROS	54
Tabel 4.14 Confusion Matrix Setelah Tomek Links	55
Tabel 4.15 Confusion Matrix Setelah ENN	55
Tabel 4.16 Confusion Matrix Setelah RUS	56
Tabel 4.17 Confusion Matrix Setelah SMOTE - Tomek Links	57
Tabel 4.18 Confusion Matrix Setelah SMOTE - ENN	58
Tabel 4.19 Confusion Matrix Setelah ADASYN - RUS	58
Tabel 4.20 Hasil Evaluasi	60
Tabel 4.21 Hasil Evaluasi Setelah SMOTE	60
Tabel 4.22 Hasil Evaluasi Setelah ADASYN	61
Tabel 4.23 Hasil Evaluasi Setelah ROS	62
Tabel 4.24 Hasil Evaluasi Setelah Tomek Links	62
Tabel 4.25 Hasil Evaluasi Setelah ENN	63
Tabel 4.26 Hasil Evaluasi Setelah RUS	64
Tabel 4.27 Hasil Evaluasi Setelah SMOTE - Tomek Links	64
Tabel 4.28 Hasil Evaluasi Setelah SMOTE - ENN	65
Tabel 4.29 Hasil Evaluasi Setelah ADASYN - RUS	66
Tabel 4.30 Perbandingan Hasil Evaluasi	66

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	19
Gambar 4.1 Flowchart Pengumpulan Data	23
Gambar 4.2 Source Code Scraping Data	23
Gambar 4.3 Source Code Filtering	24
Gambar 4.4 Flowchart Pelabelan Data	25
Gambar 4.5 Source Code Pelabelan Data	25
Gambar 4.6 Flowchart Case Folding	26
Gambar 4.7 Source Code Case Folding	26
Gambar 4.8 Flowchart Cleansing	27
Gambar 4.9 Source Code Cleansing	27
Gambar 4.10 Flowchart Stopwords Removal	29
Gambar 4.11 Source Code Stopwords Removal	29
Gambar 4.12 Flowchart Tokenizing	30
Gambar 4.13 Source Code Tokenizing	30
Gambar 4.14 Flowchart Stemming	31
Gambar 4.15 Source Code Stemming	32
Gambar 4.16 Flowchart Splitting Data	33
Gambar 4.17 Source Code Splitting Data	33
Gambar 4.18 Hasil Splitting Data	33
Gambar 4.19 Distribusi Kelas	34
Gambar 4.20 Flowchart Pembobotan TF-IDF	34
Gambar 4.21 Source Code Pembobotan TF-IDF	35
Gambar 4.22 Flowchart SMOTE	36
Gambar 4.23 Source Code SMOTE	36
Gambar 4.24 Distribusi Kelas Setelah SMOTE	37
Gambar 4.25 Flowchart ADASYN	37
Gambar 4.26 Source Code ADASYN	38
Gambar 4.27 Distribusi Kelas Setelah ADASYN	38
Gambar 4.28 Flowchart ROS	39
Gambar 4.29 Source Code ROS	39
Gambar 4.30 Distribusi Kelas Setelah ROS	40
Gambar 4.31 Flowchart Tomek Links	40
Gambar 4.32 Source Code Tomek Links	41
Gambar 4.33 Distribusi Kelas Setelah Tomek Links	41
Gambar 4.34 Flowchart ENN	42
Gambar 4.35 Source Code ENN	42
Gambar 4.36 Distribusi Kelas Setelah ENN	43
Gambar 4.37 Flowchart RUS	43
Gambar 4.38 Source Code RUS	44
Gambar 4.39 Distribusi Kelas Setelah RUS	44
Gambar 4.40 Flowchart SMOTE - Tomek Links	45
Gambar 4.41 Source Code SMOTE - Tomek Links	45
Gambar 4.42 Distribusi Kelas Setelah SMOTE - Tomek Links	46

Gambar 4.43 Flowchart SMOTE - ENN	46
Gambar 4.44 Source Code SMOTE - ENN	47
Gambar 4.45 Distribusi Kelas Setelah SMOTE - ENN	47
Gambar 4.46 Flowchart ADASYN - RUS	48
Gambar 4.47 Source Code ADASYN - RUS	48
Gambar 4.48 Distribusi Kelas Setelah ADASYN - RUS	49
Gambar 4.49 Flowchart Klasifikasi Model	50
Gambar 4.50 Source Code Klasifikasi Model	50
Gambar 4.51 Hasil Klasifikasi	50
Gambar 4.52 Flowchart Evaluasi	51
Gambar 4.53 Source Code Heatmap Confusion Matrix	51
Gambar 4.54 Heatmap Confusion Matrix	52
Gambar 4.55 Heatmap Confusion Matrix Setelah SMOTE	53
Gambar 4.56 Heatmap Confusion Matrix Setelah ADASYN	53
Gambar 4.57 Heatmap Confusion Matrix Setelah ROS	54
Gambar 4.58 Heatmap Confusion Matrix Setelah Tomek Links	55
Gambar 4.59 Heatmap Confusion Matrix Setelah ENN	56
Gambar 4.60 Heatmap Confusion Matrix Setelah RUS	56
Gambar 4.61 Heatmap Confusion Matrix Setelah SMOTE - Tomek Links	57
Gambar 4.62 Heatmap Confusion Matrix Setelah SMOTE - ENN	58
Gambar 4.63 Heatmap Confusion Matrix Setelah ADASYN - RUS	59
Gambar 4.64 Source Code Evaluasi	59

DAFTAR LAMBANG DAN SINGKATAN

Π	Notasi produk (perkalian berulang dari suatu rangkaian elemen)
\propto	Simbol proporsional (berbanding lurus dengan)
TF-IDF	Term Frequency Inverse Document Frequency
SMOTE	Synthetic Minority Oversampling Technique
ADASYN	Adaptive Synthetic Sampling
ROS	Random Oversampling
ENN	Edited Nearest Neighbor
RUS	Random Undersampling



DAFTAR ISTILAH

Data Training	Data yang digunakan untuk melatih model
Data Testing	Data yang digunakan untuk menguji akurasi model setelah dilatih
Probabilitas Prior	Probabilitas awal suatu kelas sebelum diperoleh data tambahan
Likelihood	Probabilitas suatu fitur muncul dalam kelas tertentu
Overfitting	Model terlalu menyesuaikan data training sehingga kurang akurat pada data baru
Library	Kumpulan fungsi dan modul yang digunakan untuk mempermudah pemrograman
Heatmap	Visualisasi data dengan warna untuk menunjukkan pola atau intensitas data

INTISARI

Google Gemini merupakan platform percakapan berbasis AI yang menggunakan pembelajaran mesin dan pemrosesan bahasa alami, yang dirancang untuk meningkatkan produktivitas pengguna. Seiring dengan meningkatnya penggunaan aplikasi ini, berbagai ulasan dari pengguna semakin banyak bermunculan di Google Play Store. Oleh karena itu, diperlukan pendekatan analisis sentimen untuk mengetahui apakah ulasan pengguna cenderung positif atau negatif, namun salah satu masalah dalam klasifikasi analisis sentimen adalah ketidakseimbangan data. Penelitian ini menggunakan algoritma Naive Bayes dengan menerapkan beberapa teknik resampling, seperti SMOTE, ADASYN, Random Oversampling, Tomek Links, ENN, dan Random Undersampling, untuk mengatasi ketidakseimbangan data. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki kinerja yang baik dalam mengklasifikasikan sentimen, dengan teknik ADASYN dan ENN menghasilkan akurasi tertinggi yaitu sebesar 87%. Pada teknik ADASYN, model berhasil mengklasifikasikan sebanyak 710 ulasan positif dan 159 ulasan negatif. Dan pada teknik ENN, model berhasil mengklasifikasikan sebanyak 793 ulasan positif dan 75 ulasan negatif. Hasil ini menunjukkan bahwa mayoritas ulasan bersifat positif.

Kata kunci: Analisis Sentimen, Google Gemini, Algoritma Naive Bayes

ABSTRACT

Google Gemini is an AI-based conversation platform that uses machine learning and natural language processing, designed to improve user productivity. Along with the increasing use of this application, various reviews from users are increasingly appearing on the Google Play Store. Therefore, a sentiment analysis approach is needed to find out whether user reviews tend to be positive or negative, but one of the problems in sentiment analysis classification is data imbalance. This research uses the Naive Bayes algorithm by applying several resampling techniques, such as SMOTE, ADASYN, Random Oversampling, Tomek Links, ENN, and Random Undersampling, to overcome data imbalance. The results show that the Naive Bayes algorithm performs well in classifying sentiment, with the ADASYN and ENN techniques producing the highest accuracy of 87%. In the ADASYN technique, the model successfully classified as many as 710 positive reviews and 159 negative reviews. And in the ENN technique, the model successfully classified 793 positive reviews and 75 negative reviews. These results show that the majority of reviews are positive.

Keyword: Sentiment Analysis, Google Gemini, Naive Bayes Algorithm