

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN
SUPPORT VECTOR MACHINE DALAM KLASIFIKASI
SENTIMEN REVIEW APLIKASI IDN**

SKRIPSI

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



disusun oleh

MA'RUF SYAMSUDIN ASH SHIDIQ

21.11.3957

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN
SUPPORT VECTOR MACHINE DALAM KLASIFIKASI
SENTIMEN REVIEW APLIKASI IDN**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MA'RUF SYAMSUDIN ASH SHIDIQ

21.11.3957

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT
VECTOR MACHINE DALAM KLASIFIKASI SENTIMEN REVIEW
APLIKASI IDN**

yang disusun dan diajukan oleh

Ma'ruf Syamsudin Ash Shidiq

21.11.3957

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Juli 2025

Dosen Pembimbing,



Lukman, S.Kom., M.Kom.

NIK. 190302151

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT
VECTOR MACHINE DALAM KLASIFIKASI SENTIMEN REVIEW
APLIKASI IDN

yang disusun dan diajukan oleh

Ma'ruf Syamsudin Ash Shidiq

21.11.3957

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Juli 2025

Susunan Dewan Penguji

Nama Penguji

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

Nur Aini, A.Md., S.Kom., M.Kom.
NIK. 190302066

Lukman, S.Kom., M.Kom.
NIK. 190302151

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ma'ruf Syamsudin Ash Shidiq
NIM : 21.11.3957

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI SENTIMEN REVIEW APLIKASI IDN

Dosen Pembimbing : Lukman 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, 22 Juli 2025

Yang Menyatakan,



Ma'ruf Syamsudin Ash Shidiq

HALAMAN PERSEMBAHAN

Dengan segenap rasa Syukur dan penuh cinta, karya ini kupersembahkan untuk :

1. Allah SWT, Tuhan yang maha esa atas rahmat dan ridho-Nya dalam setiap langkahku.
2. Suri tauladan, Nabi Muhammad SAW yang ajarannya menjadi cahaya dalam gelap, petunjuk dalam setiap pencarian.
3. Keluarga, terutama orang tua atas doa, dukungan, dan kasih sayang yang tak pernah berhenti.
4. Sahabat, dan teman perjuangan, yang selalu hadir dengan tawa dan semangat saat aku nyaris menyerah.
5. Dosen pembimbing dan penguji, atas bimbingan, ilmu, serta kesempatan untuk tumbuh dan belajar.
6. Diri saya sendiri karena sudah bertahan sejauh ini meski tak selalu mudah.
7. Untuk seseorang yang selalu aku kagumi, terima kasih telah memberi semangat di saat aku nyaris menyerah. Kamu tahu kamu spesial setidaknya bagiku

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul: "**Perbandingan Algoritma Naïve Bayes dan Support Vector Machine dalam Klasifikasi Sentimen Review Aplikasi IDN App**", sebagai salah satu syarat untuk memperoleh gelar Sarjana di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penyusunan skripsi ini tentunya tidak lepas dari dukungan, bimbingan, dan doa dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, penulis ingin menyampaikan apresiasi kepada:

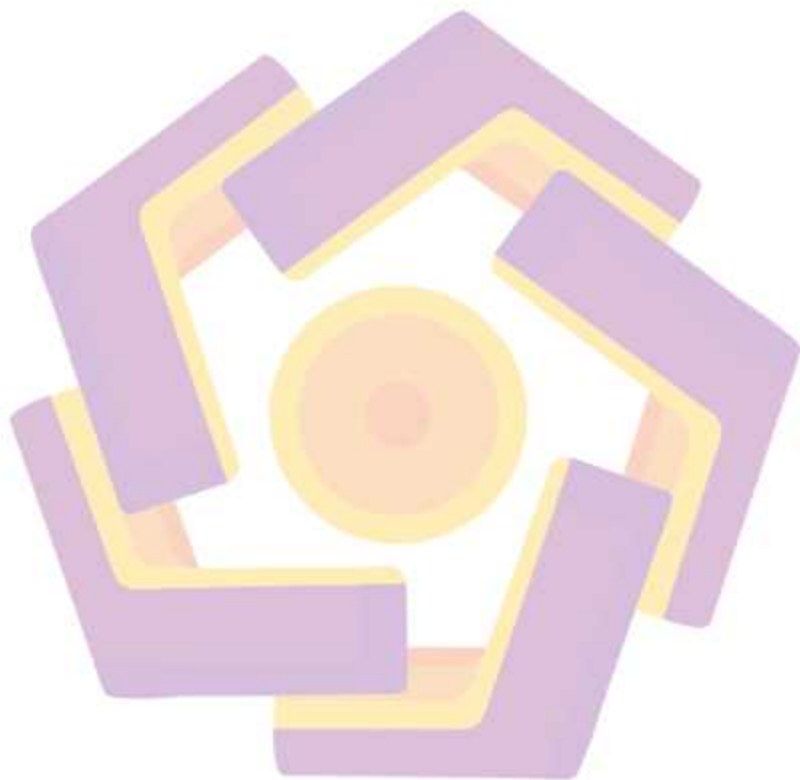
1. Allah SWT, atas segala nikmat dan kekuatan yang diberikan selama proses ini berlangsung.
2. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
3. Ibu Prof. Dr. Kusri, M.Kom, selaku Dekan Fakultas Ilmu Komputer.
4. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika.
5. Bapak Lukman S.Kom., M.Kom., selaku dosen pembimbing yang telah membimbing penulis dengan penuh kesabaran dan dedikasi.
6. Orang tua dan keluarga penulis yang senantiasa mendoakan dan memberikan dukungan moril maupun materil.
7. Teman-teman seperjuangan yang selalu hadir memberi semangat dan bantuan dalam setiap proses.
8. Seseorang yang telah menjadi semangat tambahan dalam perjalanan ini terima kasih atas dukungan yang tidak pernah gagal membuat penulis tersenyum.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis sangat terbuka terhadap kritik dan saran yang membangun demi perbaikan ke depannya. Semoga skripsi ini dapat memberikan manfaat bagi

pembaca dan menjadi kontribusi kecil dalam pengembangan ilmu pengetahuan, khususnya di bidang analisis sentimen dan pembelajaran mesin.

Yogyakarta, 22 Juli 2025

Penulis

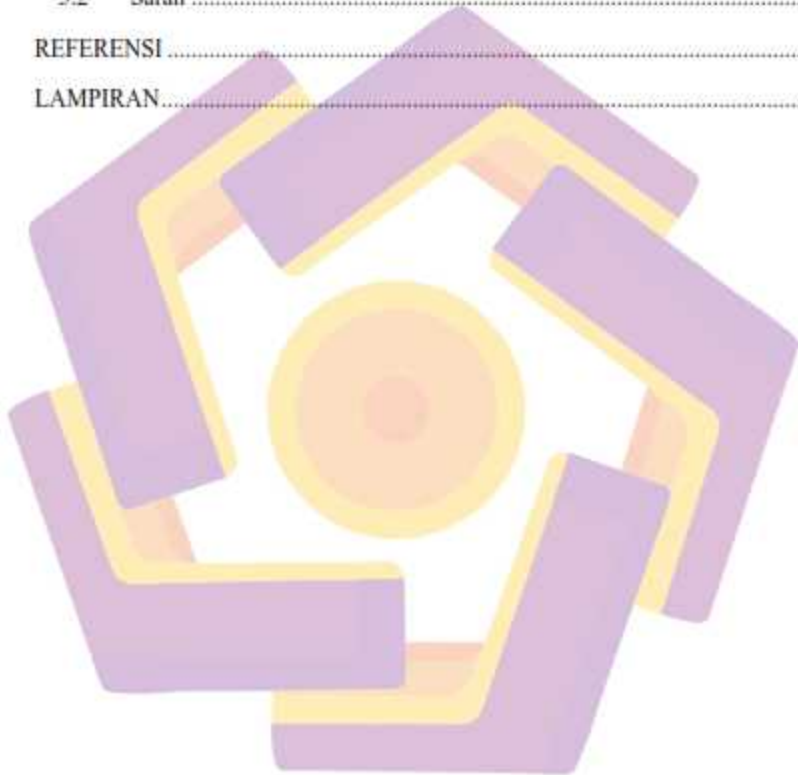


DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
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.....	9
2.2.1.	Analisis Sentimen	9
2.2.2.	Text Mining	9
2.2.3.	Google Play Store	9
2.2.4.	Pre-Processing.....	9
2.2.5.	Naïve Bayes Classifier.....	10
2.2.6.	Support Vektor Machine.....	12
BAB III METODE PENELITIAN		13
3.1	Objek Penelitian.....	13
3.2	Analisis requirement	13
3.2.1.	Data Penelitian	13
3.2.2.	Alat/Instrumen	14
3.3	Alur Penelitian	16
3.4	Uraian Penelitian.....	16
3.4.1.	Crawling Data	16
3.4.2.	Labeling	16
3.4.3.	Preprocessing data	19
3.4.4.	Penyimbangan Data (Undersampling).....	21
3.4.5.	Ekstrasi Fitur Menggunakan TF-IDF.....	22
BAB IV HASIL DAN PEMBAHASAN		26
4.1	Gambaran umum hasil penelitian	26
4.2	Hasil pemodelan dengan Naïve Bayes.....	26
4.3	Hasil pemodelan dengan SVM	28
4.4	Perbandingan Kinerja Algoritma	30
4.5	Simulasi Perhitungan Manual Klasifikasi.....	33

4.5.1	Simulasi Naïve Bayes Classifier	33
4.5.2	Simulasi Konsep SVM.....	35
BAB V PENUTUP		37
5.1	Kesimpulan	37
5.2	Saran	37
REFERENSI		39
LAMPIRAN.....		43



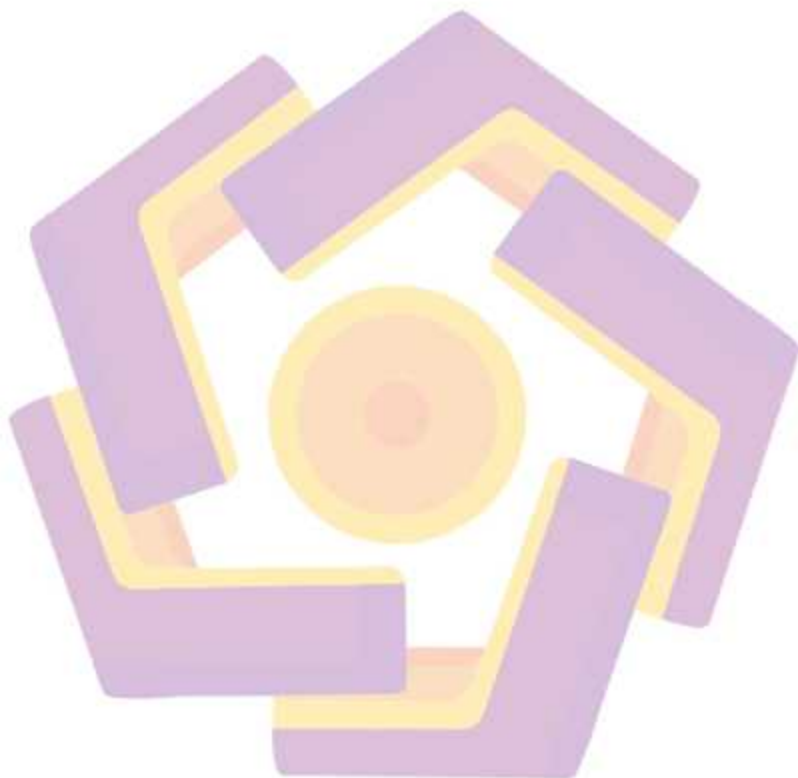
DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	7
Tabel 3.1. Contoh Hasil Labeling	18
Tabel 3.2. Hasil Case Folding	20
Tabel 3.3. Hasil Cleansing	20
Tabel 3.4. Hasil Tokenizing	21
Tabel 3.5. Hasil Stopword Removal	21
Tabel 3.6. Hasil Normalisasi	21
Tabel 3.7. Hasil Stemming	22
Tabel 3.8. Top 10 kata berdasarkan rata-rata bobot TF-IDF	25
Tabel 4.1. Hasil Evaluasi Model Naïve Bayes	26
Tabel 4.2. Confusion Matrix Naïve Bayes	27
Tabel 4.3. Hasil Evaluasi Model SVM	29
Tabel 4.4. Confusion Matrix SVM	29
Tabel 4.5. Perbandingan Hasil Naïve Bayes dan SVM	31



DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	17
Gambar 4.1. Heatmap Confusion Matrix Naïve Bayes	28
Gambar 4.2. Heatmap Confusion Matrix SVM	30



DAFTAR LAMBANG DAN SINGKATAN



SVM	Support Vector Machines
NB	Naïve Bayes
TF-IDF	Term Frequency - Inverse Document Frequency
NLP	Natural Language Processing
API	Application Programming Interface
CSV	Comma Separated Values
GPU	Graphics Processing Unit
RAM	Random Access Memory
HTML	HyperText Markup Language
TF	Term Frequency
IDF	Inverse Document Frequency
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
F1-Score	Harmonic Mean of Precision and Recall

DAFTAR ISTILAH

Sentimen atau negatif)	Perasaan atau opini yang diekspresikan dalam teks (positif
Text Mining	Proses mengekstrak informasi dari data teks tidak terstruktur
Preprocessing	Tahapan pembersihan dan persiapan data sebelum dianalisis
Case Folding	Proses mengubah huruf menjadi huruf kecil semua
Tokenizing	Proses memecah teks menjadi kata-kata atau token
Stopword "ke", "dan")	Kata-kata umum yang dianggap tidak bermakna (seperti "di",
Stemming	Proses mengubah kata ke bentuk dasar atau akar katanya
Normalisasi	Proses mengubah kata tidak baku ke bentuk baku
Confusion Matrix	Tabel evaluasi yang menunjukkan hasil prediksi vs kenyataan
Akurasi	Persentase prediksi yang benar dari total prediksi
Precision	Tingkat keakuratan prediksi positif
Recall	Kemampuan model dalam menemukan seluruh data positif
Bias hyperplane	Nilai konstan dalam model SVM untuk menggeser
Hyperplane	Garis atau bidang pemisah antara dua kelas dalam SVM
Undersampling	Teknik mengurangi jumlah data dari kelas mayoritas
Oversampling	Teknik menambah jumlah data dari kelas minoritas
Labeling	Proses memberi label (positif/negatif) pada data
Cleansing penting	Proses menghapus karakter, simbol, atau teks yang tidak
TF-IDF kepentingan kata	Metode pemberian bobot kata berdasarkan frekuensi dan

INTISARI

Penggunaan aplikasi mobile yang terus meningkat mendorong perlunya pemahaman terhadap opini pengguna melalui ulasan yang mereka berikan, khususnya di platform Google Play Store. Salah satu aplikasi yang menjadi fokus dalam penelitian ini adalah IDN App, sebuah aplikasi berbasis media digital yang menasar kalangan generasi muda. Banyaknya ulasan dari pengguna memunculkan tantangan dalam mengklasifikasikan sentimen secara otomatis. Ketidakseimbangan jumlah ulasan positif dan negatif juga menjadi masalah yang dapat mempengaruhi performa model klasifikasi. Oleh karena itu, penelitian ini penting untuk mengetahui algoritma klasifikasi yang paling optimal dalam mengenali sentimen ulasan pengguna.

Penelitian ini menggunakan dua algoritma machine learning yaitu Naïve Bayes dan Support Vector Machine (SVM). Data dikumpulkan melalui crawling menggunakan library Python dari Google Play Store, kemudian dilakukan proses labeling otomatis berbasis skor bintang, dilanjutkan dengan preprocessing teks (case folding, tokenizing, stopword removal, normalisasi, dan stemming). Untuk mengatasi ketimpangan kelas, dilakukan metode random undersampling dan selanjutnya fitur diekstraksi menggunakan metode TF-IDF. Dataset akhir berjumlah 3836 data yang terbagi rata antara sentimen positif dan negatif. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, f1-score, dan confusion matrix.

Hasil penelitian menunjukkan bahwa algoritma SVM unggul dalam hal akurasi (0.89), stabilitas performa antar kelas, dan kemampuan mengenali ulasan negative. Meskipun Naïve Bayes lebih cepat dalam pelatihan, secara keseluruhan SVM dinilai lebih efektif untuk klasifikasi sentimen pada ulasan pengguna aplikasi. Penelitian ini dapat dimanfaatkan oleh pengembang aplikasi, peneliti dibidang text mining, serta pelaku bisnis digital untuk memahami persepsi pengguna dan meningkatkan layanan berbasis data.

Kata kunci: Sentimen, IDN App, Naïve Bayes, SVM

ABSTRACT

The increasing use of mobile applications highlights the importance of understanding user opinions through their reviews, especially on the Google Play Store platform. This study focuses on IDN App, a digital media application targeting young audiences. With the abundance of user review, there is a growing need for effective sentiment classification. However, the imbalance between positive and negative reviews presents a challenge that may affect the performance of classification model. Therefore, this research aims to identify the most optimal classification algorithm for categorizing user sentiments.

This study implements two machine learning algorithms: Naïve Bayes and Support Vector Machine(SVM). Review data were collected using a web scraping method via python, followed by automatic sentiment labeling based on star ratings. The data underwent several preprocessing steps including case folding, tokenizing, stopword removal, normalization, and stemming. To address class imbalance, the random undersampling method was applied. Feature extraction was conducted using term frequency-inverse document frequency (TF-IDF). A total of 3836 balanced reviews were used for model training and evaluation. The models were assessed using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

The result show that the SVM algorithm outperforms Naïve bayes in terms of accuracy (0.89), class balance performance, and its ability to detect negative sentiment. Although Naïve Bayes demonstrated faster training time, SVM proved to be more effective overall in sentiment classification for mobile application reviews. This study can be beneficial for app developers, digital businesses, and researchers in text mining to better understand user perspectives and improve services based on data insights.

Keyword: Sentiment, IDN App, Naïve Bayes, SVM