



**PERBANDINGAN ALGORITMA NAÏVE BAYES, RANDOM
FOREST DAN SUPPORT VECTOR MACHINE: STUDI KASUS
ANALISA SENTIMEN PENGGUNA DANA**

**TUGAS AKHIR
SKRIPSI**

**REZA ADITYA PRATAMA
41519010077**

**UNIVERSITAS
MERCU BUANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



**PERBANDINGAN ALGORITMA NAÏVE BAYES, RANDOM
FOREST DAN SUPPORT VECTOR MACHINE: STUDI KASUS
ANALISA SENTIMEN PENGGUNA DANA**

**TUGAS AKHIR
SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
sarjana**

REZA ADITYA PRATAMA

41519010077

**UNIVERSITAS
MERCU BUANA**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2026

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Reza aditya Pratama
NIM : 41519010077
Program Studi : Teknik Informatika
Perbandingan Algoritma Naïve Bayes, Random
Judul Skripsi : Forest, dan Support Vector Machine: Studi Kasus
Analisa Sentimen Aplikasi Pengguna Dana.

Menyatakan bahwa Laporan Aplikatif/Tugas Akhir/Jurnal/Media Ilmiah ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 20 Agustus 2026

UNIVERSITAS
MERCU BUANA



Reza Aditya

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : Reza Aditya Pratama
NIM /Student ID Number : 41519010077
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

**“PERBANDINGAN ALGORITMA NAÏVE BAYES, RANDOM FOREST DAN
SUPPORT VECTOR MACHINE: STUDI KASUS ANALISA SENTIMEN
PENGGUNA DANA”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

30 Juli 2026

dengan nilai persentase sebesar :

/with a percentage value of:

23%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer**
Universitas Mercu Buana. */declared to meet standards in accordance with applicable
regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

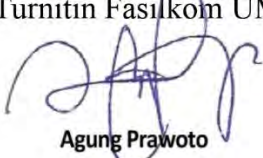
File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1-xIXy55niX6P40mfNu8D7VbSgQ0frIOy/view?usp=drive_link



Jakarta, 30 Juli 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto

Agung Prawoto, S.Kom., B.Sc

NIK : 322970503

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh :

Nama : Reza Aditya Pratama

NIM : 41519010077

Program Studi : Teknik Informatika

Judul Laporan Skripsi : Perbandingan Algoritma Naïve Bayes, Support Vector Machine dan Random Forest : Studi Kasus Analisa Sentimen Pengguna Dana

Disahkan Oleh
Pembimbing



Saruni Dwiasnati, S.T., M.M
(NIDN/NUPTK : 0325128802)

Jakarta, 18 Agustus 2026

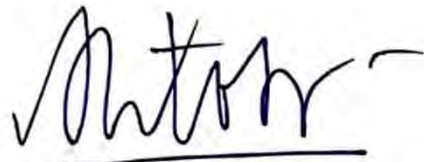
UNIVERSITAS
MERCU BUANA
Mengetahui

Dekan

Ketua Program Studi



Dr. Bambang Jokowi S, Si.MTI
NIDN/NUPTK:0320037002



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK :0225067701

KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT, Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercubuana.
4. Ibu Dosen Saruni Dwiasnati, S.T., M.M. selaku dosen pembimbing MPTI yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan proposal penelitian ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang selalu mensupport dan mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercubuana.
6. Semua teman kuliah yang selalu berbagi informasi dan memberikan dukungan dalam bentuk yang berbeda-beda.

Akhir kata, penulis berharap semoga Tuhan yang Maha Esa membalas kebaikan dan selalu mencurahkan rahmat, hidayah, serta panjang umur kepada kita semua, aamiin. Terima Kasih.

Jakarta, 20 Agustus 2026

Reza Aditya

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR DI REPOSITORI UMB**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Reza Aditya Pratama
NIM : 41519010077
Fakultas/Program Studi : Teknik Informatika
Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes, Random Forest, dan Support Vector Machine: Studi Kasus Analisa Sentimen Aplikasi Pengguna Dana.

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 20 Agustus 2026

Yang menyatakan,



(Reza Aditya Pratama)

PERBANDINGAN ALGORITMA NAÏVE BAYES SUPPORT VECTOR MACHINE, DAN RANDOM FOREST STUDI KASUS: ANALISA SENTIMEN PENGGUNA DANA

REZA ADITYA PRATAMA

ABSTRAK

Ulasan yang ditinggalkan pengguna pada Google Play Store menyajikan sumber data penting dalam mengukur tingkat kepuasan terhadap layanan aplikasi DANA. Kajian ini diangkat guna menguji dan membandingkan performa *Gaussian Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* untuk analisis sentimen berpendekatan TF-IDF. Proses pengumpulan data berhasil menjaring 25.000 ulasan via teknik *scraping*. Keseluruhan data tersebut diproses melalui *preprocessing*, pembobotan teks TF-IDF, serta pembagian rasio data latih-uji sebesar 90:10 sebelum diklasifikasikan. Penilaian presisi model diukur berdasar *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Confusion Matrix*, hingga *Area Under Curve* (AUC). Hasil pengujian mengindikasikan bahwa SVM menghasilkan capaian tertinggi (*accuracy* 89,00% dan AUC 0,938), dilanjutkan oleh Random Forest (*accuracy* 87,98% dan AUC 0,934), serta Gaussian Naïve Bayes (*accuracy* 85,77% dan AUC 0,770). Rangkaian temuan ini mempertegas posisi SVM sebagai pendekatan yang paling akurat dan efektif dalam mengklasifikasikan sentimen pengguna DANA pada eksperimen ini.

Kata kunci: Analisis Sentimen, DANA, TF-IDF, Gaussian Naïve Bayes, Support Vector Machine, Random Forest.

COMPARSION OF NAÏVE BAYES, RANDOM FOREST, AND SUPPORT VECTOR MACHINE A CASE STUDY OF SENTIMENT ANALYSIS OF DANA APPLICATION USERS

REZA ADITYA PRATAMA

ABSTRACT

User reviews on the DANA application available on the Google Play Store can be used to understand users' opinions regarding the quality of the services provided. This study aims to compare the performance of Gaussian Naïve Bayes, Support Vector Machine (SVM), and Random Forest algorithms for sentiment analysis using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting method. This study used 25,000 user reviews collected through web scraping from the Google Play Store. The research process included data preprocessing, TF-IDF weighting, splitting the dataset into training and testing sets with a 90:10 ratio, and sentiment classification using the three algorithms. Model performance was evaluated using Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and Area Under the Curve (AUC). The results showed that Support Vector Machine (SVM) achieved the best performance with an accuracy of 89.00% and an AUC of 0.938, followed by Random Forest with an accuracy of 87.98% and an AUC of 0.934, while Gaussian Naïve Bayes obtained an accuracy of 85.77% and an AUC of 0.770. These findings indicate that the SVM algorithm is the most effective method for sentiment classification of DANA user reviews in this study.

Keywords: Sentiment Analysis, DANA, TF-IDF, Gaussian Naïve Bayes, Support Vector Machine, Random Forest.

DAFTAR ISI

HALAMAN SAMPUL	0
HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORY UMB	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Teori Utama	6
2.1.1 DANA	6
2.1.2 Support Vector Machine	6
2.1.3 Naïve Bayes	7
2.1.4 Random Forest	8
2.1.5 Analisis Sentimen	8
2.2 Teori Pendukung	9
2.2.1 Text Mining	9
2.2.2 Data	9

2.2.3 Python.....	9
2.2.4 TF - IDF	10
2.2.5 Google Colaboratory	10
2.2.6 WordCloud	10
2.3 Penelitian Terdahulu	11
2.4 Gap Penelitian.....	34
2.4.1 Sumarize	34
2.4.2 Synthesize	34
2.4.3 Comparison.....	35
BAB III METODE PENELITIAN	37
3.1 Pendekatan Penelitian	37
3.2 Desain Penelitian	37
3.3 Subjek Penelitian	39
3.4 Instrumen Penelitian	39
3.5 Teknik Pengumpulan Data.....	42
3.6 Analisis Data.....	43
3.7 Prosedur Penelitian	49
3.8 Evaluasi Hasil Penelitian	52
3.9 Timeline Penelitian	53
BAB IV HASIL DAN PEMBAHASAN.....	54
4.1 Dataset	54
4.2 Hasil Pengolahan Data.....	55
4.2.1 Pelabelan Sentimen.....	55
4.2.2 Seleksi Data	56
4.2.3 Cleaning Data	56
4.2.4 Pembagian Data	57
4.3 Preprocessing.....	58
4.3.1 Lemmatization	58
4.3.2 Tokenizing	58
4.3.3 Stopword Removal	59
4.3.4 Stemming.....	60
4.3.5 Pembentukan Data Akhir.....	60
4.4 Visualiasi	61
4.4.1 Wordcloud Positif.....	61
4.4.2 Wordcloud Negatif	62
4.5 Hasil Ekstrasi Fitur TF-IDF	63
4.5.1 Hasil Vocabulary TF-IDF	63
4.5.2 Hasil Pembobotan TF-IDF.....	63
4.6 Hasil Pemodelan	64
4.6.1 Pemodelan Menggunakan Gaussian Naïve Bayes.....	65
4.6.2 Pemodelan Menggunakan Support Vector Machine (SVM)	65

4.6.3	Pemodelan Menggunakan Random Forest	66
4.7	Hasil Evaluasi Model.....	67
4.7.1	Evaluasi Model Gaussian Naïve Bayes	67
4.7.2	Evaluasi Model Support Vector Machine.....	70
4.7.3	Evaluasi Model Random Forest.....	72
BAB V	KESIMPULAN DAN SARAN	76
5.1	Kesimpulan	76
5.2	Saran	77
DAFTAR PUSTAKA.....		78
LAMPIRAN.....		81



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	11
Tabel 3.1 Variable Penelitian.....	39
Tabel 3.2 Simulasi Data	40
Tabel 3.3 Klasifikasi Data.....	41
Tabel 4.1 Atribut Dataset yang digunakan.....	54
Tabel 4.2 Pelabelan Sentimen Berdasarkan Sentimen.....	55
Tabel 4.3 Hasil Pembagian Data.....	57



DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian.....	38
Gambar 3.2 Pengambilan Kata Kunci.....	42
Gambar 3.3 Penyusain Parameter Data	43
Gambar 3.4 Prosedur Penelitian	49
Gambar 3.5 Timeline Penelitian	53
Gambar 4.1 Pelabelan Sentimen	56
Gambar 4.2 Proses Seleksi Data	56
Gambar 4.3 Hasil Cleaning Data	57
Gambar 4.4 Hasil Train - Test Split.....	58
Gambar 4.5 Lemmatization.....	58
Gambar 4.6 Tokenizing	59
Gambar 4.7 Stopword Removal.....	60
Gambar 4.8 Stemming	60
Gambar 4.9 Pembentukan Data Akhir	61
Gambar 4.10 Wordcloud Positif	62
Gambar 4.11 Wordcloud Negatif.....	62
Gambar 4.12 Proses Ekstrasi Fitur TF-IDF	63
Gambar 4.13 Jumlah Vocabulary Tf-Idf.....	63
Gambar 4.14 Hasil Pembobotan TF-IDF	64
Gambar 4.15 Pemodelan Algoritma Gaussian Naïve Bayes.....	65
Gambar 4.16 Pemodelan Algoritma Support Vector Machine	66
Gambar 4.17 Pemodelan Algoritma Random Forest	66
Gambar 4.18 Confusion Matrix Gaussian Naïve Bayes	68
Gambar 4.19 Hasil Klasifikasi Algoritma Gaussian Naïve Bayes.....	68
Gambar 4.20 Hasil ROC dan AUC Algoritma Gaussian Naïve Bayes	69
Gambar 4.21 Confusion Matrix Support Vector Machine.....	70
Gambar 4.22 Hasil Klasifikasi Algoritma Support Vector Machine	71
Gambar 4.23 Hasil ROC AUC Algoritma Support Vector Machine.....	72
Gambar 4.24 Confusion Matrix Random Forest.....	73
Gambar 4.25 Hasil Klasifikasi Algoritma Random Forest.....	74
Gambar 4.26 ROC AUC Algoritma Random Forest.....	75

DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi	81
Lampiran 2 Curriculum Vitae	82
Lampiran 3 Surat Pernyataan & Pengalihan HAKI	83
Lampiran 4 Sertifikat BNSP	85

