



**KLASIFIKASI BERITA HOAKS BERBASIS ENSEMBLE LEARNING
(NAÏVE BAYES, LOGISTIC REGRESSION, DAN RANDOM FOREST)
DENGAN TF-IDF**

TUGAS AKHIR

SKRIPSI



NADHIRA SABINA PUTRI

41822010150

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2026



**KLASIFIKASI BERITA HOAKS BERBASIS ENSEMBLE LEARNING
(NAÏVE BAYES, LOGISTIC REGRESSION, DAN RANDOM FOREST)
DENGAN TF-IDF**

TUGAS AKHIR

SKRIPSI

NADHIRA SABINA PUTRI

41822010150

**UNIVERSITAS
MERCU BUANA**

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2026

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Nadhira Sabina Putri
NIM : 41822010150
Program Studi : Sistem Informasi

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul: **"KLASIFIKASI BERITA HOAKS BERBASIS ENSEMBLE LEARNING (NAÏVE BAYES, LOGISTIC REGRESSION, DAN RANDOM FOREST) DENGAN TF-IDF"** adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila dikemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana semestinya.

Jakarta, 13 Januari 2025



Nadhira Sabina Putri.

HALAMAN SURAT KETERANGAN UJI TURNITIN



072.423.4.03.01

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	: Nadhira Sabina Putri
NIM /Student ID Number	: 41822010150
Program Studi /Study Program	: Sistem Informasi

Dengan Judul Tugas Akhir

/The title:

“KLASIFIKASI BERITA HOAKS BERBASIS ENSEMBLE LEARNING (NAÏVE BAYES, LOGISTIC REGRESSION, DAN RANDOM FOREST) DENGAN TF-IDF”

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

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

16 Januari 2026

dengan nilai persentase sebesar :

/with a percentage value of:

27%

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/1uVNXUIRn_H5BuchY5844fhDSjscvre2d/view?usp=drive_link



Jakarta, 16 Januari 2026
Admin Turnitin Fasilkom UMB

Agung Prawoto

Agung Prawoto, S.Kom., B.Sc

NIK : 322970503

Fakultas Ilmu Komputer
KAMPUS MENARA BHAKTI
Jl. Raya Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650
Telp. 021-5840816 (Hunting), Psw : 5700 Fax. 021-5840813
<http://www.mercubuana.ac.id>, e-mail : fasilkom@mercubuana.ac.id

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Nadhira Sabina Putri
NIM : 41822010150
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Klasifikasi Berita Hoaks Berbasis Ensemble Learning (NB, LR, RF) dengan TF-IDF

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan Diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing,



Misni, S.Kom, M.Kom

NIDN: 0413046802

Jakarta, 04 Februari 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer

Ketua Program Studi

Sistem Informasi



Dr. Bambang Jokonowo, S.Si., M.T.I.

NIDN: 0320037002



Wawan Gunawan, S.Kom., M.T., M.Kom.

NIDN: 0424108104

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, karunia, dan kasih sayang-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Analisis Performa Model Klasifikasi Berita Hoaks Berbasis TF-IDF dengan Pendekatan Ensemble Learning”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Dalam proses penyusunan skripsi ini, penulis menyadari sepenuhnya bahwa penyelesaian karya ini tidak terlepas dari bantuan, bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan ucapan terima kasih yang tulus, penulis menyampaikan penghargaan kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng, selaku Rektor Universitas Mercu Buana, yang telah memberikan kesempatan kepada penulis untuk menempuh pendidikan dan menyelesaikan studi di Universitas Mercu Buana.
2. Bapak Prof Dr. Bambang Jokonowo, S.Si., M.TI., selaku Dekan Fakultas Ilmu Komputer Universitas Mercu Buana, atas dukungan dan fasilitas akademik yang diberikan selama masa perkuliahan.
3. Bapak Wawan Gunawan, S.Kom., M.T., M.Kom., selaku Ketua Program Studi Sistem Informasi, atas arahan dan kebijakan akademik yang sangat membantu penulis dalam proses penyusunan skripsi ini.
4. Bapak Misni, S.Kom., M.Kom., selaku dosen pembimbing Tugas Akhir, yang dengan sabar, penuh perhatian, dan tidak pernah lelah memberikan bimbingan, arahan, serta masukan yang sangat berarti bagi penulis dari awal hingga terselesaikannya skripsi ini.
5. Ibu Kurnia Gusti Ayu, M.Kom selaku dosen pembimbing Akademik yang telah memberikan bimbingan, arahan, serta motivasi kepada penulis selama proses penyusunan karya ini.
6. Seluruh dosen dan seluruh jajaran Tata Usaha Program Studi Sistem Informasi Bapak Ajat, Bapak Efan, Bapak Agung, yang telah banyak membantu penulis dalam proses administrasi akademik, khususnya terkait pengurusan berkas, pendaftaran sidang, serta berbagai informasi penting lainnya. Penulis sangat mengapresiasi kesabaran, keramahan, dan perhatian yang diberikan, sehingga seluruh proses administrasi dapat berjalan dengan lancar.
7. Kedua orang tua tersayang, Mama dan Papa yang selalu menjadi sumber kekuatan

terbesar bagi penulis. Terima kasih atas doa yang tidak pernah putus, kasih sayang yang tak terhingga, serta dukungan tulus yang selalu mengiringi setiap langkah penulis. Secara khusus, penulis menyampaikan rasa terima kasih yang mendalam kepada Mama, atas cinta, kesabaran, dan doa-doa lembut yang selalu menguatkan penulis di setiap proses.

8. Saudara-saudara penulis, Kakak Devinia Tiara, Adik Alverenia Jasmin, Adik Azalia Arasy, dan Adik Abizar Syabani, atas dukungan, perhatian, serta canda tawa sederhana yang sering kali menjadi penghibur dan penyemangat penulis di tengah lelahnya proses penyusunan skripsi.
9. Rekan-rekan perkuliahan, Fira, Thania, Hizkia, Bulan, Nanda, dan Sheela, atas kebersamaan, dukungan, serta semangat yang saling diberikan selama masa perkuliahan dan penyusunan skripsi ini.
10. Windah Basudara, Abun Sungkar dan Callista Arum, yang melalui konten YouTube dan streaming telah menjadi teman begadang penulis di waktu-waktu larut malam. Kehadiran mereka memberikan hiburan, rasa ditemani, serta motivasi tambahan yang membantu penulis tetap bersemangat dalam menyelesaikan skripsi ini.
11. Penulis mengucapkan terima kasih kepada teman-teman Program Studi Sistem Informasi angkatan 2021, 2022, dan 2023, khususnya rekan-rekan di HIMSISFO periode 2023/2024 dan 2024/2025, atas kebersamaan, dukungan moral, serta motivasi yang telah diberikan selama masa perkuliahan.
12. Dan terakhir, untuk diri penulis sendiri, terima kasih karena telah bertahan, berusaha, dan tidak menyerah meskipun proses ini penuh tantangan. Setiap langkah kecil, lelah, dan keraguan akhirnya bermuara pada satu pencapaian yang patut dibanggakan.

Sebagai penutup, penulis berharap Tuhan Yang Maha Esa senantiasa melimpahkan balasan atas segala kebaikan semua pihak yang telah memberikan bantuan. Semoga Laporan Skripsi ini dapat memberikan kontribusi positif serta menambah wawasan dalam pengembangan ilmu pengetahuan.

Jakarta, 13 Januari 2026

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

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

Nama	: Nadhira Sabina Putri
NIM	: 41822010150
Fakultas/Program Studi	: Ilmu Komputer/ Sistem Informasi
Judul Tugas Akhir	: Klasifikasi Berita Hoaks Berbasis Ensemble Learning (NB, LR, RF) dengan TF-IDF

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/formatkan, 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 sebenar-benarnya.

Jakarta, 13 Januari 2026

Yang menandatangani



Nadhira Sabina Putri

KLASIFIKASI BERITA HOAKS BERBASIS ENSEMBLE LEARNING (NAÏVE BAYES, LOGISTIC REGRESSION, DAN RANDOM FOREST) DENGAN TF-IDF

NADHIRA SABINA PUTRI

ABSTRAK

Penelitian ini bertujuan untuk merancang sistem klasifikasi berita hoaks berbasis teks dengan menerapkan metode Ensemble Learning yang menggabungkan tiga algoritma machine learning, yaitu Naïve Bayes, Logistic Regression, dan Random Forest, serta menggunakan fitur *Term Frequency–Inverse Document Frequency* (TF-IDF). Penelitian ini dilatarbelakangi oleh tingginya peredaran hoaks di Indonesia yang diperkirakan menjangkau sekitar 30%–60% masyarakat, sementara hanya 21%–36% yang mampu mengenali informasi palsu. Penyebaran hoaks yang banyak terjadi melalui media sosial, aplikasi pesan, dan situs web menunjukkan pentingnya pemanfaatan teknologi untuk membantu masyarakat dalam melakukan verifikasi informasi secara lebih cepat dan tepat. Dataset yang digunakan berjumlah 5.446 berita berbahasa Indonesia, yang terdiri dari 2.750 berita hoaks dari TurnBackHoax dan 2.696 berita valid yang diperoleh dari Detik.com, Kompas, Liputan6, serta CNN Indonesia. Proses penelitian diawali dengan pengumpulan data melalui web scraping, kemudian dilanjutkan dengan tahap pra-pemrosesan teks yang mencakup pembersihan data, penggabungan judul dan isi berita, case folding, penghapusan URL dan tanda baca, serta tokenisasi. Selanjutnya, ekstraksi fitur dilakukan menggunakan TF-IDF dengan konfigurasi unigram, kemudian dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratified split guna menjaga keseimbangan distribusi kelas. Sistem bekerja dengan mengonversi teks berita menjadi vektor TF-IDF yang diproses secara bersamaan oleh ketiga algoritma klasifikasi, lalu hasil prediksi dari masing-masing model digabungkan melalui mekanisme hard voting untuk menentukan label akhir berupa hoaks atau valid. Hasil pengujian menunjukkan bahwa model ensemble memberikan kinerja terbaik dengan akurasi 96,69%, precision 96,15%, recall 97,22%, dan F1-score 96,68%, sehingga lebih unggul dibandingkan model tunggal. Penelitian ini menawarkan solusi berupa sistem klasifikasi otomatis berbasis Ensemble Learning yang mampu mendeteksi berita hoaks secara lebih akurat dan konsisten, sehingga diharapkan dapat membantu masyarakat dalam memverifikasi informasi serta mendukung peningkatan literasi digital.

Kata kunci: *Berita Hoaks, Klasifikasi Teks, Ensemble Learning, TF-IDF, Machine Learning.*

HOAX NEWS CLASSIFICATION USING ENSEMBLE LEARNING METHODS (NAÏVE BAYES, LOGISTIC REGRESSION, AND RANDOM FOREST) WITH TF- IDF

NADHIRA SABINA PUTRI

ABSTRACT

This study aims to develop a text-based fake news classification system by applying the Ensemble Learning method, which integrates three machine learning algorithms: Naïve Bayes, Logistic Regression, and Random Forest, utilizing the Term Frequency–Inverse Document Frequency (TF-IDF) feature representation. The study is motivated by the high circulation of fake news in Indonesia, estimated to affect approximately 30%–60% of the population, while only about 21%–36% are capable of identifying false information. The widespread distribution of fake news through social media, messaging applications, and websites highlights the need for technological solutions that can assist the public in verifying information more quickly and accurately. The dataset used in this research consists of 5,446 Indonesian-language news articles, including 2,750 fake news articles collected from TurnBackHoax and 2,696 legitimate news articles obtained from Detik.com, Kompas, Liputan6, and CNN Indonesia. The research process began with data collection through web scraping, followed by text preprocessing stages such as data cleaning, merging news titles and content, case folding, removing URLs and punctuation, and tokenization. Feature extraction was then performed using TF-IDF with a unigram configuration, after which the dataset was divided into training and testing sets with an 80:20 ratio using stratified split to maintain balanced class distribution. The system converts news text into TF-IDF vectors that are processed simultaneously by the three classification algorithms, and the predictions from each model are combined using a hard voting mechanism to determine the final label as either fake or valid news. Evaluation results indicate that the ensemble model achieved the best performance with an accuracy of 96.69%, precision of 96.15%, recall of 97.22%, and an F1-score of 96.68%, outperforming the individual models. This study proposes a solution in the form of an automated fake news classification system based on Ensemble Learning that is capable of detecting false information more accurately and consistently. The system is expected to support the public in verifying digital information and contribute to improving digital literacy.

Keywords: Fake News, Text Classification, Ensemble Learning, TF-IDF, Machine Learning.

DAFTAR ISI

HALAMAN PERNYATAAN KARYA SENDIRI.....	iii
HALAMAN SURAT KETERANGAN UJI TURNITIN	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR.....	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB	viii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian.....	4
1.4.1 Manfaat Teoritis.....	4
1.4.2 Manfaat Praktis	4
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Teori Utama	6
2.1.1 Pengertian Berita dan Berita Hoaks	6
2.1.2 Karakteristik dan Jenis Hoaks	6
2.1.3 Dampak Penyebaran Berita Hoaks.....	7

2.2	Text Mining dan Natural Language Processing untuk Klasifikasi Teks.....	7
2.2.1	Text Mining dan Klasifikasi Teks	7
2.2.2	Tahapan Umum Pengolahan Teks (pipeline).....	7
2.2.3	Representasi Teks (bag-of-words, n-gram).....	9
2.3	Ekstraksi Fitur TF-IDF	9
2.3.1	Konsep <i>Term Frequency</i> (TF).....	9
2.3.2	Konsep <i>Inverse Document Frequency</i> (IDF)	10
2.3.3	TF-IDF sebagai Pembobotan Kata untuk Klasifikasi	10
2.3.4	Konfigurasi TF-IDF (unigram, bigram, kombinasi n-gram; paraeter dasar) 12	
2.4	Algoritma Klasifikasi yang Digunakan.....	13
2.4.1	Naïve Bayes.....	13
2.4.2	Logistic Regression	14
2.4.3	Random Forest	15
2.5	Ensemble Learning	16
2.5.1	Konsep Ensemble Learning	16
2.5.2	Metode Ensemble yang Relevan	17
2.5.3	Perbandingan model tunggal dan ensemble.....	18
2.6	Evaluasi Kinerja Model Klasifikasi	19
2.7	Penelitian Terdahulu.....	21
2.8	Teori Pendukung	47
2.8.1	Klasifikasi Teks.....	47
2.8.2	Representasi Teks Menggunakan TF-IDF	47
2.8.3	Algoritma Klasifikasi Machine Learning	48
2.8.4	Ensemble Learning Berbasis Voting	49
2.8.5	Evaluasi Kinerja Model Klasifikasi	49
BAB III METODE PENELITIAN		50

3.1	Jenis Penelitian.....	50
3.2	Tahapan Penelitian	51
3.3	Instrumen Penelitian	53
3.4	Subjek Penelitian	55
BAB IV PEMBAHASAN		56
4.1	Persiapan Data	56
4.2	Pre-Processing Data	57
4.2.1	Pembersihan Teks pada Judul Berita	58
4.2.2	Penggabungan Judul dan Isi Artikel.....	58
4.2.3	Penghapusan URL, Angka, dan Tanda Baca	59
4.2.4	Penghapusan Pola Portal Berita.....	60
4.2.5	Tokenisasi.....	60
4.3	Pembagian Data (Split Data)	61
4.3.1	Perbandingan Pembagian Data	62
4.3.2	Stratified Split untuk Menjaga Keseimbangan Kelas	62
4.3.3	Penggunaan Random State untuk Reprodusibilitas	63
4.4	Ekstraksi Fitur menggunakan TF-IDF	63
4.4.1	Pengertian TF-IDF.....	63
4.4.2	Konfigurasi TfidfVectorizer	64
4.4.3	Proses Fit dan Transform.....	65
4.4.4	Hasil Ekstraksi Fitur	65
4.5	Proses Training dan Testing Model	65
4.5.1	Model Naïve Bayes.....	66
4.5.2	Model Random Forest	66
4.5.3	Model Logistic Regression.....	67
4.6	Ensemble Learning dengan Hard Voting	68
4.6.1	Konsep Hard Voting	68

4.6.2	Implementasi Ensemble Learning.....	69
4.7	Evaluasi Sistem Klasifikasi	70
4.7.1	Metrik Evaluasi.....	70
4.7.2	Fungsi Evaluasi Model	71
4.7.3	Hasil Evaluasi Model Individual	71
4.7.4	Hasil Evaluasi Ensemble Learning.....	74
4.7.5	Perbandingan Performa Model	75
4.8	Hasil Prediksi Sistem Klasifikasi	76
4.8.1	Keterbatasan Sistem	78
BAB V KESIMPULAN DAN SARAN		80
5.1	Kesimpulan.....	80
5.2	Saran	82
DAFTAR PUSTAKA		84
LAMPIRAN.....		86



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	21
Tabel 3.1 Instrumen Penelitian Perangkat Lunak (Software)	53
Tabel 3.2 Instrumen Penelitian Perangkat Keras (Hardware)	55
Tabel 4.1 Tahap Pembersihan Teks pada Judul Berita	58
Tabel 4.2 Tahap Penggabungan Judul dan Isi	59
Tabel 4.3 Penghapusan URL, Angka, dan Tanda Baca	60
Tabel 4.4 Tahap Penghapusan Pola Portal Berita	60
Tabel 4.5 Tahap Tokenisasi	61



DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian	51
Gambar 4. 1 Dataset	56
Gambar 4. 2 Confusion Matrix dari model Naive Bayes	72
Gambar 4. 3 Confusion Matrix dari model Logistic Regression	73
Gambar 4. 4 Confusion Matrix dari model Random Forest.....	74
Gambar 4. 5 Confusion Matrix dari model Ensemble Learning.....	75
Gambar 4. 6 Hasil Evaluasi Performa Model	76
Gambar 4. 7 Perbandingan Akurasi Model	76
Gambar 4. 8 Contoh Hasil Prediksi Sistem Klasifikasi Berita	77



DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi Bimbingan	86
Lampiran 2 Curriculum Vitae.....	87
Lampiran 3 Surat Keterangan BNSP	88
Lampiran 4 Hasil Cek Turnitin	89

