STUDI KOMPARASI METODE SUPPORT VECTOR MACHINE BERBASIS TF-IDF DAN FINE-TUNING INDOBERT UNTUK DETEKSI BERITA HOAKS BERBAHASA INDONESIA

TUGAS AKHIR

SKRIPSI

JUNANDA ABIMANYU

41522110057

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2026

lib.mercubuana.ac.id

STUDI KOMPARASI METODE SUPPORT VECTOR MACHINE BERBASIS TF-IDF DAN FINE-TUNING INDOBERT UNTUK DETEKSI BERITA HOAKS BERBAHASA INDONESIA

TUGAS AKHIR

SKRIPSI

JUNANDA ABIMANYU

41522110057

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2026

lib.mercubuana.ac.id

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama	:	JUNANDA ABIMANYU
NIM	:	41522110057
Fakultas/Program Studi	:	Fakultas Ilmu Komputer/Teknik Informatika
Judul Tugas Akhir	:	Studi Komparasi Metode Support Vector Machine Berbasis Tf-Idf dan Fine-Tuning Indobert untuk Deteksi Berita Hoaks Berbahasa Indonesia

Menyatakan bahwa Laporan Tugas Akhir 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 ini saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 18 Agustus 2026



Junanda Abimanyu

UNIVERSITAS
MERCU BUANA

lib.mercubuana.ac.id

UNIVERSITAS
MERCU BUANA
AKREDITASI UNGGUL

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	: Junanda Abimanyu
NIM /Student ID Number	: 41522110057
Program Studi /Study Program	: Teknik Informatika

Dengan Judul Tugas Akhir
/The title:

"STUDI KOMPARASI METODE SUPPORT VECTOR MACHINE BERBASIS TF-IDF DAN FINE-TUNING INDOBERT UNTUK DETEKSI BERITA HOAKS BERBAHASA INDONESIA"

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:
/Similarity checks have been carried out with the Turnitin system on the date:

28 Agustus 2026

dengan nilai persentase sebesar :
/with a percentage value of:

24%

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/1TVu1I4wmGhcbhQJ-lPuUyvd4kv2tqAPY/view?usp=drive_link

Jakarta, 28 Agustus 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

lib.mercubuana.ac.id

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama	:	JUNANDA ABIMANYU
NIM	:	4152110057
Fakultas/Program Studi	:	Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir	:	Studi Komparasi Metode Support Vector Machine Berbasis TF-IDF dan Fine-Tuning IndoBERT untuk Deteksi Berita Hoaks Berbahasa Indonesia

Telah berhasil dipertahankan pada sidang tanggal 18 Agustus 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing

Ilham Nugraha, S.Kom ,M.Sc

NIDN/NUPTK: 307098904

Jakarta, 18 Agustus 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer	Ketua Program Studi Teknik Informatika
Dr. Bambang Jokonowo, S.Si., M.TI.	Dr. Hadi Santoso, S.Kom., M.Kom.
NIDN/NUPTK: 0320037002	NIDN/NUPTK: 0225067701

UNIVERSITAS MERCU BUANA

lib.mercubuana.ac.id

KATA PENGANTAR

Puji syukur kehadirat Tuhan yang Maha Esa, atas segala rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika, Universitas Mercu Buana.

Penulis menyadari bahwa penyusunan Tugas Akhir ini masih jauh dari sempurna, karena kesempurnaan sejatinya hanya milik Tuhan yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Serta berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis 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. Bapak Ilham Nugraha, S.kom, M.Sc. selaku dosen pembimbing MPTI dan Tugas Akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan laporan Tugas Akhir ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang selalu mensuport dan mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercubuana..
6. Semua teman kuliah saya 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, 18 Agustus 2026



Junanda Abimanyu

lib.mercubuana.ac.id

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

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

Nama	:	JUNANDA ABIMANYU
NIM	:	41522110057
Fakultas/Program Studi	:	Fakultas Ilmu Komputer/Teknik Informatika
Judul Tugas Akhir	:	Studi Komparasi Metode Support Vector Machine Berbasis Tf-Idf dan Fine-Tuning Indobert untuk Deteksi Berita Hoaks Berbahasa Indonesia

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 diatas 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.

UNIVERSITAS MERCU BUANA

Jakarta, 18 Agustus 2026

Yang menyatakan,



Junanda Abimanyu

lib.mercubuana.ac.id

STUDI KOMPARASI METODE SUPPORT VECTOR MACHINE BERBASIS TF-IDF DAN FINE-TUNING INDOBERT UNTUK DETEKSI BERITA HOAKS BERBAHASA INDONESIA

JUNANDA ABIMANYU

ABSTRAK

Perkembangan teknologi informasi dan media digital membuat berita palsu bisa menyebar dengan cepat, sehingga diperlukan cara yang tepat untuk membedakan antara berita palsu dan berita yang benar. Penelitian ini bertujuan untuk membandingkan kemampuan Support Vector Machine (SVM) yang menggunakan TF-IDF serta SVM berbasis FineTuning IndoBERT dalam mengenali berita palsu yang ditulis dalam bahasa Indonesia.

Penelitian ini menggunakan metode penelitian kuantitatif eksperimental dengan memanfaatkan dataset berisi berita palsu dari Indonesia yang berasal dari Kaggle. Tahapan penelitian mencakup proses persiapan teks, membagi data menjadi bagian latihan, validasi, dan uji, melatih dua model, serta mengevaluasi hasil menggunakan metrik seperti Accuracy, Precision, Recall, F1-Score, Learning Curve, dan Confusion Matrix.

Penelitian menunjukkan bahwa Fine-Tuning IndoBERT memberikan hasil yang lebih baik dibandingkan metode SVM yang menggunakan TF-IDF, terlihat dari nilai akurasi, presisi, recall, dan F1-score yang lebih tinggi serta jumlah kesalahan klasifikasi yang lebih sedikit. Sementara itu, SVM yang menggunakan TF-IDF memiliki keuntungan dalam hal efisiensi pengolahan data dan kestabilan selama proses pelatihan. Oleh karena itu, fine-tuning IndoBERT lebih efektif untuk mendeteksi berita hoaks dalam bahasa Indonesia, sementara SVM berbasis TF-IDF tetap bisa menjadi pilihan yang baik jika sumber daya komputasi terbatas.

Kata Kunci: Deteksi Berita Hoaks, Support Vector Machine, TF-IDF, Fine-Tuning IndoBERT, Natural Language Processing, Klasifikasi Teks.

lib.mercubuana.ac.id

A COMPARATIVE STUDY OF TF-IDF-BASED SUPPORT VECTOR MACHINE METHODS AND INDOBERT FINE-TUNING FOR DETECTING INDONESIAN-LANGUAGE HOAX NEWS

JUNANDA ABIMANYU

ABSTRACT

The development of information technology and digital media has allowed fake news to spread rapidly, necessitating a precise method for distinguishing between fake and genuine news. This study aims to compare the performance of a Support Vector Machine (SVM) using TF-IDF and an SVM based on FineTuning IndoBERT in recognizing fake news written in Indonesian.

This study employed an experimental quantitative research method utilizing a dataset containing fake news from Indonesia sourced from Kaggle. The research stages included text preparation, dividing the data into training, validation, and testing sections, training two models, and evaluating the results using metrics such as Accuracy, Precision, Recall, F1-Score, Learning Curve, and Confusion Matrix.

The study showed that FineTuning IndoBERT provided better results than the SVM method using TF-IDF, as evidenced by higher accuracy, precision, recall, and F1-score values, as well as fewer misclassifications. Meanwhile, the SVM using TF-IDF has advantages in terms of data processing efficiency and stability during the training process. Therefore, fine-tuning IndoBERT is more effective for detecting hoax news in Indonesian, while TF-IDF-based SVM can still be a good choice if computing resources are limited.

UNIVERSITAS MERCU BUANA

Keywords: Hoax Detection, Support Vector Machine, TF-IDF, IndoBERT Fine-Tuning, Natural Language Processing, Text Classification.

lib.mercubuana.ac.id

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 REPOSITORI UMB vi
ABSTRAK vii
ABSTRACT viii
DAFTAR ISI ix
DAFTAR GAMBAR xiii
DAFTAR TABEL xiv
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Perumusan Masalah 2
1.3 Tujuan Penelitian 2
1.4 Batasan Masalah 3
1.5 Sistematika Penulisan 3
BAB II TINJAUAN PUSTAKA 5
2.1 Penelitian Terdahulu 5
2.2 Definisi Hoaks 10
2.3 Jenis-Jenis Berita Hoaks 10
2.4 Ciri – Ciri Berita Hoaks 13
2.5 Natural Language Processing (NLP) 13

lib.mercubuana.ac.id

2.6 IndoBERT 14
2.7 Representasi Teks Menggunakan TF-IDF 15
2.8 Support Vector Machine (SVM) 16
2.9 Evaluasi Model 16
2.10 Gap Penelitian 19
BAB III METODE PENELITIAN 21
3.1 Pendekatan Penelitian 21
3.2 Desain Penelitian 21
3.2.1 Pengumpulan Dataset Berita Hoaks 22
3.2.2 Preprocessing Teks 22
3.2.3 Evaluasi Model 22
3.2.4 Analisis Hasil 23
3.2.5 Diagram Alur Penelitian 24
3.3 Subjek Penelitian 25
3.4 Instrumen Penelitian 26
3.4.1 Dataset Teks Berita 26
3.4.2 Model IndoBERT dan SVM + TF-IDF 26
3.4.3 Bahasa Pemrograman Python 26
3.4.4 Library Pendukung 26
3.5 Tahapan Preprocessing Data 27
3.5.1 Seleksi Atribut & Penggabungan Dataset Hoaks dan Fakta 27
3.5.2 Case Folding 28
3.5.3 Cleaning 28
3.5.4 Penghapusan Kata dan Frasa Bias 29
3.5.5 Penghapusan Data Duplikat 31
3.5.6 Pembatasan Panjang Dokumen 31
3.5.7 Validasi Hasil Preprocessing 32

lib.mercubuana.ac.id

3.6 Pembagian Dataset 34
3.7 Pemodelan dan Parameter Eksperimen 35
3.7.1 Support Vector Machine (SVM) Berbasis TF-IDF 35
3.7.1.1 Ekstraksi Fitur TF-IDF 35
3.7.1.2 Klasifikasi Support Vector Machine (SVM) 36
3.7.1.3 Validasi Silang (Cross Validation) 37
3.7.2 Fine-Tuning IndoBERT 38
3.7.2.1 Arsitektur Model Pre-trained 38
3.7.2.2 Tokenisasi IndoBERT 38
3.7.2.3 Hyperparameter Fine-Tuning 39
3.8 Analisis Data 40
3.8.1 Accuracy (Akurasi) 40
3.8.2 Precision (Presisi) 40
3.8.3 Recall 40
3.8.4 F1-Score 40
BAB IV EKSPERIMEN DAN HASIL 41
4.1 Eksperimen 41
4.2 Implementasi SVM 42
4.2.1 Training 43
4.2.2 Hyperparameter 43
4.2.3 Hasil 44
4.2.3.1 Training Accuracy & Validation Accuracy (Learning Curve) 44
4.2.3.2 Training Loss & Validation Loss (Learning Curve) 45
4.2.4 Evaluasi Metrik 46
4.3 Implementasi IndoBERT 48
4.3.1 Training 49
4.3.2 Hyperparameter 49

UNIVERSITAS MERCU BUANA

lib.mercubuana.ac.id

4.3.3 Hasil 49
4.3.3.1 Training Accuracy & Validation Accuracy (Learning Curve) 49
4.3.3.2 Training Loss & Validation Loss (Learning Curve) 50
4.3.4 Evaluasi Metrik 52
4.4 Analisis Komparasi 54
4.4.1 Komparasi Training 54
4.4.2 Komparasi Evaluasi dan skor F1 55
4.4.3 Komparasi Learning Curve 56
4.4.3.1 Komparasi Learning Curve Accuracy SVM + TF-IDF dan Fine-tuning IndoBERT 57
4.4.3.2 Komparasi Learning Curve Loss SVM + TF-IDF dan Fine-tuning IndoBERT 58
4.4.4 Komparasi Confussion Matriks 59
4.4.5 Analisis 60
BAB V KESIMPULAN & SARAN 63
5.1 KESIMPULAN 63
5.2 SARAN 63
DAFTAR PUSTAKA 65
LAMPIRAN 67

UNIVERSITAS MERCU BUANA

lib.mercubuana.ac.id

DAFTAR GAMBAR

Gambar 2.1 Learning Curve 18
Gambar 2.2 Confusion Matriks 19
Gambar 3.3 Diagram Alur Penelitian 24
Gambar 3.4 Output Pembatasan Panjang Kata 32
Gambar 4.5 Learning Curve Accuracy SVM + TF-IDF 44
Gambar 4.6 Learning Curve Loss SVM + TF-IDF 45
Gambar 4.7 Matriks Acuracy, Precision, Recall, & F1-Score SVM + TF-IDF 46
Gambar 4.8 Confusion Matrikx SVM + TF-IDF 47
Gambar 4.9 Iterasi Eksperimen SVM + TF-IDF 47
Gambar 4.10 Learning Curve Accuracy Finetuning IndoBERT 50
Gambar 4.11 Learning Curve Loss Finetuning IndoBERT 52
Gambar 4.12 Matriks Acuracy, Precision, Recall, & F1-Score Finetuning IndoBERT 52
Gambar 4.13 Iterasi 10 Epoch Finetunig IndoBERT 53
Gambar 4.14 Confusion Matrix Finetuning IndoBERT 53
Gambar 4.15 Grafik Evaluasi Metriks Komparasi Model Finetuning IndoBERT & SVM + TF-IDF 56
Gambar 4.16 Komparasi Learning Curve Accuracy Model Finetuning IndoBERT & SVM + TF-IDF 57
Gambar 4.17 Komparasi Learning Curve Loss Model Finetuning IndoBERT & SVM + TF-IDF 58
Gambar 4.18 Komparasi Confusion Metrix Model Finetuning IndoBERT & SVM + TF-IDF 60

UNIVERSITAS MERCU BUANA

lib.mercubuana.ac.id

DAFTAR TABEL

Tabel 2.1 Daftar Jurnal Penelitian Terdahulu 5
Tabel 3.2 Tabel Deskripsi Atribut 25
Tabel 3.3 Dataset 27
Tabel 3.4 Dataset Clean 28
Tabel 3.5 Tabel Split Dataset 34
Tabel 3.6 Tabel Deskripsi Parameter Atribut SVM + TF-IDF 36
Tabel 3.7 Tabel Deskripsi Parameter Atribut SVM + TF-IDF (2) 36
Tabel 3.8 Tabel Deskripsi Parameter Atribut Finetuning IndoBERT 39
Tabel 4.9 Tabel Komparasi Evaluasi Metriks Kedua Model 55

UNIVERSITAS
MERCU BUANA

lib.mercubuana.ac.id