Jurnal Ilmiah
Teknologi Informasi dan Robotika

ISSN (Online): 2686-4339
Volume 8 Nomor 1, Juni 2026
Halaman: 10-22

Deteksi Hoaks pada Twitter Menggunakan Fitur Linguistik dan Ensemble Machine Learning

*Septiyawan Rosetya Wardhana, Gusti Eka Yuliastuti, Dian Puspita Hapsari

Institut Teknologi Adhi Tama Surabaya, Indonesia

Artikel Histori:
Disubmit: Mei 2026
Diterima: Juni 2026
Diterbitkan: juni 2026

DOI
10.33005/jifti.v8i1.218

ABSTRAK

Penyebaran hoaks di media sosial Twitter berbahasa Indonesia telah menjadi permasalahan serius yang berdampak pada opini publik dan stabilitas sosial. Penelitian ini mengusulkan metode deteksi hoaks berbasis kombinasi fitur linguistik spesifik-Indonesia dan ensemble machine learning. Fitur linguistik yang diekstrak mencakup: (1) pola leksikal hiperbola, urgensi, dan konspirasi; (2) karakteristik struktural teks; (3) fitur stilistika dan kompleksitas kalimat; serta (4) fitur TF-IDF dan n-gram karakter. Model ensemble yang diusulkan menggabungkan Random Forest, Gradient Boosting, dan Support Vector Machine melalui mekanisme soft voting. Eksperimen dilakukan pada dataset IndoFakeNews yang berisi 5.548 pasang berita asli dan hoaks. Hasil evaluasi menunjukkan bahwa metode yang diusulkan mencapai akurasi 89,3%, precision 88,7%, recall 90,1%, dan F1-score 89,4%, melampaui baseline IndoBERT fine-tuned sebesar 1,2% pada F1-score dengan kecepatan inferensi 8,3 kali lebih cepat. Hasil ini menunjukkan bahwa kombinasi fitur linguistik berbasis pengetahuan domain dengan ensemble klasik mampu menyaingi model deep learning pada tugas deteksi hoaks Bahasa Indonesia.

Kata Kunci: Deteksi Hoaks, Ensemble Learning, Fitur Linguistik, Media Sosial, NLP Bahasa Indonesia

How to Cite:
Septiyawan Rosetya Wardhana, Gusti Eka Yuliastuti, Dian Puspita Hapsari. (2026). Deteksi Hoaks pada Twitter Menggunakan Fitur Linguistik dan Ensemble Machine LEarning. *Jurnal Ilmiah Teknologi Informasi dan Robotika, 8*(1), 10-22. https://doi.org/10.33005/jifti.v8i1.218.

***Corresponding Author:**
Email : rossywardhana@itats.ac.id
Alamat : Jl Tambak Medokan Ayu Gg XI C1, Rungkut, Surabaya, Jawa Timur, 60295

PENDAHULUAN

Media sosial telah menjadi salah satu sumber informasi utama bagi masyarakat Indonesia (Pasek et al., 2022). Berdasarkan laporan Digital 2024 Indonesia, terdapat 139 juta pengguna aktif media sosial di Indonesia, dengan *platform* Twitter/X menempati posisi ketiga sebagai *platform* paling aktif digunakan (Jeong et al., 2024). Aksesibilitas informasi yang tinggi ini, sayangnya, juga diiringi oleh meningkatnya penyebaran informasi yang tidak akurat atau hoaks (Muhabatin et al., 2021).

Fenomena hoaks di Indonesia bukan sekadar masalah teknis—ia memiliki dampak sosial yang nyata (Rahmawati et al., 2023)(Agma, 2025). Kominfo mencatat lebih dari 2.000 konten hoaks yang berhasil diidentifikasi sepanjang tahun 2023, dengan kategori kesehatan, politik, dan bencana alam sebagai topik yang paling rentan (Aisyah & Yasmin, 2022)(Suherman et al., 2024). Hoaks telah terbukti memengaruhi keputusan vaksinasi masyarakat selama pandemi COVID-19, memicu kekerasan berbasis provokasi, dan mendistorsi persepsi publik terhadap kandidat dalam kontestasi demokrasi (Ohorella, 2023)(Noguera-vivo et al., 2023).

Hoaks dalam konteks media sosial didefinisikan sebagai informasi yang sengaja dibuat atau disebarkan untuk menyesatkan audiens, terlepas dari apakah pembuatnya mengetahui ketidakakuratan informasi tersebut (Lazer et al., 2018)(Halawa & Lase, 2022). Berbeda dari kesalahan informasi (*misinformation*) yang dapat terjadi tanpa niat buruk, hoaks (disinformation) secara inheren mengandung unsur kesengajaan (Iskandar et al., 2024).

Upaya deteksi hoaks secara otomatis menggunakan *Natural Language Processing* (NLP) telah banyak dieksplorasi, namun sebagian besar penelitian berfokus pada bahasa Inggris (Cendana & Permana, 2019). Tantangan khusus Bahasa Indonesia—termasuk kekayaan bahasa gaul, fenomena *code-switching*, dan keterbatasan sumber daya linguistik—membuat model berbahasa Inggris tidak dapat langsung diterapkan pada konteks Indonesia.

Beberapa pendekatan yang telah diusulkan untuk Bahasa Indonesia antara lain: analisis berbasis leksikon (Febriyanty et al., 2023), model LSTM (Rianansyah et al., 2025), serta fine-tuning IndoBERT (Koto & Baldwin, 2020). Meskipun model berbasis Transformer seperti IndoBERT memberikan akurasi tertinggi, kompleksitas komputasinya yang tinggi dan kebutuhan GPU yang besar menjadi hambatan adopsi di lingkungan dengan sumber daya terbatas, seperti institusi pendidikan atau UMKM.

Penggunaan ensemble machine learning untuk deteksi hoaks dan berita palsu telah menjadi arah penelitian yang aktif karena kemampuannya menggabungkan keunggulan beberapa pengklasifikasi sekaligus menekan varians prediksi. Hakak et al. (2021) mengusulkan kerangka ensemble yang memadukan Decision Tree, Random Forest, dan Extra Trees untuk klasifikasi berita palsu dan melaporkan peningkatan akurasi dibandingkan model tunggal. Kaliyar et al. (2020) menunjukkan bahwa pendekatan berbasis gradient boosting efektif menangkap pola fitur non-linear pada deteksi berita palsu. Untuk konteks Bahasa Indonesia, Setiawan et al. (2022) mengombinasikan fitur TF-IDF dengan ensemble voting dan memperoleh performa yang kompetitif terhadap model tunggal. Tinjauan terhadap penelitian-penelitian tersebut memperlihatkan dua celah yang menjadi motivasi penelitian ini: (1) sebagian besar pendekatan ensemble mengandalkan fitur statistik tingkat permukaan (TF-IDF/n-gram) tanpa mengintegrasikan pengetahuan domain spesifik mengenai retorika hoaks Bahasa Indonesia; dan (2) aspek efisiensi

komputasi untuk deployment praktis jarang dievaluasi secara eksplisit. Penelitian ini menutup kedua celah tersebut dengan memadukan fitur linguistik berbasis pengetahuan domain dan mekanisme soft voting berbobot, sekaligus mengukur trade-off antara akurasi dan kecepatan inferensi.

Penelitian ini mengusulkan pendekatan alternatif yang menggabungkan fitur linguistik berbasis pengetahuan domain (*domain knowledge*) dengan *ensemble machine learning* klasik. Ensemble learning merupakan paradigma yang menggabungkan prediksi beberapa model dasar untuk menghasilkan prediksi yang lebih akurat dan robust dibandingkan model tunggal (Yaghoubi et al., 2024)(V et al., 2024). Tiga strategi ensemble utama adalah bagging (Bootstrap Aggregating), boosting, dan stacking/voting.

Pemanfaatan fitur linguistik berbasis pengetahuan domain untuk deteksi hoaks berangkat dari observasi bahwa konten hoaks memiliki ciri kebahasaan yang dapat dibedakan secara sistematis dari berita faktual. Hoaks cenderung dirancang untuk memicu reaksi emosional dan mendorong penyebaran cepat, sehingga lazim menggunakan diksi hiperbola, seruan urgensi, klaim sensasional, serta narasi konspiratif. Karakteristik permukaan semacam ini meninggalkan jejak terukur pada teks, misalnya kepadatan kata bermuatan emosi, rasio penggunaan huruf kapital, frekuensi tanda seru, dan pola keterbacaan kalimat. Dengan mengkodifikasikan pola-pola tersebut menjadi fitur numerik yang dirancang secara eksplisit berdasarkan pengetahuan pakar mengenai retorika hoaks Bahasa Indonesia, model klasifikasi dapat memanfaatkan sinyal diskriminatif yang interpretable tanpa memerlukan representasi laten berdimensi tinggi seperti pada model deep learning. Pendekatan berbasis pengetahuan domain ini memberikan dua keuntungan utama: transparansi (kontribusi setiap fitur dapat ditelusuri) dan efisiensi komputasi (ekstraksi fitur dilakukan melalui operasi linguistik ringan), yang menjadikannya alternatif praktis untuk lingkungan dengan sumber daya terbatas.

Soft voting adalah metode yang menggabungkan probabilitas kelas dari beberapa model dengan rata-rata berbobot (Joses et al., 2024)(Rizka & Sari, 2025). Keunggulannya dibandingkan *hard voting* adalah kemampuannya memanfaatkan informasi confidence level setiap model, sehingga prediksi yang lebih yakin mendapatkan pengaruh lebih besar (Wu et al., 2008).

Kontribusi utama penelitian ini adalah: (1) Desain set fitur linguistik yang secara spesifik mempertimbangkan karakteristik hoaks Bahasa Indonesia, mencakup 47 fitur yang dikategorikan dalam empat kelompok; (2) Kerangka ensemble yang menggabungkan tiga algoritma klasifikasi dengan mekanisme *soft voting* berbobot; dan (3) Evaluasi komprehensif yang tidak hanya mengukur akurasi tetapi juga efisiensi komputasi, memberikan perspektif praktis untuk *deployment* di lingkungan dengan sumber daya terbatas.

METODE PENELITIAN

Dataset

Penelitian ini menggunakan dataset IndoFakeNews yang dikompilasi dari tiga sumber: (1) portal cek fakta Hoax-Slayer.id dan TurnBackHoax.id sebagai sumber berita palsu; (2) portal berita terverifikasi Detik.com, Kompas.com, dan Antara sebagai sumber berita asli;

dan (3) tweet yang ditautkan ke berita-berita tersebut. Dataset terdiri dari 5.548 sampel (2.774 hoaks dan 2.774 berita asli), mencerminkan distribusi yang seimbang untuk menghindari bias kelas.

Pembagian dataset menggunakan *stratified split*: 70% data latih (3.884 sampel), 15% data validasi (831 sampel), dan 15% data uji (833 sampel). Stratified split memastikan proporsi kelas hoaks dan non-hoaks tetap konsisten di setiap split.

Tabel 1

Distribusi Dataset IndoFakeNews

Kategori	Jumlah	Persentase	Sumber Utama
Hoaks	2774	50%	TurnBackHoax.id, Hoax-Slayer.id
Bukan Hoaks	2774	50%	Detik.com, Kompas.com, Antara
Total	5548	100%	—

Sumber: Data Diolah

Ekstraksi Fitur Linguistik

Ekstraksi fitur adalah tahap kritis dalam penelitian ini. Kami merancang 47 fitur yang dikelompokkan dalam empat kategori berdasarkan karakteristik khas hoaks dalam Bahasa Indonesia (Ratnaningsih et al., 2025).

Kelompok A – Fitur Leksikal (12 fitur): Mencakup skor kecocokan dengan kamus kata hiperbola (misal: "luar biasa", "mengejutkan", "terbukti"), kamus kata urgensi ("segera", "darurat", "langsung"), dan kamus kata konspirasi ("disembunyikan", "ditutup-tutupi"). Skor dihitung sebagai proporsi kata yang cocok terhadap total kata.

Kamus leksikal untuk ketiga kategori kata (hiperbola, urgensi, dan konspirasi) dibangun melalui prosedur semi-otomatis tiga tahap. Tahap pertama, penyusunan benih (seed words) secara manual oleh dua anotator dengan latar belakang linguistik, yang merujuk pada literatur retorika hoaks dan inspeksi terhadap subset 500 sampel hoaks pada data latih. Tahap kedua, perluasan kosakata secara otomatis menggunakan kemiripan semantik: setiap kata benih diperluas dengan kata-kata bertetangga terdekat berdasarkan word embedding FastText Bahasa Indonesia (cosine similarity ≥ 0,6), kemudian kandidat hasil perluasan diverifikasi ulang secara manual untuk membuang kata yang tidak relevan. Tahap ketiga, finalisasi dan penyaringan duplikat. Kesepakatan antar-anotator pada tahap verifikasi diukur dengan Cohen's kappa dan memperoleh nilai 0,81 (kesepakatan tinggi). Seluruh kamus disusun dalam bentuk lema dasar dan dicocokkan terhadap teks setelah proses stemming.

Tabel 2

Komposisi Kamus Leksikal yang Digunakan

Kategori Kamus	Jumlah Kata	Contoh Kata
Hiperbola	142	luar biasa, mengejutkan, terbukti, dahsyat, fenomenal
Urgensi	98	segera, darurat, langsung, sekarang juga, mendesak

Konspirasi	76	disembunyikan, ditutup-tutupi, dibungkam, rahasia, dalang
Total	316	—

Sumber: Data Diolah

Kelompok B – Fitur Struktural (8 fitur): Mencakup jumlah tanda seru, jumlah tanda tanya, rasio huruf kapital terhadap total karakter, jumlah kata dalam huruf kapital semua, dan panjang teks (jumlah karakter dan kata). Fitur-fitur ini menangkap pola sensasionalisme yang sering digunakan dalam hoaks.

Kelompok C – Fitur Stilistika (10 fitur): Mencakup skor keterbacaan (Flesch-Kincaid yang diadaptasi untuk Bahasa Indonesia), rata-rata panjang kalimat, rata-rata panjang kata, keragaman kosakata (type-token ratio), dan rasio kata konkret terhadap kata abstrak. Hoaks cenderung menggunakan bahasa yang lebih sederhana namun dramatis.

Kelompok D – Fitur Statistik Teks (17 fitur): Mencakup representasi TF-IDF dengan 15 fitur (top-15 TF-IDF terms) serta 2 fitur tambahan berupa karakter n-gram (bigram dan trigram) yang paling diskriminatif. Kelompok ini menangkap pola kata tingkat permukaan yang berkorelasi dengan hoaks.

Untuk mendukung reproduktibilitas, Tabel 3 merinci keseluruhan 47 fitur yang diekstrak beserta jumlah fitur per kelompok dan metode komputasinya. Seluruh proses ekstraksi diimplementasikan dengan Python menggunakan pustaka Sastrawi untuk stemming Bahasa Indonesia, NLTK untuk tokenisasi, dan scikit-learn untuk vektorisasi TF-IDF. Sebelum ekstraksi, teks melalui tahap prapemrosesan baku: case folding, pembersihan URL dan mention, normalisasi kata tidak baku menggunakan kamus alay, penghapusan stopword, dan stemming.

Tabel 3

Rincian Ekstraksi 47 Fitur Linguistik

Kelompok	Jumlah	Metode Ekstraksi
A – Leksikal	12	Proporsi kecocokan terhadap kamus hiperbola, urgensi, dan konspirasi (4 fitur per kamus: hitung, proporsi, kepadatan, biner keberadaan)
B – Struktural	8	Hitung tanda seru, tanda tanya, rasio huruf kapital, jumlah kata kapital penuh, panjang karakter, panjang kata, rata-rata, dan rasio tanda baca
C – Stilistika	10	Skor keterbacaan adaptasi Flesch-Kincaid, rata-rata panjang kalimat/kata, type-token ratio, rasio kata konkret-abstrak, dan kepadatan leksikal
D – Statistik Teks	17	Top-15 TF-IDF terms (unigram) + 2 fitur karakter n-gram (bigram dan trigram) paling diskriminatif berdasarkan chi-square
Total	47	—

Sumber: Data Diolah

Seluruh fitur numerik dinormalisasi menggunakan standardisasi z-score (mean nol, simpangan baku satu) sebelum dimasukkan ke model, kecuali fitur biner. Untuk

memastikan reproduktibilitas, urutan fitur, parameter vektorizer TF-IDF (ukuran vocabulary, rentang n-gram, dan ambang frekuensi dokumen), serta seluruh kamus leksikal didokumentasikan dan akan disediakan dalam repositori publik bersama kode sumber.

Arsitektur Ensemble yang Diusulkan

Model ensemble yang diusulkan menggabungkan tiga algoritma dasar: (1) Random Forest dengan 200 pohon keputusan; (2) Gradient Boosting dengan learning rate 0,1 dan 150 estimator; dan (3) Linear Support Vector Machine dengan kernel RBF. Ketiga model dilatih secara independen pada data latih yang sama.

Pemilihan ketiga algoritma ini dilandasi prinsip keberagaman (diversity) pengklasifikasi, yang merupakan syarat utama keberhasilan ensemble: model-model dasar sebaiknya memiliki bias induktif yang berbeda agar kesalahan masing-masing tidak saling berkorelasi sehingga dapat dikompensasi melalui penggabungan. Random Forest dipilih sebagai representasi paradigma bagging yang membangun banyak pohon keputusan secara paralel dengan subsampling fitur, sehingga robust terhadap overfitting dan mampu menangkap interaksi fitur non-linear. Gradient Boosting dipilih sebagai representasi paradigma boosting yang membangun pengklasifikasi secara sekuensial dengan menekankan sampel yang sulit, sehingga unggul dalam memodelkan pola halus yang sulit dipisahkan secara linear. Support Vector Machine dipilih karena efektif pada ruang fitur berdimensi relatif tinggi dan menghasilkan batas keputusan dengan margin maksimum, memberikan karakteristik generalisasi yang berbeda dari kedua model berbasis pohon. Dengan demikian, ketiga algoritma mewakili tiga keluarga pendekatan yang berbeda (bagging, boosting, dan margin-based), sehingga kombinasinya melalui soft voting diharapkan menghasilkan prediksi yang lebih akurat dan stabil dibandingkan model tunggal mana pun.

Kombinasi dilakukan melalui *soft voting* berbobot. Bobot optimal untuk setiap model (w_1, w_2, w_3) ditentukan melalui grid search pada data validasi. Probabilitas kelas akhir dihitung sebagai:

$$\boldsymbol{P(y = hoaks|x) = w_1 \cdot P_{RF} + w_2 \cdot P_{GB} + w_3 \cdot P_{SVM}} \quad (1)$$

Dimana w_1=0,35, w_2=0,40, w_3=0,25 adalah bobot optimal yang diperoleh dari grid search pada data validasi.

Mekanisme soft voting bekerja dengan menggabungkan probabilitas kelas (bukan label keras) yang dihasilkan setiap model dasar. Untuk sebuah sampel masukan, masing-masing model menghasilkan estimasi probabilitas keanggotaan kelas, misalnya P_RF (hoaks), P_GB (hoaks), dan P_SVM (hoaks). Ketiga probabilitas ini dijumlahkan secara berbobot sesuai Persamaan (1) untuk memperoleh probabilitas gabungan. Label akhir ditentukan dengan memilih kelas yang memiliki probabilitas gabungan tertinggi (argmax). Berbeda dengan hard voting yang hanya menghitung suara mayoritas dari label diskret, soft voting memanfaatkan tingkat keyakinan (confidence) tiap model, sehingga model yang sangat yakin terhadap suatu prediksi memberikan kontribusi lebih besar dibandingkan model yang ragu. Mekanisme ini terbukti lebih tahan terhadap kesalahan individual: ketika satu model keliru namun dengan keyakinan rendah, dua model lain yang benar dengan keyakinan tinggi tetap dapat mengoreksi hasil akhir. Karena SVM linear secara default menghasilkan skor jarak, probabilitas SVM dikalibrasi terlebih dahulu menggunakan

metode Platt scaling agar berada pada skala yang sebanding dengan kedua model lainnya sebelum digabungkan.

Penentuan bobot (w_1, w_2, w_3) dilakukan melalui pencarian grid (grid search) yang sistematis pada data validasi. Ruang pencarian didefinisikan dengan menumbuhkan setiap bobot pada rentang [0; 1] dengan langkah (step) 0,05, dengan batasan $w_1 + w_2 + w_3 = 1$ sehingga hanya kombinasi yang memenuhi kendala simpleks yang dievaluasi. Untuk setiap kombinasi bobot kandidat, ensemble dievaluasi pada data validasi (831 sampel) dan dihitung F1-score-nya. Kombinasi bobot yang menghasilkan F1-score validasi tertinggi dipilih sebagai konfigurasi optimal, yaitu w_1=0,35 (Random Forest), w_2=0,40 (Gradient Boosting), dan w_3=0,25 (SVM). Hasil ini konsisten dengan performa individual masing-masing model pada data validasi, di mana Gradient Boosting memberikan kontribusi tunggal terkuat sehingga memperoleh bobot terbesar, sedangkan SVM yang relatif paling lemah memperoleh bobot terkecil. Penggunaan data validasi yang terpisah dari data uji memastikan bahwa pemilihan bobot tidak menyebabkan kebocoran informasi (data leakage) terhadap evaluasi akhir.

Baseline Perbandingan

Untuk mengevaluasi kontribusi metode yang diusulkan secara objektif, eksperimen membandingkan dengan lima baseline: (1) Logistic Regression dengan TF-IDF; (2) Naive Bayes Multinomial; (3) LSTM dengan embedding FastText; (4) IndoBERT fine-tuned; dan (5) masing-masing komponen ensemble secara individual.

Kelima baseline tersebut dipilih agar mewakili spektrum pendekatan yang representatif dalam literatur deteksi hoaks dan klasifikasi teks, sekaligus memungkinkan perbandingan yang adil pada beberapa tingkat kompleksitas. Logistic Regression dan Naive Bayes Multinomial mewakili model klasik berbasis fitur statistik yang umum digunakan sebagai baseline standar pada tugas klasifikasi teks karena kesederhanaan dan kecepatannya. LSTM dengan embedding FastText mewakili pendekatan deep learning berbasis sequence yang lazim diadopsi pada penelitian deteksi berita palsu Bahasa Indonesia sebelumnya, sehingga relevan sebagai pembanding dari paradigma neural. IndoBERT fine-tuned dipilih sebagai baseline state-of-the-art karena merupakan model bahasa pra-latih (pretrained) terkuat untuk Bahasa Indonesia dan menjadi acuan akurasi tertinggi pada banyak tugas NLP Bahasa Indonesia. Terakhir, perbandingan terhadap masing-masing komponen ensemble secara individual (Random Forest, Gradient Boosting, dan SVM) berfungsi sebagai ablation pada level model untuk membuktikan bahwa penggabungan ensemble memberikan peningkatan nyata di atas model tunggal terbaik. Dengan demikian, pemilihan baseline tidak hanya mengacu pada satu penelitian tertentu, tetapi disusun untuk mencakup baseline klasik, neural, pretrained Transformer, dan komponen internal, sehingga keunggulan metode yang diusulkan dapat dinilai secara menyeluruh.

HASIL DAN PEMBAHASAN

Hasil Evaluasi Keseluruhan

Hasil performa seluruh metode yang dibandingkan pada data uji seperti ditunjukkan pada Tabel 4. Metode ensemble yang diusulkan (Proposed-Ensemble) lebih unggul untuk semua

baseline pada metrik F1-score, yang merupakan metrik utama karena distribusi kelas yang seimbang.

Tabel 4
Perbandingan Performa Metode

Metode	Accuracy	Precision	Recall	F1-Score	Waktu Inferensi *
Logistic Regression + TF-IDF	80,2%	79,8%	81,1%	80,4%	0,03 ms
Naive Bayes Multinomial	76,5%	75,2%	78,3%	76,7%	0,01 ms
Random Forest (tunggal)	85,1%	84,6%	85,9%	85,2%	0,12 ms
Gradient Boosting (tunggal)	86,4%	85,9%	87,0%	86,4%	0,18 ms
SVM Linear (tunggal)	83,7%	83,1%	84,5%	83,8%	0,09 ms
LSTM + FastText	85,8%	85,2%	86,6%	85,9%	2,14 ms
IndoBERT Fine-tuned	88,5%	87,9%	89,2%	88,2%	20,7 ms
Proposed-Ensemble	89,3%	88,7%	90,1%	89,4%	0,25 ms

**Waktu inferensi per sampel diukur pada CPU Intel Core i5-1135G7, tanpa GPU.*

Sumber: Data Diolah

Analisis Kontribusi Kelompok Fitur

Untuk memahami kontribusi masing-masing kelompok fitur terhadap performa keseluruhan, dilakukan ablation study dengan melatih ulang ensemble menggunakan subset fitur. Hasilnya disajikan pada Tabel 5.

Tabel 5

Ablation Study: Kontribusi Fitur

Konfigurasi Fitur	Accuracy	F1-Score	Δ F1 vs Full
Semua fitur (Full)	89,3%	89,4%	—
Tanpa Fitur Leksikal (−A)	86,1%	86,3%	−3,1%
Tanpa Fitur Struktural (−B)	88,0%	88,1%	−1,3%
Tanpa Fitur Stilistika (−C)	87,8%	87,9%	−1,5%
Tanpa TF-IDF/n-gram (−D)	86,9%	87,0%	−2,4%
Hanya Fitur Leksikal (A saja)	78,4%	78,6%	−10,8%
Hanya TF-IDF/n-gram (D saja)	83,5%	83,7%	−5,7%

Sumber: Data Diolah

Hasil *ablation study* menunjukkan bahwa Fitur Leksikal (Kelompok A) memberikan kontribusi terbesar dengan penurunan F1-score sebesar 3,1% ketika dihilangkan. Hal ini mengkonfirmasi hipotesis bahwa pola leksikal spesifik—khususnya kata-kata hiperbola dan urgensi—merupakan penanda diskriminatif yang kuat untuk hoaks Bahasa Indonesia.

Fitur TF-IDF dan n-gram (Kelompok D) juga memberikan kontribusi signifikan (−2,4%), menunjukkan bahwa pola statistik tingkat permukaan masih relevan meskipun lebih rentan terhadap variasi bahasa dibandingkan fitur leksikal berbasis kamus.

Analisis Error dan Kasus Sulit

Inspeksi manual terhadap 89 sampel yang salah diklasifikasikan mengungkap tiga pola kesalahan utama:

1. Hoaks tersamarkan (*camouflaged hoax*): Hoaks yang ditulis dengan gaya bahasa jurnalistik formal dan menghindari kata-kata sensasional, sehingga fitur leksikal gagal mendeteksinya. Sebanyak 38 dari 89 kesalahan (42,7%) termasuk dalam kategori ini.

2. Berita asli yang sensasional: Berita asli tentang bencana atau krisis yang secara alami menggunakan bahasa darurat dan dramatis, sehingga menyerupai pola hoaks. Sebanyak 31 dari 89 kesalahan (34,8%) berasal dari kategori ini.

3. Ketergantungan konteks eksternal: Hoaks yang kebenarannya hanya dapat dinilai dengan informasi eksternal (misalnya tanggal kejadian atau lokasi spesifik) yang tidak tersedia dalam teks itu sendiri. Sebanyak 20 dari 89 kesalahan (22,5%) memerlukan verifikasi fakta eksternal.

Untuk memberikan gambaran yang lebih konkret mengenai pola kesalahan klasifikasi, Tabel 6 merangkum distribusi 89 sampel yang salah diprediksi berdasarkan kategori kesalahan beserta jenis kesalahannya (False Positive atau False Negative). False Positive merujuk pada berita asli yang keliru diklasifikasikan sebagai hoaks, sedangkan False Negative merujuk pada hoaks yang lolos terdeteksi sebagai berita asli.

Tabel 6

Distribusi Kesalahan Klasifikasi Berdasarkan Kategori

Kategori Kesalahan	Jumlah	Proporsi	Jenis
Hoaks tersamarkan	38	42,7%	False Negative
Berita asli sensasional	31	34,8%	False Positive
Ketergantungan konteks eksternal	20	22,5%	FN/FP
Total kesalahan	89	100%	—

Sumber: Data Diolah

Untuk mengilustrasikan ketiga kategori tersebut, berikut disajikan contoh representatif dari sampel yang salah diklasifikasikan beserta analisisnya:

Contoh False Negative (hoaks tersamarkan). Teks: “Kementerian Kesehatan resmi mengumumkan penundaan program vaksinasi nasional tahap dua karena alasan evaluasi distribusi.” Berita ini sesungguhnya hoaks, namun ditulis dengan gaya jurnalistik formal tanpa kata hiperbola maupun seruan urgensi. Akibatnya, fitur leksikal (Kelompok A) menghasilkan skor mendekati nol dan model memprediksi kelas “bukan hoaks” dengan probabilitas gabungan 0,78. Kasus ini menunjukkan keterbatasan fitur berbasis kamus ketika hoaks sengaja menyamarkan diri dalam register bahasa formal.

Contoh False Positive (berita asli sensasional). Teks: "DARURAT! Gempa magnitudo 6,2 guncang wilayah pesisir, warga diminta SEGERA menjauhi pantai dan waspada potensi tsunami!" Berita ini faktual dan berasal dari sumber resmi BMKG, tetapi secara alami memuat kata urgensi ("darurat", "segera"), huruf kapital, dan tanda seru yang merupakan penanda kuat hoaks pada model. Fitur struktural (Kelompok B) dan leksikal terpicu sehingga model keliru memprediksi kelas "hoaks" dengan probabilitas 0,71. Kasus ini memperlihatkan bahwa konteks kebencanaan yang sah dapat menyerupai pola retorika hoaks.

Contoh ketergantungan konteks eksternal. Teks: "Pemerintah menggratiskan seluruh tarif tol selama libur panjang pekan ini." Kebenaran klaim ini tidak dapat ditentukan hanya dari teks karena bergantung pada kebijakan aktual pada periode tertentu. Tanpa verifikasi fakta eksternal, baik model maupun pembaca manusia kesulitan menilai, sehingga sampel jenis ini berkontribusi pada kesalahan FN maupun FP tergantung pada arah prediksi.

Selain analisis pada level ensemble, dilakukan inspeksi terhadap kasus split, yaitu sampel di mana ketiga algoritma anggota ensemble menghasilkan prediksi yang tidak bulat (terjadi ketidaksepakatan). Tabel 7 menyajikan tiga contoh kasus split beserta probabilitas hoaks yang dihasilkan tiap algoritma dan hasil akhir soft voting.

Tabel 7

Contoh Kasus Split Antar-Algoritma Ensemble

Sampel	P(RF)	P(GB)	P(SVM)	Hasil Akhir
S1 – hoaks tersamarkan	0,41	0,58	0,39	Bukan Hoaks
S2 – berita asli sensasional	0,63	0,47	0,66	Hoaks
S3 – hoaks bahasa gaul	0,72	0,69	0,44	Hoaks

Sumber: Data Diolah

Pola yang teramati dari kasus split memperkuat justifikasi pemilihan ketiga algoritma. Pada sampel S1, Gradient Boosting cenderung lebih sensitif terhadap pola halus sehingga memberikan probabilitas hoaks tertinggi, namun karena Random Forest dan SVM lebih konservatif, hasil soft voting condong ke "bukan hoaks", dan inilah kesalahan FN pada hoaks tersamarkan. Pada S2, dua model berbasis margin dan bagging (SVM dan RF) terpicu oleh fitur sensasional sehingga mengalahkan GB, menghasilkan FP. Sebaliknya pada S3, kekuatan RF dan GB dalam menangkap interaksi fitur leksikal berhasil mengoreksi keraguan SVM, menghasilkan klasifikasi hoaks yang benar. Observasi ini menegaskan bahwa keberagaman perilaku antar-algoritma adalah pedang bermata dua: ia mengoreksi kesalahan pada sebagian besar kasus, namun pada kasus ambigu, distribusi bobot menjadi penentu hasil akhir.

Perbandingan Efisiensi Komputasi

Salah satu keunggulan kunci metode yang diusulkan adalah efisiensi komputasi. IndoBERT, meskipun memiliki F1-score yang sedikit lebih rendah (88,2% vs 89,4%), memerlukan waktu inferensi 20,7 ms per sampel dibandingkan hanya 0,25 ms untuk Proposed-Ensemble—perbedaan 83 kali lipat. Pada skenario monitoring real-time dengan volume 10.000 tweet per menit, Proposed-Ensemble dapat memproses seluruh volume

tersebut hanya dalam 2,5 detik, sementara IndoBERT memerlukan 207 menit dengan CPU standar.

Selain itu, Proposed-Ensemble tidak memerlukan GPU, dapat dijalankan pada server standar, dan ukuran model yang hanya 45 MB dibandingkan 500 MB untuk IndoBERT menjadikannya pilihan yang jauh lebih praktis untuk deployment produksi di lingkungan dengan sumber daya terbatas.

SIMPULAN

Penelitian ini telah mengusulkan dan mengevaluasi metode deteksi hoaks berbahasa Indonesia berbasis kombinasi fitur linguistik spesifik-Indonesia dan ensemble machine learning. Metode yang diusulkan berhasil mencapai F1-score 89,4% pada dataset IndoFakeNews, melampaui IndoBERT fine-tuned sebesar 1,2 poin persentase dengan kecepatan inferensi 83 kali lebih cepat.

Temuan utama penelitian ini adalah: (1) Fitur leksikal berbasis kamus domain merupakan penanda diskriminatif paling kuat untuk hoaks Bahasa Indonesia, berkontribusi 3,1% pada F1-score; (2) *Ensemble* dengan *soft voting* berbobot secara konsisten mengungguli model tunggal terbaik (*Gradient Boosting*) sebesar 3,0 poin F1; dan (3) Tradeoff antara akurasi dan efisiensi komputasi mendukung penggunaan *Proposed-Ensemble* untuk skenario *deployment* dengan sumber daya terbatas.

Keterbatasan penelitian ini mencakup ketergantungan kamus leksikal yang perlu diperbarui secara berkala seiring perkembangan bahasa, serta ketidakmampuan mendeteksi hoaks yang memerlukan verifikasi fakta eksternal. Penelitian lanjutan disarankan untuk mengeksplorasi: (1) Integrasi fitur jaringan sosial (pola penyebaran tweet) sebagai sinyal tambahan; (2) Penggunaan teknik augmentasi data untuk meningkatkan robustness terhadap variasi bahasa; dan (3) Pengembangan sistem deteksi hoaks multi-modal yang mengintegrasikan analisis gambar dan teks.

DAFTAR PUSTAKA

Agma, A. R. (2025). Hoaks dan Disinformasi di Media Sosial: Strategi Literasi Digital untuk Meningkatkan Kesadaran Publik Jurnal Komunikasi dan Media. *Jurnal Komunikasi Dan Media*, *1*(1), 30–36.

Aisyah, S., & Yasmin, A. (2022). Hoax News and Future Treats : A Study of the Constitution , Pancasila , and the Law. *Indonesian Journal of Pancasila Dan Global Constitutionalism*, *1*(1), 171–238.

Cendana, M., & Permana, S. D. H. (2019). Pra-Pemrosesan Teks pada Grup Whatsapp untuk Pemodelan Topik. *Junal Mantik Penusa*, *3*(3), 107–116.

Febriyanty, N. E., Hariyadi, M. A., & Crysdian, C. (2023). Hoax Detection News Using Naïve Bayes and Support Vector Machine Algorithm. *International Journal of Advances in Data and Information Systems*, *4*(2), 191–200. https://doi.org/10.25008/ijadis.v4i2.1306

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify

fake news. *Future Generation Computer Systems*, *117*(April), 47–58. https://doi.org/10.1016/j.future.2020.11.022

Halawa, N., & Lase, F. (2022). Mengentaskan Hoax Dengan Membaca Pemahaman Di Era Digital. *Educativo: Jurnal Pendidikan*, *1*(1), 235–243.

Iskandar, D., Suryawati, I., Suratno, G., Liliyana, L., Muhtadi, M., & Ngimadudin, N. (2024). Public Communication Model in Combating Hoaxes and Fake News in Ahead of the 2024 General Election. *International Journal of Environmental Sustainability and Social Science*, *4*(5), 1505–1518.

Jeong, U., Sheth, P., Tahir, A., Alatawi, F., Bernard, H. R., & Liu, H. (2024). Exploring Platform Migration Patterns between Twitter and Mastodon: A User Behavior Study. *Proceeding of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)*, 738–750.

Joses, S., Quinevera, S., Mardianto, R., Yulvida, D., & Shiddiqi, A. M. (2024). Pendekatan Metode Ensemble Learning untuk Deteksi Serangan DDoS Menggunakan Soft Voting Classifier. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, *10*(1), 79–87.

Kaliyar, R. K., Goswami, A., & Narang, P. (2020). FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research*, *61*, 32–44. https://doi.org/10.1016/j.cogsys.2019.12.005

Koto, F., & Baldwin, T. (2020). *IndoLEM and IndoBERT : A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. 757–770.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). *Supplementary Materials for The science of fake news*. *1094*(March), 1–12. https://doi.org/10.1126/science.aao2998

Muhabatin, H., Prabowo, C., Ali, I., Lukman Rohmat, C., Rizki Amalia, D., Sitasi, C., & Rizki, D. (2021). Klasifikasi Berita Hoax Menggunakan Algoritma Naive Bayes Berbasis PSO. *Informatics for Educators and Professionals*, *5*(2), 156–165.

Noguera-vivo, J. M., Grandío-pérez, M. M., Villar-rodríguez, G., & Martín, A. (2023). Disinformation and Vaccines on Social Networks: Behavior of Hoaxes on Twitter. *RLCS, Revista Latina de Comunicación Social*, *19*(81), 44–62.

Ohorella, N. R. (2023). The Utilization of Twitter in The Anticipation of Covid-19 Hoax News in Yogyakarta City. *Journal of Communication Studies and Society*, *2*(1), 26–33. https://doi.org/10.38043/commusty.v2i1.4979

Pasek, P., Mahawardana, O., Arya, G., Agus, I. P., & Pratama, E. (2022). Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap "Figure Pemimpin" Menggunakan Python. *JITTER - Jurnal Ilmiah Teknologi Dan Komputer*, *3*(1).

Rahmawati, D., Setyo, R., Robawa, P., Abiyyi, M. F. Al, Rf, P. D. N., Nugraha, R. I., & Margono, F. P. (2023). Analisis Hoaks dalam Konteks Digital: Implikasi dan Pencegahannya di Indonesia. *INNOVATIVE: Journal of Social Science Research*, *3*(2).

Ratnaningsih, D., Iskandar, I., & Anwar, M. (2025). Derivasi Verba dalam Bahasa Lampung melalui Pendekatan Linguistik Komputasional. *Sienna*, *6*(2).

Rianansyah, A., Utami, E., & Ariatmanto, D. (2025). Implementation of Word Embedding in Detecting Political Fake News in indonesia using Long Short-Term Memory Algorithm. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *10*(4), 3124–3135.

Rizka, A. S., & Sari, V. (2025). Penerapan Teknik Soft Voting Ensemble pada Klasifikasi Rating Film. *Indonesian Journal of Applied Statistics (IJAS)*, *May*, 24–37. https://doi.org/10.13057/ijas.v8i1.100904.1.

Setiawan, R., Ponnam, V. S., Sengan, S., Anam, M., Subbiah, C., Phasinam, K., Vairaven, M., & Ponnusamy, S. (2022). Certain investigation of fake news detection from Facebook and Twitter using ensemble machine learning approach. *Wireless Personal Communications*, *127*, 1737–1762. https://doi.org/10.1007/s11277-021-08928-9

Suherman, A., Ferdi, F., Maulana, H. F., & Putra, M. R. A. (2024). Sentiments Analysis on Twitter Towards Hoax Information on Social Media News. *Jurnal Ilmu Komunikasi*, *22*(1), 61–76.

V, C. T., Yadav, S. T., Chaitanya, C., A, N. B., & S, S. M. (2024). Cardiac Arrhythmia Classification using Ensemble Machine Learning Algorithms with PPG and ECG signals. *2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, *2018*, 1–5. https://doi.org/10.1109/ICRASET63057.2024.10894999

Wu, X., Kumar, V., Quinlan, R. J., Ghosh, J., Yang, Q., Motoda, H., Mclachlan, G., Ng, A., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 Algorithms in Data Mining. *Knowledge Information System*, *14*, 1–37.

Yaghoubi, E., Yaghoubi, E., Khamees, A., & Hossein, A. (2024). A systematic review and meta-analysis of artificial neural network , machine learning , deep learning , and ensemble learning approaches in field of geotechnical engineering. In *Neural Computing and Applications* (Vol. 36, Issue 21). Springer London. https://doi.org/10.1007/s00521-024-09893-7