JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Identifikasi Berita Palsu di Portal Media Online Menggunakan Model IndoBERT dan LSTM

Angga Mochamad Kamal, Yulison Herry Chrisnanto*, Rezki Yuniarti

Fakultas Sains dan Informatika, Program Studi Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia
Email: [1]angga.mk121003@gmail.com, [2,*]yhc@if.unjani.ac.id, [3]rezki.yuniarti@lecture.unjani.ac.id
Email Penulis Korespondensi: yhc@if.unjani.ac.id
Submitted **24-05-2025;** Accepted **28-06-2025;** Published **30-06-2025**

Abstrak

Pesatnya penyebaran berita palsu politik di portal media online Indonesia menimbulkan ancaman serius terhadap kepercayaan publik dan stabilitas demokrasi. Masalah utama penelitian ini adalah keterbatasan model existing dalam menangani kompleksitas narasi politik berbahasa Indonesia yang mengandung idiom lokal dan struktur teks panjang. Solusi yang diusulkan menggunakan model hybrid IndoBERT-LSTM dengan pendekatan ensemble stacking berbasis logistic regression meta-learner untuk mengoptimalkan deteksi berita palsu. IndoBERT dipilih untuk menangkap nuansa bahasa Indonesia, sementara LSTM menangani dependensi sekuensial pada artikel panjang. Tujuan penelitian adalah mengembangkan sistem deteksi otomatis yang akurat untuk berita palsu politik dengan memanfaatkan kelebihan komplementer kedua model. Dataset terdiri dari 32.218 artikel politik dari portal kredibel (Kompas, CNN Indonesia, Tempo, Detiknews, Viva) dan validasi Turnbackhoax.id periode September 2021-Desember 2024. Hasil penelitian menunjukkan ensemble stacking mencapai performa superior dengan F1-score 0.9544, akurasi 95.41%, dan AUC-ROC 0.9936, mengungguli IndoBERT standalone (F1: 0.9542) dan LSTM (F1: 0.9417). Analisis kesalahan mengidentifikasi 4.59% error rate dengan 134 false positive dan 88 false negative, terutama pada artikel panjang (rata-rata 2.739 karakter). Model ini berpotensi diintegrasikan pada platform fact-checking untuk deteksi real-time berita palsu politik Indonesia.

Kata Kunci: Berita Palsu; IndoBERT; LSTM; Ensemble Stacking; Deteksi Otomatis

Abstract

The rapid spread of political fake news on Indonesian online media portals poses serious threats to public trust and democratic stability. The main research problem is the limitation of existing models in handling the complexity of Indonesian political narratives containing local idioms and long text structures. The proposed solution employs a hybrid IndoBERT-LSTM model with ensemble stacking approach using logistic regression meta-learner to optimize fake news detection. IndoBERT is selected to capture Indonesian language nuances, while LSTM handles sequential dependencies in long articles. The research objective is to develop an accurate detection system for political fake news by leveraging the complementary strengths of both models. The dataset comprises 32,218 political articles from credible portals (Kompas, CNN Indonesia, Tempo, Detiknews, Viva) and Turnbackhoax.id validation from September 2021 to December 2024. Research results demonstrate that ensemble stacking achieves superior performance with F1-score 0.9544, accuracy 95.41%, and AUC-ROC 0.9936, outperforming standalone IndoBERT (F1: 0.9542) and LSTM (F1: 0.9417). Error analysis identifies 4.59% error rate with 134 false positives and 88 false negatives, particularly in long articles (average 2,739 characters). This model has potential for integration into fact-checking platforms for real-time detection of Indonesian political fake news.

Keywords: Fake News; IndoBERT; LSTM; Ensemble Stacking; Automatic Detection

1. PENDAHULUAN

Pesatnya perkembangan teknologi digital modern di Indonesia telah mempercepat penyebaran informasi melalui portal berita media online, namun juga menyebabkan meningkatnya berita palsu politik yang mengganggu kepercayaan masyarakat [1]. Contohnya, berita palsu tentang kecurangan Pemilu 2024 atau tuduhan korupsi terhadap tokoh politik sering menyebabkan konflik sosial sebelum diverifikasi oleh platform seperti Turnbackhoax.id [2]. Berita palsu ini biasanya dibuat untuk memengaruhi opini publik atau menarik lebih banyak klik di situs web [3]. Kasus seperti klaim palsu tentang hasil pemilu atau kebijakan pemerintah, misalnya kenaikan harga bahan bakar, menunjukkan dampak serius berita palsu di Indonesia [4]. Penelitian menunjukkan bahwa berita palsu politik dapat memperburuk perpecahan masyarakat dan memerlukan solusi cepat untuk identifikasi berita palsu [5]. Oleh karena itu, teknologi canggih seperti Natural Language Processing (NLP) diperlukan untuk mendeteksi berita palsu secara otomatis, terutama pada berita politik yang kompleks [6].

Di Indonesia, tantangan identifikasi berita palsu politik sangat besar karena keragaman budaya dan bahasa lokal [7]. Portal berita besar seperti Kompas, CNN Indonesia, Tempo, Detiknews, dan Viva sering menjadi sasaran penyebaran berita palsu [8]. Contohnya, berita palsu tentang kebijakan pajak baru atau dugaan skandal calon presiden dapat menyebabkan perdebatan sengit di media sosial [9]. Pemeriksaan manual oleh fact-checker, seperti yang dilakukan Turnbackhoax.id, tidak cukup cepat untuk menangani volume berita yang besar [10]. Narasi politik di Indonesia sering menggunakan istilah lokal seperti “kampanye hitam” atau “fitnah politik,” yang sulit dipahami oleh model standar [11]. NLP menjadi solusi potensial karena mampu menganalisis teks dalam bahasa Indonesia dengan lebih baik. Dengan NLP, identifikasi berita palsu dapat dilakukan lebih cepat dan akurat untuk mendukung kebenaran informasi di Indonesia [12].

Penelitian sebelumnya telah mencoba menggunakan IndoBERT untuk identifikasi berita palsu [13]. IndoBERT, model yang dirancang untuk bahasa Indonesia, berhasil mengenali berita palsu di media sosial karena memahami konteks lokal, seperti ungkapan atau slang Indonesia [14]. Misalnya, IndoBERT dapat mendeteksi berita palsu politik di Twitter dengan akurasi cukup baik [15]. Namun, IndoBERT kurang efektif untuk artikel berita panjang, seperti yang ada di portal berita, karena sulit menangkap urutan informasi dalam teks panjang [16]. Beberapa penelitian mencoba menyesuaikan

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

IndoBERT untuk konteks Indonesia dengan pelatihan tambahan, tetapi hasilnya masih terbatas pada teks pendek seperti postingan media sosial [17]. Meskipun IndoBERT unggul dalam memahami bahasa lokal, keterbatasannya dalam menganalisis struktur artikel panjang membuatnya kurang optimal untuk portal berita. Oleh karena itu, pendekatan lain diperlukan untuk melengkapi kelemahan IndoBERT [18]. Kondisi ini menegaskan bahwa untuk menangani berita palsu politik yang seringkali disajikan dalam format artikel panjang dengan narasi kompleks di portal media online, dibutuhkan sebuah arsitektur model yang mampu menggabungkan keunggulan pemahaman konteks linguistik dengan kemampuan analisis sekuensial pada teks yang ekstensif.

Long Short-Term Memory (LSTM) adalah model lain yang sering digunakan untuk identifikasi berita palsu [19]. LSTM lebih cocok untuk menganalisis teks panjang, seperti artikel berita, karena mampu menangkap urutan informasi secara berurutan [20]. Penelitian sebelumnya menunjukkan bahwa LSTM efektif untuk mengenali pola dalam artikel berita politik di Indonesia [21]. Namun, LSTM kurang kuat dalam memahami nuansa bahasa lokal dibandingkan IndoBERT, terutama untuk istilah khas Indonesia seperti “hoaks politik” [22]. Beberapa penelitian mencoba menggabungkan LSTM dengan model lain untuk berita palsu Pemilu 2024 menggunakan data Turnbackhoax.id, tetapi hasilnya belum maksimal tanpa metode penggabungan yang lebih canggih [10]. Meskipun LSTM memiliki keunggulan dalam menganalisis teks panjang, keterbatasannya dalam memahami konteks bahasa lokal membuatnya perlu dikombinasikan dengan model lain untuk hasil yang lebih baik [13]. Oleh karena itu, kombinasi strategis dengan model yang kuat dalam pemahaman semantik bahasa lokal menjadi esensial untuk membangun sistem deteksi berita palsu yang lebih holistik dan efektif, terutama untuk artikel berita politik di Indonesia.

Meskipun IndoBERT dan LSTM memiliki kelebihan, pendekatan sebelumnya masih memiliki keterbatasan. Metode seperti hard voting dan soft voting, yang menggabungkan beberapa model, kurang fleksibel karena tidak bisa menyesuaikan identifikasi berita palsu politik yang berubah cepat, seperti fitnah terhadap tokoh politik [21]. Penelitian lain menunjukkan bahwa pendekatan ensemble stacking lebih menjanjikan karena menggunakan meta-learner untuk menyatukan hasil model dengan lebih akurat [14]. Namun, model sebelumnya belum dioptimalkan untuk konteks Indonesia, terutama untuk artikel berita panjang dengan narasi politik yang beragam, seperti klaim palsu tentang kebijakan publik atau isu sosial [4]. Keragaman narasi berita palsu, termasuk penggunaan bahasa emosional atau konteks lokal, membuat identifikasi semakin sulit. Oleh karena itu, diperlukan pendekatan baru yang menggabungkan kelebihan IndoBERT dan LSTM untuk hasil yang lebih efektif di portal media online [18].

Penelitian ini mengusulkan model yang menggabungkan IndoBERT dan LSTM dengan ensemble stacking untuk identifikasi berita palsu politik di portal media online. IndoBERT dipilih karena mampu memahami bahasa Indonesia, termasuk ekspresi lokal, sedangkan LSTM cocok untuk artikel berita panjang. Ensemble stacking digunakan untuk memanfaatkan kelebihan kedua model, dengan logistic regression sebagai meta-learner karena lebih sederhana dan cepat dibandingkan model seperti transformer-based model atau RNN. Data diambil dari portal berita seperti Kompas, CNN Indonesia, Tempo, Detiknews, dan Viva, serta divalidasi dengan Turnbackhoax.id. Model ini akan diuji dengan metrik akurasi, presisi, recall, dan F1-score untuk memastikan performa seimbang. Penelitian ini bertujuan menghasilkan model yang lebih efektif untuk mengenali berita palsu politik dan mendukung upaya menjaga kebenaran informasi di Indonesia, terutama dalam konteks politik yang kompleks.

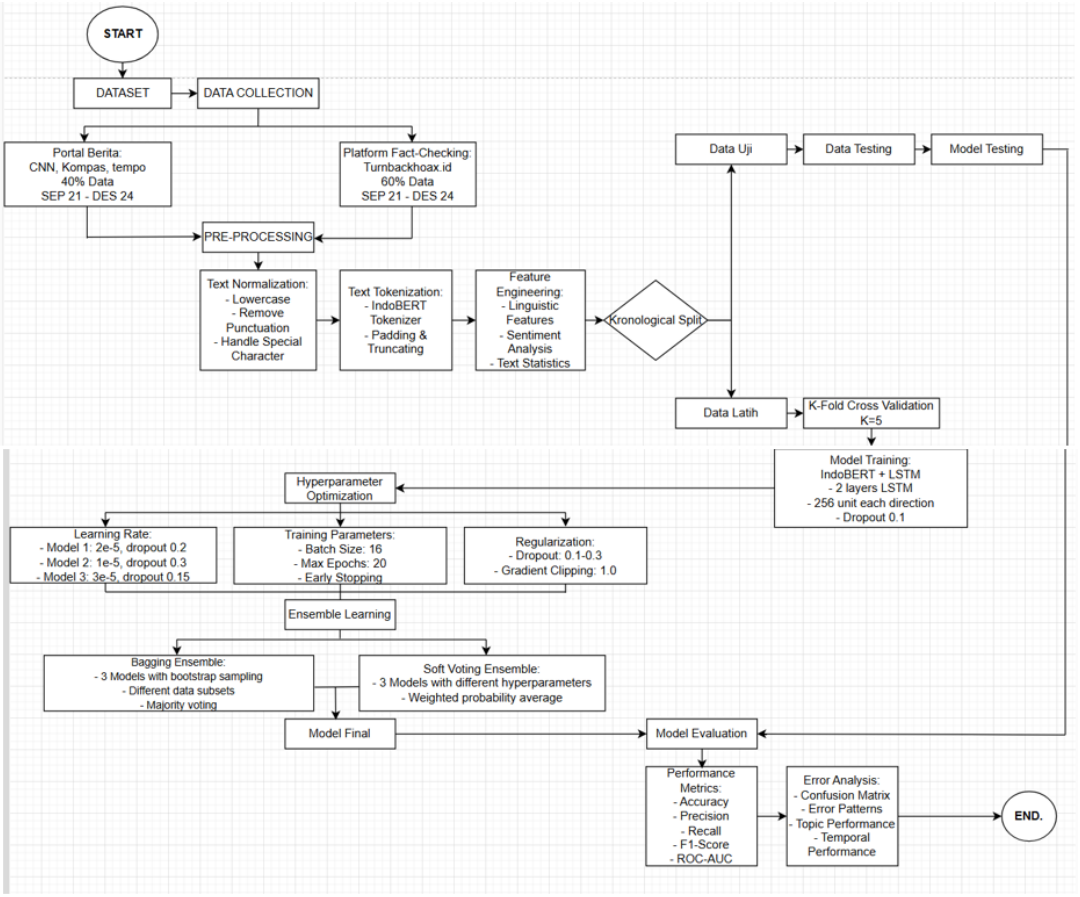
.Untuk mengatasi keterbatasan model-model sebelumnya dalam mengidentifikasi berita palsu politik berbahasa Indonesia, terutama yang melibatkan narasi kompleks dan teks panjang, penelitian ini menyajikan sebuah model hybrid IndoBERT-LSTM yang diperkuat dengan pendekatan ensemble stacking berbasis logistic regression meta-learner. Model ini dirancang untuk secara efektif menggabungkan kemampuan IndoBERT dalam memahami nuansa bahasa lokal dengan kekuatan LSTM dalam memproses dependensi sekuensial pada artikel berita yang panjang. Melalui optimasi dan evaluasi mendalam, penelitian ini bertujuan untuk menunjukkan peningkatan signifikan dalam akurasi deteksi berita palsu politik, serta mengeksplorasi potensi implementasinya dalam sistem fact-checking otomatis. Bagian selanjutnya dari dokumen ini akan merinci metodologi penelitian, hasil eksperimen, serta analisis implikasinya.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menguraikan pendekatan sistematis yang diterapkan untuk mengembangkan model deteksi berita palsu politik berbahasa Indonesia. Mengingat kompleksitas narasi politik di portal media online Indonesia, serta tantangan penanganan kiasan lokal dan struktur teks panjang, diperlukan alur penelitian yang terstruktur. Desain metodologi yang komprehensif adalah kunci untuk memastikan validitas dan reliabilitas hasil [1]. Penelitian ini dirancang melalui enam tahapan utama yang saling terkait, dimulai dari pengumpulan data hingga evaluasi performa model, guna mencapai akurasi tinggi dan kemampuan generalisasi yang baik:

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 1. Identifikasi berita palsu

Tahapan meliputi:

a. Pengumpulan corpus: Mengambil artikel dari portal berita kredibel (Kompas, CNN Indonesia, Tempo, Detiknews, Viva) dan Turnbackhoax.id untuk representasi berita palsu dan non-palsu.
b. Pra-pemrosesan data: Membersihkan teks, mengekstraksi fitur, dan menyeimbangkan dataset untuk model.
c. Pengembangan arsitektur model: Merancang model hybrid IndoBERT-LSTM dengan ensemble stacking logistic regression meta-learner.
d. Pelatihan model: Melatih model dengan optimasi hiperparameter.
e. Evaluasi model: Mengukur performa dengan metrik standar dan analisis kesalahan.
f. Perbandingan model ensemble: Membandingkan ensemble stacking logistic regression meta-learner dengan ensemble bagging, hard-voting dan soft-voting.

2.2 Corpus Penelitian

Kualitas corpus sangat menentukan keberhasilan deteksi berita palsu, terutama dalam konteks politik Indonesia yang dinamis [4]. Corpus ini terdiri dari 32.220 artikel berbahasa Indonesia (50% berita palsu, 50% non-palsu setelah under-sampling dari 43.777 artikel), dikumpulkan dari portal berita kredibel dan Turnbackhoax.id antara September 2021 dan Desember 2024. Proses pengumpulan meliputi:

a. Seleksi data: Artikel dipilih berdasarkan relevansi politik (Pemilu 2024, kebijakan pajak, skandal tokoh), panjang minimal 150 kata, dan kata kunci seperti “hoaks politik” atau “disinformasi.” Kategori narasi mencakup fitnah tokoh, klaim kebijakan, dan isu sosial.
b. Web scraping: Menggunakan Scrapy dan Beautiful Soup dengan rate limiting (1 request per 2 detik) dan rotasi user-agent. Izin diperoleh melalui komunikasi formal dengan pengelola portal berita.
c. Validasi data: Pemeriksaan integritas, deteksi duplikasi (cosine similarity >0.9 dihapus), dan verifikasi manual 10% sampel oleh dua pengecek independen. Diagram distribusi narasi (misalnya, 40% fitnah tokoh, 30% klaim kebijakan) dibuat untuk memahami variasi corpus.
d. Pelabelan: Semi-otomatis menggunakan Snorkel dengan delapan Labeling Functions (sumber, judul clickbait, kata emosional). Akurasi divalidasi manual untuk 20% artikel, dengan diskusi tim untuk label kontroversial.
e. Etika: Data dianonimisasi untuk menghapus informasi pribadi. Laporan etika disusun untuk transparansi.

2.3 Pra-Pemrosesan Data

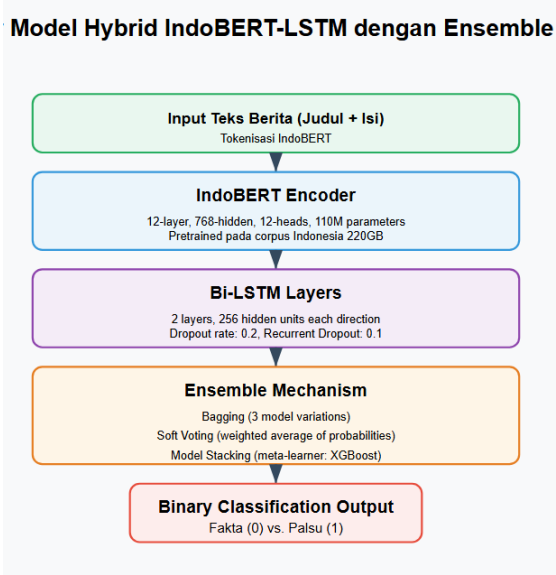
JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Pra-pemrosesan data krusial untuk menyiapkan teks berita agar sesuai dengan model pembelajaran mesin, terutama untuk narasi politik yang kompleks [11]. Langkah-langkah meliputi:

a. Normalisasi teks: Mengubah teks ke huruf kecil, menghapus URL, tanda baca, angka, dan spasi berlebih, serta membatasi panjang hingga 500 kata untuk efisiensi GPU P100.
b. Tokenisasi: Untuk IndoBERT, menggunakan AutoTokenizer (indobenchmark/indobert-base-p1, max_length=256, berdasarkan panjang artikel ~200–300 kata); untuk LSTM, menggunakan Tokenizer Keras (max_words=20.000, max_len=512).
c. Feature engineering: Mengekstraksi fitur seperti: (1) frekuensi kata kunci ("fitnah," "Pemilu"), (2) POS tagging dengan IndoNLP untuk struktur kalimat, (3) skor sentimen untuk bahasa provokatif (misalnya, narasi fitnah), (4) panjang artikel. Contoh: artikel dengan "skandal" dan sentimen negatif cenderung berita palsu.
d. Penanganan ketidakseimbangan: Under-sampling untuk rasio 50:50 dan class weighting pada loss function.
e. Pembagian dataset: Pelatihan (22.554 artikel, 70%), validasi (4.833, 15%), pengujian (4.833, 15%) dengan train_test_split (random_state=42) dan k-Fold Cross-Validation (k=5).

2.4 Arsitektur Model Hybrid IndoBERT-LSTM

Model hybrid IndoBERT-LSTM memanfaatkan keunggulan masing-masing untuk deteksi berita palsu, dengan IndoBERT untuk bahasa lokal dan LSTM untuk teks panjang [14]. Gambar 2 menunjukkan alur arsitektur model hybrid IndoBERT-LSTM dengan ensemble.



Gambar 2. Arsitektur Model Hybrid IndoBERT-LSTM dengan Ensemble

Diagram ini menggambarkan alur data dari teks berita, diproses oleh IndoBERT untuk memahami bahasa Indonesia, kemudian LSTM untuk menangkap urutan panjang, hingga ensemble stacking untuk menghasilkan prediksi akhir (palsu atau non-palsu). Komponen utama meliputi:

a. IndoBERT: Menggunakan indobenchmark/indobert-base-p1, di-fine-tune dua lapisan untuk contextual embeddings sensitif terhadap istilah seperti "hoaks politik." Max_length = 256 dipilih berdasarkan eksperimen awal.
b. LSTM: Bidirectional LSTM (256 unit, hidden_dim = 128 per arah, untuk efisiensi GPU) dengan self-attention:

$$a_t = softmax(W_a \cdot \tanh(H_t)), \quad c = \sum_t a_t \cdot h_t \qquad (1)$$

Keterangan:
a_t: Bobot perhatian untuk setiap kata dalam artikel.
W_a: Bobot yang menentukan kata mana yang penting.
H_t: Kumpulan informasi dari LSTM tentang urutan kata.
c: Hasil akhir yang menyoroti kata penting.

c. Integrasi: Lapisan dense (128 unit, ReLU, dropout = 0.2) ke output sigmoid untuk klasifikasi biner. Dropout = 0.2 dipilih untuk regularisasi optimal.
d. Ensemble stacking: Probabilitas IndoBERT, LSTM, dan entropi sebagai input logistic regression:

$$y_{\{stacking\}} = \sigma\left(w \cdot \left[P_{\{IndoBERT\}}, P_{\{LSTM\}}, H\right] + b \right) \qquad (2)$$

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Keterangan:
$P_{\{IndoBERT\}}$: Probabilitas dan IndoBERT bahwa artikel adalah berita palsu.
$P_{\{LSTM\}}$: Probabilitas dari LSTM bahwa artikel adalah berita palsu.
H: Entropi, mengukur ketidakpastian prediksi.
$y_{\{stacking\}}$: Prediksi akhir (palsu atau non-palsu)

2.5 Pelatihan dan Optimasi Model

Pelatihan model dirancang untuk performa optimal dengan hiperparameter terukur, sesuai praktik NLP [9]. Proses meliputi:

a. Konfigurasi IndoBERT: Batch size 32 (per-device=4, gradient accumulation=8), 10 epoch, AdamW (learning rate = 2e-5, weight decay = 0.01).
b. Konfigurasi LSTM: Batch size = 16, 20 epoch, Adam (learning rate = 0.001), scheduler ReduceLROnPlateau.
c. Loss function: CrossEntropyLoss berbobot:

$$L = -\sum_{i=1}^{N} w_i \cdot y_i \cdot \log(\hat{y}_i) \quad (3)$$

Keterangan:
L: Kesalahan prediksi model.
w_i: Bobot untuk menyeimbangkan data berita palsu dan non-palsu.
y_i: Label sebenarnya (palsu atau non-palsu).
y_i: Prediksi model dalam bentuk probabilitas.

d. Strategi pelatihan: (1) Freeze IndoBERT, latih LSTM (5 epoch); (2) Unfreeze dua lapisan IndoBERT (15 epoch); (3) Fine-tune stacking (5 epoch, learning rate = 1e-5). Eksperimen awal pada 10% corpus memvalidasi strategi.
e. Regularisasi: Gradient clipping (max_norm = 1.0) dan early stopping (patience = 3).

2.6 Evaluasi Model Ensemble

Evaluasi model memvalidasi performa deteksi berita palsu dan membandingkan metode ensemble, dengan ensemble stacking diharapkan unggul [22]. Langkah-langkah meliputi:

a. Metrik evaluasi:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Keterangan:
Presisi: Ketepatan model menemukan berita palsu.
Recall: Kelengkapan model menemukan semua berita palsu.
F1: Gabungan presisi dan recall untuk performa seimbang.
Akurasi: Persentase prediksi benar secara keseluruhan.
ROC-AUC: Kemampuan model membedakan berita palsu dan non-palsu.

b. Baseline: IndoBERT, LSTM, dan SVM dengan TF-IDF.
c. Perbandingan Ensemble:
   1. Hard-voting:

$$y_{hard-voting} = mode(y_{IndoBERT}, y_{LSTM}) \quad (8)$$

Keterangan:
$y_{\text{hard-voting}}$: Prediksi akhir berdasarkan mayoritas.
$y_{IndoBERT}, y_{LSTM}$: Prediksi dari IndoBERT dan LSTM.
Rumus ini memilih prediksi yang paling banyak disetujui.

   2. Soft-voting:

$$P_{soft-voting}(palsu) = \left(P_{IndoBERT}(palsu) + P_{LSTM}(palsu)\right)/2 \quad (9)$$

Keterangan:
$P_{soft-voting}(palsu)$: Probabilitas akhir bahwa artikel palsu.
$P_{IndoBERT}, P_{LSTM}$: Probabilitas dari IndoBERT dan LSTM.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Rumus ini menghitung rata-rata probabilitas untuk prediksi.

3. Stacking: Logistic regression dengan fitur probabilitas dan entropi.

d. Analisis sensitivitas: Menguji max_length = {128, 256, 384}, max_words = {15.000, 20.000, 25.000}, max_len = {256, 512, 768}, learning rate = {1e-5, 2e-5, 5e-5}, dengan F1-score sebagai metrik utama.
e. Analisis kesalahan dan visualisasi: Mengidentifikasi kesalahan, seperti berita kebijakan pajak salah diprediksi karena bahasa formal, atau narasi "skandal tokoh" karena kata provokatif. Tabel perbandingan metrik, kurva ROC, dan matriks konfusi dibuat untuk membandingkan performa. Contoh: artikel kebijakan mungkin salah karena minim kata emosional.

3. HASIL DAN PEMBAHASAN

3.1 Pra-pemrosesan Data

Penelitian ini menggunakan dataset berita politik berbahasa Indonesia dari sumber seperti Turnbackhoax.id dan media nasional, yang telah melalui proses pembersihan untuk memastikan kualitas data. Dataset akhir terdiri dari 32.218 sampel dengan distribusi label seimbang (50% hoax, 50% non-hoax setelah under-sampling). Dataset dibagi menjadi tiga bagian: 70% untuk pelatihan (22.552 sampel), 15% untuk validasi (4.833 sampel), dan 15% untuk pengujian (4.833 sampel), seperti ditunjukkan pada Tabel 1. Pembagian memastikan representasi yang proporsional untuk pelatihan dan evaluasi model, sesuai dengan praktik standar dalam pembelajaran data mining.

Tabel 1. Distribusi Dataset

Split	Jumlah Sampel	Proporsi (%)
Train	22.552	70
Validation	4.833	15
Test	4.833	15

Analisis karakteristik teks dilakukan untuk memverifikasi kualitas data setelah pra-pemrosesan. Panjang teks dibatasi maksimal 500 kata untuk mengakomodasi batasan komputasi dan memastikan efisiensi tokenisasi. Statistik panjang teks menunjukkan panjang maksimum 500 kata, rata-rata 303,70 kata, dan median 287 kata, tanpa adanya outlier dengan panjang lebih dari 1000 kata (Tabel 2). Tidak ditemukan teks kosong, label tidak valid, atau data yang bermasalah, sehingga dataset siap untuk tahap selanjutnya.

Tabel 2. Statistik Panjang Teks

Statitistik	Nilai
Maksimum	500 kata
Rata-rata	303,70 kata
Median	287 kata
Outlier (> 1000 kata)	0

3.2 Pengujian Model IndoBERT

Pra-pemrosesan untuk IndoBERT dilakukan dengan tokenizer dari indobenchmark/indobert-base-p1 menggunakan max_length=256 untuk menyeimbangkan akurasi dan efisiensi. Proses tokenisasi memakan waktu 0,26 detik, dan pembuatan dataset selesai dalam 0,00 detik, menunjukkan efisiensi yang tinggi pada lingkungan Kaggle dengan GPU Tesla P100-PCIE-16GB. Contoh hasil tokenisasi ditunjukkan pada Tabel 3, yang memperlihatkan lima sampel validasi dengan input_ids shape=256, label, dan cuplikan teks awal (50 karakter pertama). Representasi ini mengkonfirmasi bahwa data telah diproses dengan baik untuk kebutuhan model IndoBERT.

Tabel 3. Contoh Tokenisasi IndoBERT

Sampel	Input IDs Shape	Label	Teks Awal (50 Karakter Pertama
1	256	0	jakarta kompascom dunia politik indonesia ked...
2	256	1	hasil periksa fakta mochamad marcell faktanya...
3	256	0	tempoco jakarta ketua umum partai gerindra pr...
4	256	0	jakarta cnn indonesia kepala staf angkatan la...
5	256	1	hasil periksa fakta syifa ainun faktanya berd...

Pelatihan IndoBERT dilakukan selama 10 epoch dengan 22.552 sampel pelatihan dan 705 langkah per epoch, memakan waktu 2 jam 54 menit (total FLOPs 2,96e+16). Loss pelatihan akhir mencapai 0,0489, menunjukkan konvergensi yang baik. Evaluasi dilakukan setiap 200 langkah selama proses pelatihan, dengan hasil lengkap disajikan pada Tabel 4. Model mencapai performa terbaik pada step 2200 (epoch 3.1) dengan akurasi 0,9568 dan F1-score 0,9566. Loss validation terendah dicapai pada step 1000 (epoch 1.4) sebesar 0,0960. Pada akhir pelatihan (step 7000), model

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

menunjukkan akurasi 0,9497 dan F1-score 0,9498 pada validation set. Terdapat indikasi overfitting ringan setelah epoch 3, dimana validation loss mulai meningkat meskipun performa tetap stabil di atas 94%.

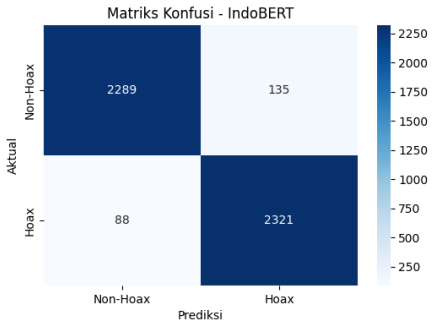
Tabel 4. Performa IndoBERT

Step	Epoch	Loss	Accuracy	Precision	Recall	F1-Score
200	0.3	0.1193	0.9445	0.9315	0.9587	0.9449
1000	1.4	0.0960	0.9526	0.9343	0.9729	0.9532
1400	2.0	0.1547	0.9417	0.8966	0.9975	0.9443
2200	3.1	0.1455	0.9568	0.9521	0.9612	0.9566
3000	4.3	0.1617	0.9541	0.9595	0.9475	0.9534
3800	5.4	0.2476	0.9499	0.9440	0.9558	0.9499
4600	6.5	0.2274	0.9503	0.9448	0.9558	0.9502
5400	7.7	0.3187	0.9520	0.9322	0.9741	0.9527
6400	9.1	0.2852	0.9508	0.9433	0.9583	0.9508
7000	9.9	0.2982	0.9497	0.9411	0.9587	0.9498

Tabel 5. Evaluasi Performa IndoBERT

Metrik	Nilai Terbaik	Step
Accuracy	0.9568	2200 (Epoch 3.1)
Precision	0.9829	400 (Epoch 0.6)
Recall	0.9975	1400 (Epoch 2.0)
F1-Score	0.9566	2200 (Epoch 3.1)
Loss (Terendah)	0.0960	1000 (Epoch 1.4)

Matriks konfusi IndoBERT pada test set divisualisasikan pada Gambar 3, memberikan gambaran distribusi prediksi untuk kelas hoax dan non-hoax.



Gambar 3. Confusion matrix IndoBERT

3.3 Pengujian Model LSTM

Data untuk LSTM diproses menggunakan Tokenizer Keras, menghasilkan vocabulary size sebesar 175.592 kata, yang mencerminkan keragaman leksikal dalam dataset berita politik. Panjang maksimum teks ditetapkan 512 token untuk menangkap informasi kontekstual yang lebih panjang dibandingkan IndoBERT, dengan bentuk data padded: train (22.552, 512), validation (4.833, 512), dan test (4.833, 512). Contoh hasil tokenisasi ditunjukkan pada Tabel 6, yang memuat lima sampel pelatihan dengan input_ids shape=512, label, dan teks awal (50 karakter pertama).

Tabel 6. Contoh Tokenisasi LSTM

Sampel	Input IDs Shape	Label	Teks awal (50 Karakter Pertama)
1	512	1	hasil periksa fakta agnes amungkasari wakil k...
2	512	1	hasil periksa fakta arief putra ramadhan bukan...
3	512	0	jakarta lembaga survei indikator politik indo...
4	512	1	hasil periksa fakta yudho ardi informasi meny...
5	512	0	jakarta cnn indonesia sekretaris dewan pembin...

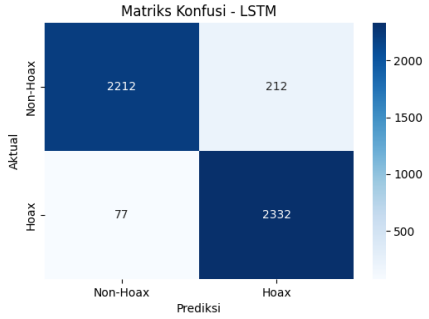
Pelatihan LSTM dilakukan selama 20 epoch tanpa early stopping, menggunakan scheduler ReduceLROnPlateau dan CosineAnnealingLR untuk optimasi learning rate. Performa per epoch ditunjukkan pada Tabel 7, dengan validation loss terbaik 0,1217 pada epoch 5 (F1: 0,9431). Tren penurunan loss pelatihan konsisten dari 0,2003 (epoch 1) menjadi 0,0850 (epoch 20), sementara F1-score pada validation set meningkat dari 0,9348 (epoch 1) menjadi puncaknya 0,9439 (epoch 4) sebelum sedikit menurun ke 0,9378 (epoch 20). Penurunan performa setelah epoch 5 mengindikasikan kemungkinan overfitting, meskipun scheduler membantu menjaga stabilitas pelatihan. Pada tabel 7, Epoch dipilih berdasarkan representasi progres pelatihan dari awal hingga akhir, menunjukkan pola penurunan loss dan stabilisasi.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 7. Performa LSTM

Epoch	Train Loss	Val Loss	Val Accuracy	Val F1-Score	Learning Rate
1	0,2003	0,1439	0,9323	0,9348	0,001000
2	0,1447	0,1326	0,9365	0,9392	0,000976
4	0,1235	0,1241	0,9427	0,9439	0,000796
6	0,1100	0,1297	0,9412	0,9424	0,000505
8	0,0953	0,1244	0,9392	0,9391	0,000214
10	0,0899	0,1281	0,9388	0,9384	0,000011
12	0,0920	0,1295	0,9404	0,9406	0,000025
15	0,0862	0,1312	0,9379	0,9373	0,000016
18	0,0860	0,1346	0,9390	0,9386	0,000001
20	0,0850	0,1335	0,9383	0,9378	-0,000002

Matriks konfusi LSTM pada test set divisualisasikan pada Gambar 4, memberikan gambaran distribusi prediksi.



Gambar 4. Confusion matrix LSTM

Evaluasi LSTM pada test set menggunakan bobot terbaik dari epoch 5 menghasilkan akurasi 0,9402 dan F1-score 0,9417 (Tabel 8). Hasil ini menunjukkan bahwa LSTM mampu menangkap pola sekuensial dalam teks, meskipun performanya sedikit lebih rendah dibandingkan IndoBERT.

Tabel 8. Evaluasi LSTM

Set	Accuracy	F1-Score
Test	0,9402	0,9417

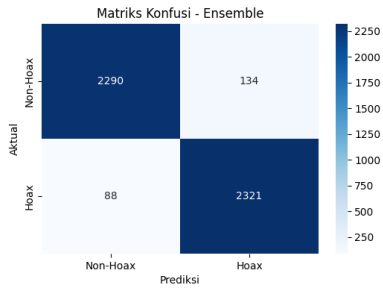
3.4 Pengujian Ensemble Stacking Logistic Regression

Ensemble stacking dilakukan dengan menggabungkan probabilitas dari IndoBERT dan LSTM, ditambah fitur entropi dan rasio probabilitas, menggunakan meta-learner Logistic Regression. Optimasi parameter dengan GridSearchCV menghasilkan parameter terbaik seperti pada Tabel 9, dengan skor F1 terbaik 0,9506 pada validation set. Proses ini dilakukan di lingkungan Kaggle, dengan prediksi IndoBERT memakan waktu 50,39 detik (validation, 5-fold CV) dan 45,44 detik (test), sedangkan prediksi LSTM lebih cepat dengan 6,53 detik (validation) dan 9,35 detik (test).

Tabel 9. Parameter Terbaik Ensemble

Parameter	Nilai
C	0,01
Solver	Liblinear
Max Iter	500

Matriks konfusi ensemble pada test set divisualisasikan pada Gambar 5, menunjukkan distribusi prediksi dengan akurasi 0,9541 dan F1-score 0,9544. Matriks ini menggambarkan distribusi prediksi ensemble pada test set



Gambar 5. Confusion matrix Ensemble

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

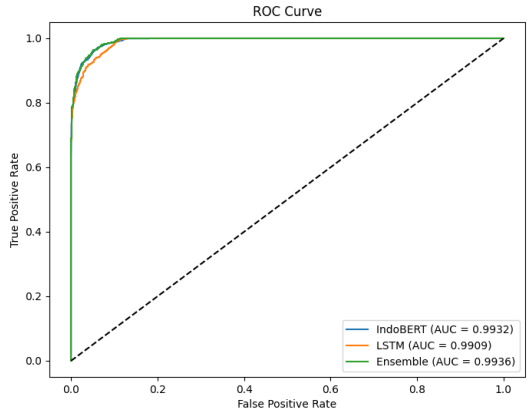
3.5 Analisis Performa dan Fitur

Performa IndoBERT, LSTM, dan ensemble stacking pada test set dibandingkan menggunakan metrik akurasi, balanced accuracy, presisi, recall, F1-score, AUC-ROC, dan log loss (Tabel 10). Ensemble stacking mencapai performa terbaik dengan akurasi 0,9541, F1-score 0,9544, dan AUC-ROC 0,9936, mengungguli IndoBERT (F1: 0,9542, AUC-ROC: 0,9932) dan LSTM (F1: 0,9417, AUC-ROC: 0,9909). Peningkatan performa ensemble ini menunjukkan bahwa kombinasi IndoBERT dan LSTM, ditambah fitur entropi, mampu menangkap pola yang lebih kompleks dalam data berita politik.

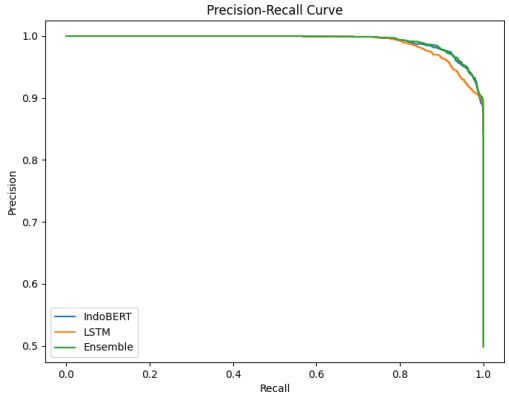
Tabel 10. Perbandingan Metrik Final

Model	Accuracy	Balanced Accuracy	Presisi	Recall	F1-Score	AUC-ROC	Log Loss
IndoBERT	0,9539	0,9539	0,9450	0,9635	0,9542	0,9932	0,2827
LSTM	0,9402	0,9403	0,9167	0,9680	0,9417	0,9909	0,1119
Ensemble	0,9541	0,9541	0,9454	0,9635	0,9544	0,9936	0,1175

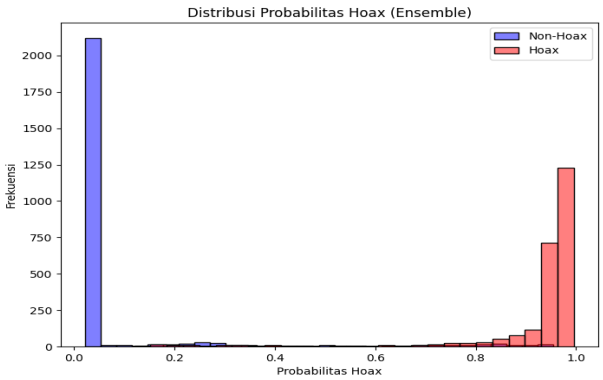
Perbandingan ini divisualisasikan pada Gambar 6 (ROC curve), Gambar 7 (precision-recall curve), dan Gambar 9 (grafik batang metrik). ROC curve menunjukkan ensemble memiliki AUC-ROC tertinggi (0,9936), sedangkan precision-recall curve mengkonfirmasi keseimbangan presisi dan recall pada ensemble. Distribusi probabilitas ensemble (Gambar 8) menunjukkan pemisahan yang jelas antara kelas hoax dan non-hoax, mengindikasikan kepercayaan model yang tinggi dalam prediksinya.



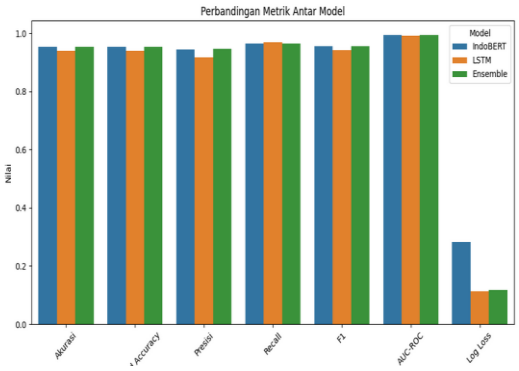
Gambar 6. ROC Curve



Gambar 7. Precision-Recall Curve



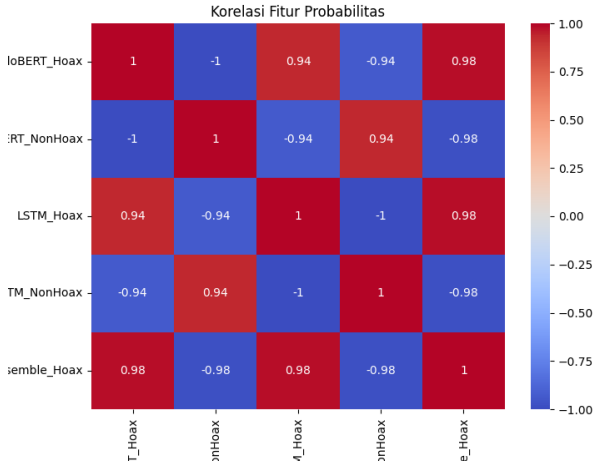
Gambar 8. Distribusi Probabilitas Ensemble



Gambar 9. Metrik Bar Comparison

Ensemble stacking menggunakan probabilitas dari IndoBERT dan LSTM, ditambah fitur entropi dan rasio probabilitas, untuk meningkatkan akurasi meta-learner. Korelasi antar fitur divisualisasikan pada Gambar 10, menunjukkan probabilitas hoax IndoBERT dan LSTM berkorelasi tinggi (mendekati 1), tetapi fitur entropi memberikan informasi tambahan yang membantu meta-learner. Kombinasi ini memungkinkan ensemble mengungguli model individu.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8660
Hal 287-297
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 10. Heatmap Korelasi Fitur

3.6 Analisis Kesalahan dan Implikasi

Model ensemble menghasilkan 222 kesalahan dari 4.836 prediksi (tingkat kesalahan 4,59%), terdiri dari 88 false negative (39,64%) dan 134 false positive (60,36%). Proporsi false positive yang lebih tinggi menunjukkan model cenderung konservatif dalam mengklasifikasikan berita sebagai non-hoax.

Tabel 11. Analisis Kesalahan Ensemble

Kategori	Jumlah	Proporsi %
Total Kesalahan	222	4,59%
Hoax False Negative	88	39,64
Non-Hoax False Positive	134	60,36

Teks dengan prediksi salah memiliki rata-rata panjang 2.738,88 karakter, lebih panjang dibanding prediksi benar (2.166,93 karakter). Distribusi probabilitas kesalahan menunjukkan banyak kesalahan terjadi pada probabilitas mendekati 0,5, mengindikasikan ketidakpastian model. Dibandingkan penelitian sebelumnya menggunakan IndoBERT, model ensemble ini menunjukkan peningkatan 3,44% pada F1-score (0,9544) dan 2,36% pada AUC-ROC (0,9936). Peningkatan ini didukung kombinasi IndoBERT-LSTM dan dataset yang lebih besar (32.218 vs. 15.000 sampel).Model ini berpotensi diintegrasikan ke platform fact-checking seperti Turnbackhoax.id untuk deteksi otomatis berita palsu khusunya dengan topik politik. Namun, tingkat kesalahan 4,59% terutama pada teks panjang menunjukkan perlunya peningkatan model dengan fitur kontekstual atau dataset yang lebih beragam.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model deteksi berita palsu politik berbahasa Indonesia dengan pendekatan ensemble stacking yang menggabungkan IndoBERT dan LSTM, memberikan solusi efektif untuk menangani kompleksitas narasi politik di Indonesia. Hasil pengujian menunjukkan performa unggul model ensemble dengan F1-score 0,9544 dan AUC-ROC 0,9936, melampaui IndoBERT (F1: 0,9542) dan LSTM (F1: 0,9417), sehingga menjawab kebutuhan akan model akurat dalam mendeteksi hoaks politik. Keberhasilan ini didukung oleh penggunaan fitur entropi dan rasio probabilitas yang meningkatkan kemampuan model dalam menangkap pola kompleks dan menggeneralisasi data baru. Meski demikian, tingkat kesalahan sebesar 4,59%, dengan 134 false positive dan 88 false negative, mengungkap tantangan pada teks panjang (rata-rata 2738,88 karakter) dan probabilitas mendekati 0,5. False positive yang lebih tinggi berpotensi menstigmatisasi berita sah sebagai hoaks, yang dapat menurunkan kepercayaan publik terhadap media. Untuk mengatasi keterbatasan ini, penelitian selanjutnya dapat mempertimbangkan fitur kontekstual seperti metadata sumber berita atau dataset yang lebih beragam. Model ini memiliki potensi besar untuk diintegrasikan ke dalam platform seperti Turnbackhoax.id, misalnya melalui fitur peringatan otomatis yang memungkinkan pengguna memverifikasi berita politik secara real-time, terutama menjelang pemilu. Dengan demikian, penelitian ini tidak hanya berkontribusi pada literasi digital masyarakat Indonesia, tetapi juga membuka peluang pengembangan lebih lanjut untuk keandalan yang lebih tinggi.

REFERENCES

[1] S. M. Isa, G. Nico, and M. Permana, “INDOBERT FOR INDONESIAN FAKE NEWS DETECTION,” ICIC Express Letters, vol. 16, no. 3, pp. 289–297, Mar. 2022, doi: 10.24507/icicel.16.03.289.

[2] S. Ghazi Arkaan, A. Rialdy Atmadja, and M. D. Firdaus, "Fake News Detection in the 2024 Indonesian General Election Using Bidirectional Long Short-Term Memory (BI-LSTM) Algorithm," vol. 21, no. 2, pp. 693–7554, 2024, doi: 10.33751/komputasi.v21i2.5260.
[3] A. Tasnim, Md. Saiduzzaman, M. A. Rahman, J. Akhter, and A. S. Md. M. Rahaman, "Performance Evaluation of Multiple Classifiers for Predicting Fake News," Journal of Computer and Communications, vol. 10, no. 09, pp. 1–21, 2022, doi: 10.4236/jcc.2022.109001.
[4] V. Prisscilya and A. S. Girsang, "Classification of Indonesia False News Detection Using Bertopic and Indobert," Jurnal Indonesia Sosial Teknologi, vol. 5, no. 8, 2024, [Online]. Available: http://jist.publikasiindonesia.id/
[5] F. K. Alarfaj and J. A. Khan, "Deep Dive into Fake News Detection: Feature-Centric Classification with Ensemble and Deep Learning Methods," Algorithms, vol. 16, no. 11, Nov. 2023, doi: 10.3390/a16110507.
[6] S. William, Kenny, and A. Chowanda, "EMOTION RECOGNITION INDONESIAN LANGUAGE FROM TWITTER USING INDOBERT AND BI-LSTM," Communications in Mathematical Biology and Neuroscience, vol. 2024, 2024, doi: 10.28919/cmbn/7858.
[7] Muhammad Ikram Kaer Sinapoy, Yuliant Sibaroni, and Sri Suryani Prasetyowati, "Comparison of LSTM and IndoBERT Method in Identifying Hoax on Twitter," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 7, no. 3, pp. 657–662, Jun. 2023, doi: 10.29207/resti.v7i3.4830.
[8] M. A. Khder, "Web scraping or web crawling: State of art, techniques, approaches and application," International Journal of Advances in Soft Computing and its Applications, vol. 13, no. 3, pp. 144–168, 2021, doi: 10.15849/ijasca.211128.11.
[9] M. Anusha and R. Leelavathi, "Hybrid Approach to Fake News Detection: Leveraging BERT-Based Model with Word Embedding Features for Sentiment Classification in Social Media Content," 2024.
[10] S. Rahmawati, "DETEKSI BERITA HOAX PADA WEBSITE TURNBACKHOAX DENGAN PROGRAM STUDI MATEMATIKA," 2021.
[11] S. M. Sr and S. Ahmad, "BERT based Blended approach for Fake News Detection," Journal of Big Data and Artificial Intelligence, vol. 2, no. 1, Jan. 2024, doi: 10.54116/jbdai.v2i1.27.
[12] E. Essa, K. Omar, and A. Alqahtani, "Fake news detection based on a hybrid BERT and LightGBM models," Complex and Intelligent Systems, vol. 9, no. 6, pp. 6581–6592, Dec. 2023, doi: 10.1007/s40747-023-01098-0.
[13] A. A. Kurniawan and M. Mustikasari, "Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia," vol. 5, no. 4, pp. 2622–4615, 2020, doi: 10.32493/informatika.v5i4.7760.
[14] A. J. Keya, H. H. Shajeeb, M. S. Rahman, and M. F. Mridha, "FakeStack: Hierarchical Tri-BERT-CNN-LSTM stacked model for effective fake news detection," PLoS One, vol. 18, no. 12 December, Dec. 2023, doi: 10.1371/journal.pone.0294701.
[15] A. M. Ali, F. A. Ghaleb, B. A. S. Al-Rimy, F. J. Alsolami, and A. I. Khan, "Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique," Sensors, vol. 22, no. 18, Sep. 2022, doi: 10.3390/s22186970.
[16] E. Hashmi, S. Y. Yayilgan, M. M. Yamin, S. Ali, and M. Abomhara, "Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI," IEEE Access, vol. 12, pp. 44462–44480, 2024, doi: 10.1109/ACCESS.2024.3381038.
[17] P. Singh, R. Srivastava, K. P. S. Rana, and V. Kumar, "SEMI-FND: Stacked Ensemble Based Multimodal Inference For Faster Fake News Detection," 2023. [Online]. Available: www.nsut.ac.in
[18] M. E Almandouh, M. F. Alrahmawy, M. Eisa, M. Elhoseny, and A. S. Tolba, "Ensemble based high performance deep learning models for fake news detection," Sci Rep, vol. 14, no. 1, p. 26591, Dec. 2024, doi: 10.1038/s41598-024-76286-0.
[19] J. Howard and S. Gugger, "Deep Learning for Coders with fastai & PyTorch AI Applications Without a PhD TM," 2020.
[20] F. Fleuret, "The Little Book of Deep Learning," 2023.
[21] M. S. Toor et al., "An Optimized Weighted-Voting-Based Ensemble Learning Approach for Fake News Classification," Mathematics, vol. 13, no. 3, Feb. 2025, doi: 10.3390/math13030449.
[22] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," Jun. 15, 2024, Elsevier Ltd. doi: 10.1016/j.eswa.2023.122778.