Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

Deteksi Hoaks Berita Berbahasa Indonesia Menggunakan IndoBERT dengan Penanganan Ketidakseimbangan Kelas Berbasis Gabungan Class Weighting dan Focal Loss

Riadhul Muttaqin[1*], Muhammad Edya Rosadi[2], Muhammad Iqbal Firdaus[3], Dian Agustini[4]
[1]Program Studi Teknik Industri, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin
[2,3,4]Program Studi Teknik Informatika, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin
*riadhul_muttaqin@uniska-bjm.ac.id

Abstrak

Penyebaran hoaks dalam ruang digital berbahasa Indonesia meningkat signifikan dalam lima tahun terakhir dengan dampak sosial luas. Sebagian besar penelitian deteksi hoaks berbahasa Indonesia berfokus pada arsitektur model, sementara ketidakseimbangan kelas pada data lapangan kurang diperhatikan. Penelitian ini menyajikan pendekatan terpusat pada data yang menggabungkan IndoBERT dengan strategi penanganan ketidakseimbangan kelas berbasis gabungan class weighting dan focal loss. Dataset yang digunakan adalah korpus indonesiafalsenews berisi 4.231 artikel terlabel dengan rasio kelas hoaks dan fakta mendekati 4,5:1. Evaluasi dilakukan dengan lima seed serta pengujian signifikansi melalui uji McNemar dan paired bootstrap. Konfigurasi terbaik mencapai F1-macro 0,7240, akurasi 0,8392, ROC-AUC 0,8083, dan koefisien korelasi Matthews 0,4510, serta secara statistik mengungguli seluruh basis klasik berbasis TF-IDF dengan nilai p kurang dari 0,05. Penelitian ini juga menunjukkan augmentasi teks tidak memberikan tambahan kinerja yang konsisten, sehingga penanganan ketidakseimbangan kelas menjadi faktor pengungkit yang lebih efektif untuk deteksi hoaks berbahasa Indonesia

Kata kunci : deteksi hoaks, focal loss, IndoBERT, ketidakseimbangan kelas, pemrosesan bahasa alami.

Abstract

The spread of hoaxes in Indonesian-language digital media has risen sharply in the past five years with wide societal harm. Most prior work on Indonesian hoax detection emphasizes model architecture, while the class-imbalance problem in field data receives less attention. This study presents a data-centric approach combining the IndoBERT pretrained language model with a hybrid class-imbalance objective of class weighting and focal loss. The dataset is the indonesiafalsenews corpus of 4231 labelled articles with an approximate 4.5 to 1 hoax-to-fact ratio. Evaluation uses five-seed runs with McNemar and paired-bootstrap significance testing. The proposed configuration achieves an F1-macro of 0.7240, accuracy of 0.8392, ROC-AUC of 0.8083, and a Matthews correlation coefficient of 0.4510, significantly outperforming every TF-IDF baseline at the 0.05 level. Text augmentation does not provide consistent gains, which implies that imbalance handling is the more effective lever for Indonesian hoax detection..

Keywords : *class imbalance, focal loss, hoax detection, IndoBERT, natural language processing.*

1. Pendahuluan

Hoaks dalam ruang digital menjadi tantangan informasi paling mendesak satu dekade terakhir. Kementerian Komunikasi dan Informatika mencatat lebih dari sebelas ribu hoaks teridentifikasi pada rentang 2018 hingga 2023 seputar kesehatan, politik elektoral, dan keagamaan, sehingga sistem deteksi otomatis menjadi prioritas riset untuk mendukung pemeriksa fakta. Model berbasis transformer telah meraih kinerja kompetitif pada klasifikasi teks, termasuk klasifikasi hoaks lintas bahasa[1].

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

Pada bahasa Indonesia, IndoBERT yang dilatih pada korpus berskala besar menyediakan representasi kontekstual yang kuat[2], [3] dan terbukti pada beragam tugas hilir. Namun capaiannya pada data hoaks masih menunjukkan disparitas antara metrik agregat dan kinerja per kelas ketika rasio kelas timpang[4]. Ketidakseimbangan kelas tidak dapat diabaikan karena salah klasifikasi pada kelas fakta yang minoritas berdampak besar, sementara korpus hoaks di lapangan cenderung timpang karena terkumpul dari pelaporan yang lebih banyak memuat hoaks daripada fakta[5]. Akibatnya, model tanpa pertimbangan distribusi kelas bias ke mayoritas dengan akurasi tinggi yang rapuh per kelas.

Komunitas pemrosesan bahasa alami menempatkan paradigma terpusat pada data sebagai pelengkap penting paradigma terpusat pada model[6], [7], karena perlakuan data seperti augmentasi maupun penyesuaian fungsi kerugian dapat berkontribusi melebihi modifikasi arsitektur[8]. Pada teks berbahasa Indonesia, kajian sistematis mengenai kombinasi strategi penanganan ketidakseimbangan kelas masih sangat terbatas.

Kesenjangan ini menyisakan beberapa persoalan. Pertama, penelitian sebelumnya berhenti pada metrik agregat tanpa menelaah sistematis bagaimana ketidakseimbangan kelas menekan kinerja kelas minoritas. Kedua, strategi penanganan ketidakseimbangan umumnya diuji terpisah dan di luar teks berbahasa Indonesia, sehingga belum ada perbandingan strategi tingkat data dan algoritma pada satu basis eksperimen setara. Ketiga, banyak studi tidak menyertakan pengujian signifikansi statistik. Persoalan ini mendesak karena sistem yang bias ke mayoritas dapat menandai informasi benar sebagai hoaks dan menurunkan kepercayaan pada pemeriksaan fakta.

Penelitian ini berkontribusi dalam tiga aspek. Pertama, studi empiris atas tujuh metode augmentasi dan delapan strategi penanganan ketidakseimbangan kelas pada satu basis model yang sama sehingga perbandingannya adil. Kedua, konfigurasi terpadu berbasis IndoBERT dengan gabungan class weighting dan focal loss yang dirumuskan secara eksplisit. Ketiga, analisis signifikansi statistik melalui uji McNemar dan paired bootstrap pada lima seed. Kebaruannya terletak pada penyatuan ketiga aspek tersebut dalam satu studi deteksi hoaks berbahasa Indonesia.

2. Tinjauan Pustaka

Kajian pustaka pada bagian ini terbagi dua. Subbab 2.1 menyajikan penelitian terkait dalam lima tahun terakhir beserta celahnya, sedangkan Subbab 2.2 menguraikan landasan teori mengenai model bahasa praterlatih, penanganan ketidakseimbangan kelas, augmentasi teks, dan

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

pengujian signifikansi yang mendasari metode pada Bagian 3.

2.1. Penelitian Terkait

Studi deteksi hoaks berbahasa Indonesia berkembang seiring tersedianya korpus berlabel dan model praterlatih khusus. Ridho dan Yulianti membandingkan fine-tuning IndoBERT dengan sejumlah model pembelajaran mesin berbasis TF-IDF pada klasifikasi berita hoaks dan menemukan IndoBERT unggul secara konsisten[4], namun evaluasinya bertumpu pada metrik agregat tanpa menelaah kinerja per kelas pada data timpang. Sinapoy dkk. melaporkan IndoBERT lebih baik daripada long short-term memory pada deteksi hoaks media sosial[9], tetapi tanpa menyertakan strategi penanganan ketidakseimbangan kelas maupun uji signifikansi statistik. Yefferson dkk. menunjukkan model hibrida IndoBERT dengan long short-term memory mengungguli model tunggal[5], meskipun fokusnya tetap pada modifikasi arsitektur sehingga ketimpangan distribusi kelas tidak ditangani secara eksplisit.

Pada sisi strategi data, Muftie dan Haris menunjukkan augmentasi kontekstual berbasis IndoBERT lebih stabil daripada metode tingkat token pada klasifikasi teks berbahasa Indonesia[10], namun dampaknya pada korpus hoaks dengan distribusi kelas timpang belum diuji. Tian dkk. mengusulkan oversampling semantik dengan kendala mutual information untuk klasifikasi teks tidak seimbang[11], tetapi pengujiannya terbatas pada korpus berbahasa Inggris. Kelima studi tersebut menegaskan belum adanya perbandingan sistematis strategi penanganan ketidakseimbangan kelas untuk deteksi hoaks berbahasa Indonesia pada satu basis eksperimen yang setara, dan celah itulah yang diisi penelitian ini.

2.2. Landasan Teori

1. Model bahasa praterlatih IndoBERT

IndoBERT diperkenalkan pada IndoLEM sebagai model praterlatih khusus bahasa Indonesia[2] dan diperluas melalui IndoNLU dengan korpus serta tugas evaluasi yang lebih beragam[3]. Tantangan utamanya adalah keragaman dialek dan minimnya korpus berlabel domain spesifik, sehingga adaptasi domain menjadi langkah yang lazim ditempuh.

2. Penanganan ketidakseimbangan kelas

Penanganan ketidakseimbangan kelas dapat dikelompokkan menjadi pendekatan tingkat data dan tingkat algoritma. Pada tingkat data, random oversampling cenderung memperbesar risiko overfitting sementara random undersampling dapat menghilangkan informasi penting kelas mayoritas[11].

Pada tingkat algoritma, class weighting menyesuaikan kontribusi loss tiap kelas terhadap kebalikan frekuensinya[12], focal loss menambah faktor pelemahan bagi sampel mudah agar perhatian terarah ke kelas minoritas[13], class

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

balanced loss menggunakan jumlah efektif sampel[14], dan LDAM loss memberi margin berbeda antar kelas[15].

3. Augmentasi teks

Augmentasi teks memperbesar variasi pelatihan tanpa biaya pelabelan tambahan, mencakup metode tingkat token dan tingkat semantik seperti back translation dan augmentasi kontekstual berbasis masked language model[7]. Keuntungannya sangat bergantung pada karakteristik data dan kapasitas model[16], dan pada bahasa Indonesia metode kontekstual berbasis IndoBERT cenderung lebih stabil daripada metode tingkat token sederhana[10].

4. Pengujian signifikansi statistik

Pengujian signifikansi diperlukan agar kesimpulan tidak bertumpu pada satu run pelatihan. Uji McNemar membandingkan dua model pada test set yang sama dengan memperhatikan ketidaksetujuan prediksi[17], sedangkan paired bootstrap pada F1-macro memberi estimasi interval kepercayaan empiris tanpa asumsi distribusi[18]. Pelaporan multi seed dengan standar deviasi juga dianjurkan agar varians inisialisasi terpisah dari perbedaan antar metode[19].

3. Metode Penelitian

3.1. Tahapan Penelitian

Penelitian dilaksanakan dalam tiga tahap sebagaimana ditunjukkan pada Gambar 1. Tahap pertama adalah penyiapan data, mencakup pemuatan korpus, pra-pemrosesan, pembagian stratified, dan penyiapan lima seed. Tahap kedua menjalankan empat ablasi pada basis model yang sama, yaitu eksplorasi data awal, basis TF-IDF, ablasi augmentasi, dan ablasi strategi ketidakseimbangan kelas. Tahap ketiga menetapkan konfigurasi proposed dari hasil ablasi, membandingkannya dengan seluruh basis melalui uji signifikansi, dan melakukan analisis kekeliruan.

Gambar 1. Tahapan penelitian.

3.2. Instrumen Penelitian

Eksperimen diimplementasikan dalam bahasa Python dan dijalankan pada komputer dengan akselerator GPU. Pustaka PyTorch dan Transformers digunakan untuk pelatihan IndoBERT, scikit-learn untuk basis TF-IDF dan perhitungan metrik, serta SciPy untuk pengujian statistik. Instrumen utamanya adalah model

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

IndoBERT-base sebagaimana diuraikan pada Subbab 3.4.

3.3. Metode Pengumpulan Data

1. Sumber dan karakteristik data

Penelitian ini menggunakan korpus indonesiafalsenews dari platform Kaggle yang berisi 4.231 artikel berbahasa Indonesia dari pemeriksa fakta nasional, dengan label hoaks (satu) dan fakta (nol). Hanya kolom judul dan narasi yang digunakan, digabung dengan pemisah spasi karena judul kerap memuat indikator emosional sedangkan narasi memberi konteks. Statistik korpus disajikan pada Tabel 1.

Tabel 1. Statistik korpus indonesiafalsenews.

Atribut	Nilai
Jumlah sampel total	4.231
Jumlah sampel hoaks	3.465
Jumlah sampel fakta	766
Rasio ketidakseimbangan	4,52 : 1
Panjang teks rata-rata (karakter)	238
Panjang teks minimum (karakter)	23
Panjang teks maksimum (karakter)	1.471

Tabel 1 memperlihatkan korpus yang timpang dengan rasio sekitar 4,5 banding 1 serta panjang teks yang sangat bervariasi, sehingga menjadi pertimbangan dalam pemilihan panjang token maksimum.

2. Pra-pemrosesan dan pembagian data

Pra-pemrosesan meliputi normalisasi karakter, penghapusan URL dan tag HTML, perapian spasi, dan pembersihan teks kosong, menyisakan 4.228 sampel valid. Data dibagi secara stratified menjadi 70 persen pelatihan, 10 persen validasi, dan 20 persen pengujian, sebagaimana Tabel 2.

Tabel 2. Distribusi sampel pada subset pelatihan, validasi, dan pengujian.

Subset	Total	Hoaks	Fakta	Proporsi minoritas
Pelatihan	2.959	2.424	535	0,181
Validasi	423	346	77	0,182
Pengujian	846	693	153	0,181

Stratifikasi menjaga proporsi kelas tiap subset hampir identik dengan korpus penuh, dan pembagian diulang untuk lima seed guna estimasi varians yang andal.

3.4. Teknik Analisis Data

Analisis data mencakup enam komponen berikut.

1. Arsitektur model

Model dasar yang digunakan adalah IndoBERT versi base dengan dua belas lapisan encoder, 768 dimensi tersembunyi, dan 110 juta parameter, sebagaimana dirilis oleh Koto dkk. melalui repositori IndoLEM[2].

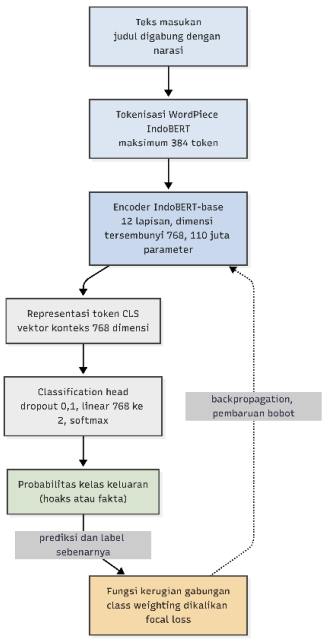
Representasi token klasifikasi dipakai sebagai vektor agregat dokumen, di atasnya ditambahkan dropout 0,1, lapisan linear dari 768 dimensi ke jumlah kelas, dan softmax. Bobot diperbarui melalui propagasi balik yang digerakkan fungsi kerugian gabungan pada Persamaan (3). Arsitektur ringkas konfigurasi proposed ditunjukkan pada Gambar 2.

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

IndoBERT dengan panjang maksimum 384 token, dipilih dari sebaran panjang teks agar minimal 95 persen teks tidak terpotong.



Gambar 2. Arsitektur model usulan

2. Strategi penanganan ketidakseimbangan kelas

Penelitian ini menelaah delapan strategi penanganan ketidakseimbangan kelas, yaitu class weighting, weighted sampler, random oversampling, random undersampling, focal loss, class balanced loss, LDAM loss, dan strategi gabungan, yang seluruhnya dibandingkan terhadap kondisi tanpa perlakuan sebagai basis. Notasi ny menyatakan jumlah sampel kelas y, N jumlah seluruh sampel pelatihan, py probabilitas prediksi kelas y, dan C jumlah kelas.

Class weightin atau bobot kelas dihitung sebagai kebalikan frekuensi yang dinormalisasi. Bentuknya ditunjukkan pada Persamaan (1).

$$w_y = N / (C \cdot n_y) \quad (1)$$

Dengan demikian, sampel kelas minoritas mendapat bobot lebih besar; pada korpus ini bobot kelas fakta sekitar 2,76 dan hoaks sekitar 0,61.

Focal loss memperkenalkan faktor pelemahan untuk sampel yang mudah diklasifikasikan, sebagaimana diusulkan oleh Lin dkk.[13]. Persamaan focal loss adalah seperti pada Persamaan (2).

$$FL(p_y) = -\alpha_y (1 - p_y)^{\gamma} \log(p_y) \quad (2)$$

Pada persamaan tersebut, α_y adalah faktor penyeimbang kelas, γ adalah parameter fokus yang mengatur seberapa kuat pelemahan diterapkan, dan p_y adalah probabilitas yang diprediksi model untuk label sebenarnya. Pada penelitian ini, $\alpha_y = 0{,}25$ untuk kelas hoaks dan satu dikurangi nilai tersebut untuk kelas fakta, sedangkan γ diatur pada nilai dua.

Pada strategi gabungan (model yang diusulkan), dua mekanisme dipadukan, yaitu class weighting pada faktor α_y dari Persamaan (2) dan focal loss pada faktor pelemahan dengan γ sama dengan dua. Bentuk akhirnya ditunjukkan pada Persamaan (3).

$$L_{proposed} = -w_y (1 - p_y)^{\gamma} \log(p_y) \quad (3)$$

Pada persamaan tersebut, w_y diperoleh dari Persamaan (1) sedangkan eksponen pelemahan diatur sama dengan dua.

3. Augmentasi teks sebagai pembanding

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

Sebagai pembanding, tujuh metode augmentasi dievaluasi pada konfigurasi terbaik IndoBERT, yaitu penggantian sinonim, penyisipan, penghapusan, dan penukaran token, augmentasi kontekstual berbasis masked language model, back translation lewat pivot bahasa Inggris, serta tanpa augmentasi sebagai referensi, semuanya diterapkan hanya pada kelas minoritas dengan rasio satu.

4. Prosedur pelatihan

Pelatihan menggunakan AdamW dengan ukuran batch efektif enam belas hasil akumulasi gradien dua langkah dari batch delapan. Pelatihan berjalan hingga sepuluh epoch dengan penghentian dini berbasis F1-macro validasi dan kesabaran tiga epoch serta autocast presisi campuran. Setiap konfigurasi dijalankan pada lima seed, yaitu 7, 42, 101, 1234, dan 1137, dengan metrik dilaporkan sebagai rata-rata.

Nilai hyperparameter tidak ditetapkan lewat pencarian otomatis seperti grid search, random search, maupun optimasi Bayesian, melainkan mengikuti nilai baku literatur penyetelan model berbasis BERT dan rekomendasi publikasi asli tiap fungsi kerugian karena ruang eksperimen sudah luas. Laju pembelajaran 2 kali sepuluh pangkat minus lima, peluruhan bobot 0,01, pemanasan linier sepuluh persen, dan dropout 0,1 mengikuti konvensi; panjang token 384 mengikuti sebaran panjang teks pada Tabel 1; batch efektif enam belas menyesuaikan memori GPU; gamma dua, alpha 0,25, beta 0,999, margin maksimum 0,5, dan skala 30 mengikuti rekomendasi pengusulnya. Konfigurasi terbaik dipilih pada tingkat strategi melalui ablasi terkendali berkriteria F1-macro, bukan penalaan hyperparameter individual.

5. Metrik evaluasi

Evaluasi memakai beberapa metrik komplementer. Akurasi cenderung menyesatkan pada data tidak seimbang sehingga metrik utama adalah F1-macro, yaitu rata-rata aritmatika F1 tiap kelas. Sebagai pelengkap, koefisien korelasi Matthews juga dilaporkan[20]. Selain itu, ROC-AUC dan PR-AUC dihitung untuk mengevaluasi kualitas peringkat probabilitas, sedangkan presisi dan recall per kelas membedakan kinerja hoaks dan fakta.

6. Uji signifikansi

Perbandingan antar model memakai uji McNemar yang memperhatikan sampel yang hanya salah pada satu model[17] serta paired bootstrap dengan seribu pengambilan ulang pada test set untuk estimasi delta F1-macro beserta interval kepercayaan 95 persen[18]. Perbedaan dianggap signifikan bila nilai p bootstrap kurang dari 0,05 dan interval tidak mencakup nol.

3.5. Lokasi Penelitian

Penelitian komputasional ini dilaksanakan di Program Studi Teknik Informatika, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

Banjarmasin, dengan data sekunder yang diperoleh secara daring dari platform Kaggle sebagaimana diuraikan pada Subbab 3.3.

4. Hasil dan Pembahasan

4.1 Hasil Penelitian

1. Eksplorasi data awal

Eksplorasi data menunjukkan kelas hoaks mendominasi seluruh subset sekitar 81,9 persen berbanding 18,1 persen fakta, dengan distribusi panjang teks berekor kanan bermodus 60 sampai 200 karakter. Sebagian kecil sampel melebihi 1.000 karakter sehingga panjang maksimum 384 token sudah memadai. Kata kunci yang sering muncul pada kelas hoaks didominasi istilah pandemi, politik elektoral, dan tokoh publik, sementara kelas fakta lebih beragam karena bersumber dari rilis resmi lembaga.

2. Kinerja basis TF-IDF

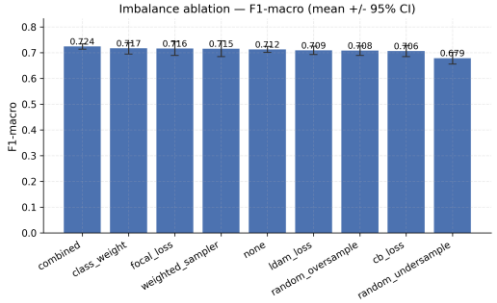
Empat klasifikator berbasis TF-IDF, yaitu Logistic Regression, Linear SVM, Random Forest, dan XGBoost, dijalankan sebagai pembanding klasik dengan class_weight balanced agar setara dengan strategi penanganan ketidakseimbangan pada IndoBERT. Kinerja keempatnya disajikan bersama seluruh model pada Tabel 3 di bagian perbandingan penuh. Logistic Regression terbaik pada F1-macro di antara basis TF-IDF, XGBoost berakurasi tertinggi namun recall faktanya rendah, dan Random Forest memiliki MCC terendah yang menandakan bias ke kelas mayoritas. Hal ini menegaskan akurasi semata tidak memadai dan basis TF-IDF kesulitan menangkap pola kelas minoritas.

3. Ablasi augmentasi

Ablasi augmentasi dijalankan pada IndoBERT-base dengan strategi ketidakseimbangan dimatikan agar efek murni augmentasi teramati. Dari tujuh metode yang diuji, hanya augmentasi kontekstual yang sedikit mengungguli kondisi tanpa augmentasi (0,7164 berbanding 0,7124, selisih 0,004 masih dalam interval kepercayaan), sedangkan lima metode lain justru menurunkan kinerja, dengan penukaran token sebagai yang terburuk pada 0,6788. Hasil ini sejalan dengan temuan bahwa keuntungan augmentasi sangat sensitif terhadap karakteristik data dan model[16], sehingga augmentasi belum memberi tambahan kinerja yang konsisten pada konfigurasi ini.

4. Ablasi penanganan ketidakseimbangan kelas

Ablasi penanganan ketidakseimbangan kelas dilakukan dengan augmentasi dimatikan agar efek murni masing-masing strategi terukur. Peringkat F1-macro kedelapan strategi beserta kondisi tanpa perlakuan ditunjukkan pada Gambar 3.



Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

Gambar 3. Perbandingan F1-macro delapan strategi penanganan ketidakseimbangan kelas dan kondisi tanpa perlakuan.

Strategi gabungan yang diusulkan menempati peringkat F1-macro tertinggi di antara seluruh strategi berperlakuan, dengan recall kelas fakta naik dari 0,386 tanpa perlakuan menjadi 0,461 dengan biaya kecil pada recall hoaks. Random undersampling menaikkan recall fakta hingga 0,541, namun menurunkan akurasi karena banyak sampel mayoritas dibuang[11].

5. Perbandingan penuh dan uji signifikansi

Konfigurasi proposed mengungguli semua pembanding pada F1-macro dan MCC sekaligus mempertahankan akurasi kompetitif. Recall hoaks proposed 0,907 sedikit di bawah basis tanpa perlakuan 0,930, tetapi diimbangi kenaikan recall fakta yang lebih besar sehingga F1-macro dan MCC bergerak positif. Tabel 3 menyandingkan proposed dengan seluruh basis pada metrik utama. Uji paired bootstrap menunjukkan keunggulan proposed signifikan pada taraf 0,05 terhadap empat dari enam pembanding. Keunggulan terhadap Logistic Regression berada pada batas, sedangkan terhadap basis tanpa perlakuan tidak signifikan secara statistik. Meski demikian, proposed memiliki MCC dan recall fakta lebih tinggi sehingga tetap layak direkomendasikan untuk konteks yang memprioritaskan perlindungan kelas fakta.

Tabel 3. Perbandingan akhir antara proposed dan seluruh basis.

Model	F1-macro	Akurasi	MCC	ROC-AUC	PR-AUC
Proposed (IndoBERT + gabungan)	**0,7240 ± 0,0118**	**0,8392**	**0,4510**	**0,8083**	**0,9388**
IndoBERT + augmentasi kontekstual	0,7164 ± 0,0311	0,8364	0,4378	0,7986	0,9353
IndoBERT tanpa perlakuan	0,7124 ± 0,0133	0,8449	0,4323	0,7936	0,9329
Logistic Regression	0,6878 ± 0,0135	0,8118	0,3764	0,7836	0,9328
Linear SVM	0,6708 ± 0,0107	0,8210	0,3471	0,7660	0,9269
XGBoost	0,6695 ± 0,0073	0,8478	0,3865	0,7462	0,9156
Random Forest	0,5897 ± 0,0230	0,8364	0,2879	0,7624	0,9234

6. Analisis kekeliruan

Analisis kekeliruan pada konfigurasi proposed memperlihatkan tingkat kesalahan pada kelas fakta jauh lebih tinggi, yaitu 82 dari 153 sampel fakta (0,536) salah diklasifikasikan sebagai

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

hoaks, berbanding 52 dari 693 sampel hoaks (0,075).

4.2 Pembahasan

Dari sisi mekanisme, strategi gabungan unggul karena memadukan dua komponen yang saling melengkapi, yaitu class weighting yang menaikkan kontribusi kelas minoritas secara global dan focal loss yang mengarahkan pembelajaran pada sampel sulit di dekat batas keputusan, sehingga recall fakta meningkat dengan penurunan recall hoaks yang kecil tanpa membuang informasi kelas mayoritas. Sebaliknya, random undersampling menaikkan recall fakta secara tajam karena dominasi mayoritas berkurang, namun pembuangan sampel menghilangkan variasi linguistik yang dibutuhkan untuk mengenali hoaks sehingga akurasi turun.

Pemeriksaan kualitatif terhadap sampel fakta yang salah klasifikasi menemukan tiga pola berulang, yaitu narasi fakta bergaya lugas tanpa penekanan emosional sehingga mirip hoaks pendek bergaya netral, sampel fakta sangat pendek yang miskin sinyal linguistik sehingga model condong pada prior mayoritas, serta sampel yang merujuk peristiwa kontroversial sehingga representasi kontekstualnya berdekatan dengan cluster hoaks pada ruang embedding. Pola tersebut menyarankan penambahan data fakta dari sumber yang lebih beragam atau augmentasi terarah pada kelas minoritas dengan kontrol kualitas ketat.

Augmentasi tidak konsisten membantu karena korpus relatif besar sehingga kapasitas IndoBERT-base sudah memadai tanpa sampel sintetik, metode penukaran dan penyisipan acak merusak struktur sintaktis teks pendek, dan augmentasi kontekstual menghasilkan parafrase yang terlalu mirip aslinya. Implikasinya, pada korpus ribuan sampel berketimpangan moderat, intervensi pada fungsi kerugian lebih berpengaruh dan stabil sehingga sebaiknya diprioritaskan.

Dibandingkan dengan penelitian terkait pada Subbab 2.1, temuan ini menunjukkan bahwa perbaikan kinerja tidak selalu menuntut modifikasi arsitektur seperti pendekatan hibrida; penyesuaian fungsi kerugian gabungan sudah cukup memperbaiki recall kelas minoritas, dan intervensi tingkat algoritma lebih menjanjikan daripada augmentasi data pada korpus ribuan sampel.

5. Kesimpulan

Penelitian ini menyajikan studi empiris pendekatan terpusat pada data untuk deteksi hoaks berbahasa Indonesia dan menghasilkan tiga simpulan. Pertama, IndoBERT-base mengungguli basis TF-IDF pada F1-macro dan MCC dengan margin signifikan pada sebagian besar pembanding. Kedua, strategi gabungan class weighting dan focal loss mencapai F1-

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

macro tertinggi yaitu 0,7240 dan signifikan terhadap basis TF-IDF serta augmentasi kontekstual pada taraf 0,05, tetapi tidak signifikan terhadap IndoBERT-base tanpa perlakuan berdasarkan uji McNemar dan paired bootstrap, sehingga kontribusinya terletak pada perbaikan recall kelas fakta dan MCC, bukan keunggulan menyeluruh. Ketiga, augmentasi teks tidak memberi kontribusi konsisten sehingga penelitian lanjutan sebaiknya memprioritaskan intervensi pada fungsi kerugian.

Beberapa keterbatasan dicatat. Korpus terbatas pada bahasa Indonesia berukuran ribuan sampel dengan rasio ketidakseimbangan sekitar 4,5 banding 1 dan hanya satu varian model, yaitu IndoBERT-base, sehingga generalisasi pada distribusi atau model lain perlu kehati-hatian. Arah lanjutan mencakup augmentasi kontekstual yang menjaga kualitas semantik pada teks pendek, integrasi sinyal eksternal seperti metadata sumber, serta evaluasi generalisasi lintas domain dan lintas waktu.

6. Daftar Pustaka

[1] M. F. Mridha, A. J. Keya, Md. A. Hamid, M. M. Monowar, and Md. S. Rahman, “A Comprehensive Review on Fake News Detection With Deep Learning,” *IEEE Access*, vol. 9, pp. 156151–156170, 2021, doi: 10.1109/ACCESS.2021.3129329.

[2] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020, pp. 757–770, doi: 10.18653/v1/2020.coling-main.66.

[3] B. Wilie et al., “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, Suzhou, China: Association for Computational Linguistics, 2020, pp. 843–857, doi: 10.18653/v1/2020.aacl-main.85.

[4] M. Y. Ridho, and E. Yulianti, “From Text to Truth: Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 10, no. 3, pp. 544–555, 2024.

[5] D. Y. Yefferson, V. Lawijaya, and A. S. Girsang, “Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 2, pp. 1913–1924, 2024, doi: 10.11591/ijai.v13.i2.pp1913-1924.

[6] S. Henning, W. Beluch, A. Fraser, and A. Friedrich, “A Survey of Methods for Addressing Class Imbalance in Deep-Learning Based Natural Language Processing,” 2023, pp. 523–540, doi: 10.18653/v1/2023.eacl-main.38.

[7] M. Bayer, M.-A. Kaufhold, and C. Reuter, “A Survey on Data Augmentation for Text Classification,” *ACM Computing Surveys*, vol. 55, no. 7, p. 146, 2022, doi: 10.1145/3544558.

[8] D. Zha et al., “Data-centric Artificial Intelligence: A Survey,” *ACM Computing Surveys*, vol. 57, no. 5, 2025, doi: 10.1145/3711118.

Jurnal Informatika dan Teknologi

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.786-797

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35214 Link : https://dx.doi.org/10.29408/jit.v9i2.35214

[9] M. I. K. Sinapoy, Y. Sibaroni, and S. S. 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, 2023.

[10] F. Muftie, and M. Haris, "IndoBERT Based Data Augmentation for Indonesian Text Classification," 2023, pp. 128–132, https://ieeexplore.ieee.org/document/10250061/.

[11] J. Tian et al., "Re-embedding Difficult Samples via Mutual Information Constrained Semantically Oversampling for Imbalanced Text Classification," 2021, doi: 10.18653/v1/2021.emnlp-main.252.

[12] K. R. M. Fernando, and C. P. Tsokos, "Dynamically Weighted Balanced Loss: Class Imbalanced Learning and Confidence Calibration of Deep Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 2940–2951, 2022, doi: 10.1109/TNNLS.2020.3047335.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017)*, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.324.

[14] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-Balanced Loss Based on Effective Number of Samples," 2019, pp. 9268–9277, doi: 10.1109/CVPR.2019.00949.

[15] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss," 2019, pp. 1567–1578, https://proceedings.neurips.cc/paper/2019/hash/621461af90cadfdaf0e8d4cc25129f91-Abstract.html.

[16] J. Chen, D. Tam, C. Raffel, M. Bansal, and D. Yang, "An Empirical Survey of Data Augmentation for Limited Data Learning in NLP," *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 191–211, 2023, doi: 10.1162/tacl_a_00542.

[17] T. G. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, 1998, doi: 10.1162/089976698300017197.

[18] D. Ulmer, C. Hardmeier, and J. Frellsen, "deep-significance: Easy and Meaningful Statistical Significance Testing in the Age of Neural Networks," 2022, doi: 10.48550/arXiv.2204.06815.

[19] J. Dodge, S. Gururangan, D. Card, R. Schwartz, and N. A. Smith, "Show Your Work: Improved Reporting of Experimental Results," 2019, pp. 2185–2194, doi: 10.18653/v1/D19-1224.

[20] D. Chicco, and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, p. 6, 2020, doi: 10.1186/s12864-019-6413-7.

.