

Klasifikasi Berita Hoaks Bencana Alam Menggunakan Representasi IndoBERT dan Algoritma XGBoost

¹Musthofa Ilmi, ²Ardi Sanjaya, ³Risky Aswi Ramadhani

^{1, 2, 3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ilmimusthofa8189@gmail.com, dersky@gmail.com, riskyaswiramadhani@gmail.com

Penulis Korespondens : Musthofa Ilmi

Abstrak— Penyebaran berita hoaks saat bencana alam dapat menimbulkan kepanikan dan mengganggu penanganan darurat. Penelitian ini mengembangkan sistem klasifikasi otomatis untuk mendeteksi berita hoaks berbahasa Indonesia bertema bencana alam. Dataset diperoleh melalui web scraping dari situs turnbackhoax.id untuk berita hoaks dan detik.com untuk berita valid, yang dibagi menjadi data latih, validasi, dan uji. Sistem menggunakan IndoBERT untuk ekstraksi fitur semantik dan XGBoost sebagai algoritma klasifikasi. Hasil pengujian evaluasi dari model menunjukkan akurasi sebesar 99,11%, f1-score sebesar 0,99 pada kedua kelas, serta recall 0,98 untuk hoaks dan 1,00 untuk valid. Sistem ini dikembangkan dalam antarmuka web berbasis React.js dan Flask, serta menunjukkan hasil klasifikasi yang konsisten dan cepat. Pendekatan ini terbukti efektif dalam mengenali pola linguistik berita hoaks, serta berpotensi diterapkan dalam platform verifikasi informasi pada situasi bencana di Indonesia.

Kata Kunci— Berita Palsu, Deteksi Hoaks, IndoBERT, Pemrosesan Bahasa Alami, XGBoost

Abstract— The spread of hoax news during natural disasters can cause panic and disrupt emergency response. This research develops an automatic classification system to detect hoax news in Indonesian with the theme of natural disasters. Datasets are obtained through web scraping from turnbackhoax.id for hoax news and detik.com for valid news, which are divided into training, validation, and test data. The system uses IndoBERT for semantic feature extraction and XGBoost as a classification algorithm. The evaluation test results of the model showed an accuracy of 99.11%, f1-score of 0.99 in both classes, and recall of 0.98 for hoax and 1.00 for valid. The system is developed in a React.js and Flask-based web interface, and showed consistent and fast classification results. This approach proved effective in recognising the linguistic patterns of hoax news, and has the potential to be applied in information verification platforms in disaster situations in Indonesia.

Keywords— Fake News, IndoBERT, Hoax Detection, Natural Language Processing, XGBoost

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Transformasi dalam bidang teknologi informasi dan komunikasi telah mengubah cara masyarakat memperoleh dan menyebarkan informasi, baik secara langsung maupun melalui media daring [4]. Di era digital saat ini, media sosial menjadi salah satu saluran utama yang memungkinkan informasi tersebar dengan sangat cepat dan luas jangkauannya [1]. Namun, kemudahan distribusi informasi ini juga membuka peluang bagi beredarnya konten yang menyesatkan, termasuk berita palsu atau hoaks. Dalam konteks bencana alam, penyebaran hoaks sangat berisiko karena dapat menciptakan kepanikan, mengacaukan koordinasi bantuan, dan menghambat proses evakuasi secara efektif [7]. Narasi yang digunakan dalam berita palsu semacam ini sering kali bersifat hiperbolik atau menakutkan, seperti menyebut "tsunami raksasa" atau "gempa dahsyat", yang bertujuan menarik perhatian publik namun justru memperburuk situasi [3].

Fenomena meningkatnya penyebaran hoaks, khususnya dalam konteks berita bencana alam, mendorong perlunya pengembangan sistem otomatis yang mampu mengidentifikasi informasi palsu secara efisien dan akurat. Deteksi berita hoaks berbasis teks telah menjadi fokus penelitian di bidang *Natural Language Processing (NLP)*, terutama karena kemampuan NLP dalam memahami struktur kalimat dan konteks semantik dari teks berbahasa alami [6]. Salah satu pendekatan mutakhir dalam NLP adalah penggunaan model berbasis transformer seperti BERT (*Bidirectional Encoder Representations from Transformers*), yang telah terbukti unggul dalam berbagai tugas klasifikasi teks.

Dalam konteks bahasa Indonesia, model BERT telah disesuaikan melalui pelatihan khusus menjadi IndoBERT, yang memanfaatkan korpus besar berbahasa Indonesia untuk menangkap nuansa semantik lokal secara lebih akurat [2]. IndoBERT menghasilkan representasi kontekstual dari teks dalam bentuk vektor, yang memungkinkan model memahami makna kalimat secara keseluruhan. Namun, untuk melakukan klasifikasi akhir terhadap teks tersebut, diperlukan algoritma klasifikasi yang dapat mengolah fitur kompleks dengan efisiensi tinggi.

Salah satu algoritma yang memenuhi kriteria tersebut adalah XGBoost (*Extreme Gradient Boosting*), yang dikenal sebagai algoritma ensemble tree boosting yang sangat efisien dalam hal waktu pelatihan dan akurasi prediksi [10]. Kombinasi antara representasi kontekstual dari IndoBERT dan kekuatan klasifikasi dari XGBoost telah menunjukkan hasil yang menjanjikan dalam berbagai tugas NLP, termasuk deteksi berita hoaks [8]. Meski demikian, penerapan kombinasi ini dalam konteks berita hoaks bertema bencana alam masih belum banyak dieksplorasi, sehingga membuka peluang kontribusi baru dalam bidang ini [9].

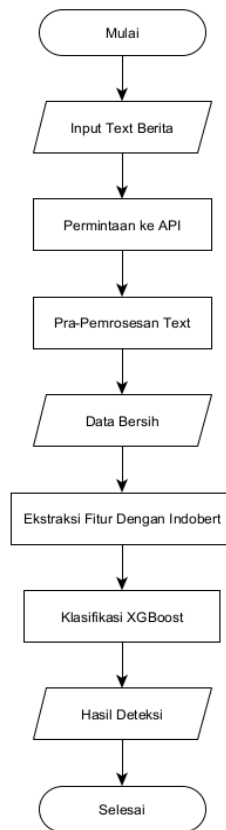
Penelitian ini bertujuan untuk mengembangkan sistem deteksi berita hoaks bertema bencana alam berbahasa Indonesia menggunakan pendekatan gabungan IndoBERT untuk ekstraksi fitur semantik dan XGBoost sebagai algoritma klasifikasi. Sistem ini dirancang dalam bentuk aplikasi berbasis web, dengan menggunakan *framework React.js* sebagai antarmuka pengguna dan *Flask* sebagai *backend API*, sehingga memungkinkan integrasi antara model deteksi dan interaksi langsung dengan pengguna secara real-time. Penelitian ini diharapkan dapat memberikan kontribusi dalam mendukung literasi digital dan membantu proses mitigasi bencana dengan mengurangi dampak negatif dari penyebaran informasi palsu.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk membangun sistem klasifikasi otomatis yang mampu membedakan antara berita hoaks dan berita valid dalam konteks bencana alam. Model yang dikembangkan menggabungkan representasi semantik berbasis IndoBERT dan algoritma klasifikasi XGBoost. Sistem juga diimplementasikan dan diuji melalui

antarmuka web berbasis *React.js* dan *Flask API* untuk menilai dalam skenario penggunaan nyata.

A. Alur Kerja Sistem



Gambar 1. *Flowchart Sistem*

Diagram alur di atas menggambarkan tahapan proses sistem deteksi berita hoaks berbasis web yang dibangun menggunakan kombinasi IndoBERT dan XGBoost. Proses diawali saat pengguna memasukkan teks berita melalui antarmuka aplikasi. Input teks ini kemudian dikirimkan sebagai permintaan (*request*) ke API berbasis Flask, yang mengelola seluruh alur pemrosesan data. Setelah menerima data, sistem menjalankan tahap pra-pemrosesan teks yang mencakup pembersihan karakter khusus, konversi huruf ke format standar (*lowercase*), *tokenisasi*, penghapusan *stopword*, serta proses *stemming*. Tahapan ini menghasilkan data teks yang bersih, siap untuk diproses lebih lanjut. Selanjutnya, data bersih diekstraksi menggunakan model IndoBERT untuk memperoleh representasi vektor dari teks menggunakan output token [CLS]. Vektor hasil ekstraksi ini menjadi input bagi algoritma XGBoost, yang akan mengklasifikasikan teks ke dalam salah satu dari dua kelas: hoaks atau valid. Hasil klasifikasi dikembalikan ke sisi antarmuka sebagai output berupa label deteksi serta persentase probabilitas. Sistem secara otomatis menyajikan hasil kepada pengguna dalam bentuk visualisasi dan keterangan prediksi, menandai berakhirnya proses klasifikasi.

B. Proses Pengumpulan Data

Data yang digunakan dalam penelitian ini dikumpulkan melalui proses *web scraping* dari dua sumber utama, yaitu situs *turnbackhoax.id* sebagai sumber berita hoaks dan *detik.com* sebagai sumber berita valid. Seluruh berita difokuskan pada topik bencana alam seperti gempa bumi, banjir, dan letusan gunung berapi. Untuk menjaga relevansi dan kesesuaian konteks, pengumpulan data dibatasi pada berita yang diterbitkan dalam periode Mei 2020 hingga Mei

2024. Dari proses ini diperoleh total 3004 data, yang kemudian dibagi menjadi tiga kelompok, yaitu 2102 data pelatihan, 451 data validasi, dan 451 data pengujian. Setiap entri dalam dataset diberi label biner: 0 untuk berita hoaks dan 1 untuk berita valid. Sebagai ilustrasi, contoh data dari masing-masing kelas disajikan pada Tabel berikut, yang mencakup informasi berupa label, judul berita (*headline*), serta isi utama berita (*body*) untuk memberikan gambaran mengenai perbedaan karakteristik antara berita hoaks dan non-hoaks.

Tabel 1. Contoh Data Hoaks Dan Valid

Label	Headline	Body
0	“Gempa Dahsyat Akan Terjadi Lagi di Jawa Timur!”	Sebuah pesan berantai menyebutkan akan terjadi gempa besar susulan di wilayah Malang...
1	“BMKG: Tidak Ada Ancaman Tsunami Setelah Gempa”	BMKG menyatakan bahwa gempa bumi yang terjadi di selatan Jawa tidak berpotensi tsunami...

C. Pra-Pemrosesan Data

Sebelum digunakan dalam pemodelan, data terlebih dahulu diproses melalui beberapa tahapan pra-pemrosesan untuk memastikan kualitas dan konsistensinya. Proses ini dimulai dengan pembersihan teks, yaitu menghapus karakter khusus dan mengubah semua huruf menjadi huruf kecil guna menghindari duplikasi akibat perbedaan penulisan. Selanjutnya, dilakukan tokenisasi untuk memecah teks menjadi unit-unit kata yang lebih kecil, serta penghapusan *stopword* agar kata-kata yang tidak memiliki makna signifikan dapat dieliminasi. Setelah itu, kata-kata yang telah ditokenisasi dinormalisasi melalui proses *stemming*, sehingga kata-kata dengan akar yang sama dikembalikan ke bentuk dasarnya. Terakhir, teks yang telah dibersihkan dan dinormalisasi diubah menjadi representasi *vektor* numerik menggunakan *tokenizer* dari model IndoBERT, yang memungkinkan teks dimengerti oleh model pembelajaran mesin.

D. Ekstraksi Fitur dengan IndoBERT

Ekstraksi fitur dilakukan dengan memanfaatkan *IndoBERT Base*, yaitu model *pre-trained* berbasis arsitektur BERT yang dirancang khusus untuk bahasa Indonesia. Model ini diunduh melalui pustaka *HuggingFace Transformers*, yang menyediakan berbagai model *state-of-the-art* untuk pemrosesan bahasa alami. IndoBERT dilatih pada korpus besar berbahasa Indonesia, termasuk berita, artikel ilmiah, dan percakapan informal, sehingga mampu memahami konteks linguistik lokal dengan lebih baik daripada model generik [2]. Dalam penelitian ini, teks berita yang telah melalui tahap pra-pemrosesan diubah menjadi representasi numerik berupa vektor embedding. Representasi ini diperoleh dari *pooled output token* khusus [CLS] pada lapisan akhir model, yang dirancang untuk menangkap konteks semantik keseluruhan dokumen [5]. Vektor [CLS] ini kemudian digunakan sebagai fitur masukan untuk tahap klasifikasi, karena secara empiris terbukti efektif untuk berbagai tugas klasifikasi teks.

E. Klasifikasi dengan XGBoost

Vektor embedding hasil ekstraksi dari IndoBERT kemudian digunakan sebagai input untuk algoritma XGBoost (*Extreme Gradient Boosting*). XGBoost adalah algoritma klasifikasi berbasis ensemble tree yang terkenal karena kecepatan dan akurasi yang tinggi dalam berbagai kompetisi data sains [10]. Pada penelitian ini, XGBoost digunakan untuk melakukan klasifikasi biner terhadap teks berita, dengan dua kelas utama: hoaks (0) dan valid (1). Model XGBoost dilatih menggunakan parameter yang telah disesuaikan untuk performa optimal, yaitu *max_depth* = 6 untuk mengatur kedalaman pohon keputusan, *learning_rate* = 0.1 untuk mengontrol

kontribusi masing-masing pohon terhadap model akhir, dan $n_estimators = 100$ yang menentukan jumlah pohon yang dibangun selama pelatihan. XGBoost dipilih karena kemampuannya mengatasi *overfitting*, efisiensi dalam pemrosesan data skala besar, serta fleksibilitas dalam pengaturan *hyperparameter*. Kombinasi IndoBERT dan XGBoost ini terbukti efektif dalam penelitian serupa untuk tugas-tugas klasifikasi teks, termasuk deteksi berita palsu dan spam [8].

F. Pengembangan Sistem melalui Antarmuka Web

Untuk mengevaluasi performa model dalam konteks penggunaan nyata, sistem klasifikasi yang telah dibangun diimplementasikan dalam bentuk antarmuka web. Antarmuka pengguna (*frontend*) dikembangkan menggunakan *React.js*, sedangkan *backend* dikembangkan menggunakan Flask yang berfungsi sebagai API untuk memproses input teks dan memberikan hasil prediksi. Pada tahap pengujian ini, pengguna dapat memasukkan judul dan isi berita melalui form pada antarmuka web. Data yang dimasukkan kemudian dikirim ke backend Flask yang akan memproses input menggunakan model IndoBERT + XGBoost. Hasil klasifikasi ditampilkan kembali ke pengguna dalam bentuk label (hoaks atau valid) beserta probabilitasnya dalam format grafik pie.

G. Evaluasi Model

Evaluasi performa model dilakukan dengan menggunakan confusion matrix serta empat metrik utama, yaitu akurasi, presisi, *recall*, dan *f1-score*. Akurasi mengukur jumlah prediksi yang benar dibandingkan dengan total data uji, sementara presisi menunjukkan proporsi prediksi berita hoaks yang benar dari seluruh prediksi hoaks yang dibuat oleh model. *Recall* mengukur sejauh mana model mampu menangkap semua berita hoaks yang sebenarnya ada dalam data, dan *f1-score* digunakan sebagai ukuran harmonisasi antara presisi dan recall, terutama ketika distribusi kelas tidak seimbang.

III. HASIL DAN PEMBAHASAN

A. Hasil Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas dan konsistensi data teks sebelum dimasukkan ke model. Teks berita awal yang diperoleh dari scraping masih mengandung karakter khusus, kapitalisasi tidak seragam, serta kata-kata umum (*stopwords*) yang tidak relevan. Setelah pembersihan, tokenisasi, penghapusan *stopwords*, dan *stemming*, data menjadi lebih terstruktur dan siap diproses lebih lanjut.

Tabel 2. Hasil Pra-Pemrosesan Data

Label	Headline	Body
0	“Gempa Dahsyat Akan Terjadi Lagi di Jawa Timur!”	gempa dahsyat akan terjadi jawa timur
1	“BMKG: Tidak Ada Ancaman Tsunami Setelah Gempa”	bmkg tidak ada ancaman tsunami setelah gempa selatan jawa

B. Ekstraksi Fitur dengan IndoBERT

Setelah tahap pra-pemrosesan, teks diubah menjadi representasi numerik menggunakan IndoBERT. Setiap dokumen berita dikonversi menjadi *vektor embedding* yang diperoleh dari *pooled output token* [CLS]. Representasi ini menangkap makna kontekstual dari keseluruhan teks, bukan hanya makna kata per kata.

Tabel 3. Hasil Ekstraksi Fitur dengan IndoBERT

Label	Headline	Body
0	gempa dahsyat akan terjadi jawa timur	[0.234, -0.152, 0.556, 0.091, -0.327, ...]
1	bmkg tidak ada ancaman tsunami setelah gempa jawa	[0.187, -0.211, 0.643, 0.076, -0.289, ...]

C. Klasifikasi dengan XGBoost

Vektor embedding dari IndoBERT kemudian digunakan sebagai input untuk model XGBoost. Setelah proses pelatihan, model mampu memetakan vektor input ke dalam dua kelas: hoaks (0) atau valid (1). Contoh hasil klasifikasi dapat dilihat pada Tabel 4.

Tabel 4. Hasil Klasifikasi dengan XGBoost

Label	Headline	Hasil Prediksi
0	gempa dahsyat akan terjadi jawa timur	Hoaks (0)
1	bmkg tidak ada ancaman tsunami setelah gempa jawa”	Valid (1)

D. Hasil Evaluasi Model

Setelah model dinyatakan memiliki performa tinggi secara evaluasi numerik, sistem diuji dalam bentuk implementasi nyata menggunakan antarmuka web. Antarmuka dikembangkan menggunakan *React.js* sebagai *frontend* dan *Flask* sebagai backend API yang menghubungkan pengguna dengan model klasifikasi. Dalam pengujian ini, pengguna memasukkan judul dan isi berita secara langsung ke dalam form yang tersedia, dan hasil klasifikasi ditampilkan secara real-time dalam bentuk label (hoaks atau valid) beserta persentase probabilitasnya.



Gambar 2. Hasil Klasifikasi Sistem Untuk Berita Valid



Gambar 3. Hasil Klasifikasi Sistem Untuk Berita Hoaks

Pada gambar tersebut, salah satu pengujian dilakukan pada berita yang berasal dari BMKG, yang berisi penjelasan teknis mengenai gempa bumi tanpa narasi berlebihan. Sistem mengklasifikasikan berita ini sebagai valid dengan probabilitas 94%. Sebaliknya, ketika diuji dengan teks yang bersifat emosional dan tidak bersumber resmi, sistem mengklasifikasikannya sebagai hoaks dengan probabilitas 98%. Hasil ini memperlihatkan bahwa model tidak hanya unggul dalam evaluasi data statis, tetapi juga mampu mempertahankan konsistensi klasifikasi dalam skenario penggunaan nyata.

E. Hasil Evaluasi Model

Setelah dilakukan proses pelatihan model menggunakan gabungan IndoBERT sebagai ekstraktor fitur dan XGBoost sebagai algoritma klasifikasi, diperoleh hasil evaluasi yang sangat tinggi. Model diuji dengan data uji sebanyak 451 data. Evaluasi dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan f1-score. Berikut adalah hasil evaluasi model berdasarkan data uji.

Tabel 5. Hasil Evaluasi Model IndoBERT dan XGBoost

Kelas	Precision	Recall	F1-Score	Support
Hoaks (0)	1,00	0,98	0,99	226
Valid (1)	0,98	1,00	0,99	225
Rata-Rata	0,99	0,99	0,99	451

Berdasarkan hasil evaluasi, model menunjukkan performa yang sangat tinggi dengan akurasi keseluruhan sebesar 99,11%. Nilai *f1-score* sebesar 0,99 pada kedua kelas mengindikasikan bahwa model mampu menjaga keseimbangan yang sangat baik antara presisi dan *recall*. Presisi sempurna (1.00) pada kelas hoaks menunjukkan bahwa seluruh berita yang diklasifikasikan sebagai hoaks oleh sistem memang benar-benar hoaks, sementara *recall* sebesar 0,98 menunjukkan bahwa sebagian besar berita hoaks berhasil dikenali dengan sangat tepat oleh model.

F. Confusion Matrix

Untuk memahami detail kesalahan klasifikasi, digunakan *confusion matrix* yang ditunjukkan pada tabel berikut.

Tabel 6. *Confusion Matrix Model IndoBERT + XGBoost*

	Hoaks	Valid
Hoaks	222	4
Valid	0	225

Hasil di atas menunjukkan bahwa dari 226 berita hoaks, hanya 4 yang salah diklasifikasikan sebagai valid. Sementara semua berita valid berhasil diklasifikasikan dengan benar. Nilai *precision* dan *recall* yang hampir sempurna pada kedua kelas menunjukkan bahwa model tidak hanya sensitif terhadap berita hoaks, tetapi juga sangat akurat dalam mengidentifikasi berita yang benar.

IV. KESIMPULAN

Penelitian ini berhasil membuktikan bahwa kombinasi IndoBERT dan XGBoost merupakan pendekatan yang efektif untuk tugas klasifikasi berita hoaks berbahasa Indonesia, khususnya dalam domain berita bencana alam. Dengan memanfaatkan kemampuan IndoBERT dalam menangkap makna kontekstual dari teks dan keunggulan XGBoost dalam melakukan klasifikasi secara efisien, sistem yang dikembangkan mampu mencapai akurasi sebesar 99,11%, dengan *f1-score* sebesar 0,99 untuk kedua kelas, yaitu hoaks dan valid. Selain itu, model menunjukkan *recall* sebesar 0,98 pada kelas hoaks dan 1,00 pada kelas valid, yang menandakan bahwa model mampu mendeteksi hampir seluruh berita hoaks serta mengenali seluruh berita valid tanpa false positive. Hasil ini menunjukkan bahwa sistem tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga seimbang dalam hal sensitivitas dan presisi terhadap kedua jenis berita.

Kontribusi utama dari penelitian ini terhadap bidang ilmu komputer dan teknologi informasi terletak pada pengembangan pendekatan terpadu yang efisien dan akurat untuk klasifikasi berita hoaks berbahasa Indonesia dalam konteks bencana alam, yang hingga kini masih relatif jarang menjadi fokus penelitian. Sistem telah diimplementasikan dan diuji melalui antarmuka web berbasis React.js dan Flask, memungkinkan pengguna untuk secara langsung menguji berita dengan hasil yang cepat, konsisten, dan akurat sesuai dengan hasil evaluasi model.

Ke depan, sistem ini dapat dikembangkan lebih lanjut dengan memperluas cakupan data dari berbagai sumber media daring, menambahkan fitur klasifikasi multi-label, serta menyertakan analisis tingkat kepercayaan (*confidence score*) terhadap prediksi model. Dengan demikian, hasil penelitian ini tidak hanya memberikan kontribusi teoritis dalam bidang pemrosesan bahasa alami dan pembelajaran mesin, tetapi juga membuka peluang implementasi praktis dalam mendukung literasi digital dan meningkatkan ketahanan informasi masyarakat Indonesia terhadap penyebaran hoaks, khususnya pada situasi darurat seperti bencana alam.

DAFTAR PUSTAKA

- [1] A. R. I. Fauzy and E. B. Setiawan, "Detecting Fake News on Social Media Combined with the CNN Methods," *J. RESTI*, vol. 7, no. 2, pp. 271–277, 2023.
- [2] B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," *arXiv preprint arXiv:2009.05387*, 2020.
- [3] I. Ginting, "Pentingnya Daya Kritis Masyarakat Tangkal HOAX," *babelprov.go.id*, 2024.
- [4] J. Amalia et al., "Model Klasifikasi Berita Palsu Menggunakan Bidirectional LSTM dan Word2Vec Sebagai Vektorisasi," *J. Tek. Inform. dan Sist. Inform.*, vol. 9, no. 4, 2022.
- [5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [6] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," *ACM SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, 2017.
- [7] Kementerian Komunikasi dan Informatika Republik Indonesia, "Kominfo Tangani Ribuan Isu Hoaks," *kominfo.go.id*, 2024. [Online]. Available: <https://www.kominfo.go.id>.
- [8] R. Juarto and I. Yulianto, "Deteksi Berita Hoaks Bahasa Indonesia Menggunakan IndoBERT," *J. Ilmu Komput.*, vol. 12, no. 1, pp. 45–54, 2023.
- [9] R. S. H. Pratama, D. I. Ardiyanto, and A. Ayu, "Klasifikasi Hoax Vs Non-Hoax pada Berita Bencana Alam Berbahasa Indonesia Menggunakan Word Embedding," *J. RESTI*, vol. 6, no. 6, pp. 1084–1091, 2022.
- [10] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.