

IMPLEMENTASI ALGORITMA INDOBERT DAN XGBOOST UNTUK DETEKSI BERITA PALSU

SKRIPSI

Diajukan Guna Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer (S. Kom.)
Pada Program Studi Teknik Informatika



Oleh :

Musthofa Ilmi
NPM : 20113020172

**FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS NUSANTARA PGRI KEDIRI
2025**

Skripsi oleh:

Musthofa Ilmi
NPM : 20113020172

Judul :

**IMPLEMENTASI ALGORITMA INDOBERT DAN XGBOOST UNTUK
DETEKSI BERITA PALSU**

Telah Disetujui Untuk Diajukan Kepada Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri

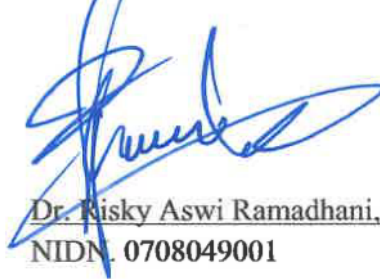
Tanggal : 15 Juli 2025

Pembimbing I



Ardi Sanjaya, M.Kom.
NIDN. 0706118101

Pembimbing II



Dr. Risky Aswi Ramadhani, M. Kom.
NIDN. 0708049001

Skripsi oleh:

Musthofa Ilmi
NPM : 20113020172

Judul :

**IMPLEMENTASI ALGORITMA INDOBERT DAN XGBOOST UNTUK
DETEKSI BERITA PALSU**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer

Universitas Nusantara PGRI Kediri

Pada tanggal : 15 Juli 2025

Dan Dinyatakan telah Memenuhi Syarat

Panitia Penguji :

1. Ketua : Ardi Sanjaya, M.Kom
2. Penguji I : Siti Rochana, M.Pd
3. Penguji II : Dr. Risky Aswi Ramadhani, M. Kom



Mengetahui,
Dekan FTIK

Dr. Sulistiono, M.Si.
NIDN. 0007076801

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Musthofa Ilmi
Jenis Kelamin : Laki - Laki
Tempat/Tgl Lahir : Nganjuk, 24 Mei 2002
NPM : 2113020172
Fakultas/Prodi : Teknik dan Ilmu Komputer/ Teknik Informatika

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak dapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 18 Juli 2025
Yang Menyatakan



Musthofa Ilmi
NPM : 2113020172

HALAMAN PERSEMBAHAN

Penulisan skripsi ini dengan tulus saya dedikasikan kepada:

1. Kedua orang tua saya, yang dengan penuh kesabaran senantiasa mendoakan, memberikan dukungan terbaik, serta menjadi sumber motivasi tak henti-hentinya dalam perjalanan menyelesaikan skripsi ini.
2. Adik-adik saya, yang selalu memberikan semangat dan dukungan moral, sehingga saya dapat menyelesaikan proses ini dengan baik.
3. Seluruh dosen Universitas Nusantara PGRI Kediri, yang telah memberikan bimbingan, ilmu, dan pelajaran berharga, baik dalam ranah akademik maupun kehidupan sehari-hari.
4. Teman-teman seperjuangan di kampus, yang menjadi tempat berbagi suka dan duka, serta saling menyemangati selama menjalani masa perkuliahan hingga tahap akhir ini.
5. Almamater tercinta, Universitas Nusantara PGRI Kediri, yang telah menjadi tempat saya tumbuh, belajar, dan mengembangkan diri selama masa perkuliahan.
6. Semua pihak yang tidak dapat saya sebutkan satu per satu, yang telah memberikan kontribusi dalam berbagai bentuk dukungan selama proses penyusunan skripsi ini.

Semoga dedikasi ini dapat menjadi penghormatan atas semua dukungan, doa, dan kebersamaan yang telah diberikan.

HALAMAN MOTTO

"A lie can travel halfway around the world while the truth is still putting on its shoes." — **Mark Twain**

"Media literacy is not just an option, it is a necessity in the digital age." — **Renee Hobbs**

"In a time of deceit, telling the truth is a revolutionary act."
— *George Orwell*

RINGKASAN

Musthofa Ilmi Implementasi Algoritma Indobert dan XGBoost untuk Deteksi Berita Palsu

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

Penelitian ini dilatarbelakangi oleh meningkatnya penyebaran berita hoaks di Indonesia, khususnya pada saat terjadi bencana alam seperti gempa bumi, banjir, dan letusan gunung berapi. Berita palsu yang tersebar di media sosial sering kali mengandung narasi menyesatkan dan tidak berbasis sumber resmi, sehingga dapat menimbulkan kepanikan dan mengganggu proses penanganan darurat. Penyaringan informasi secara manual dianggap tidak lagi memadai karena volume dan kecepatan distribusi berita digital yang sangat tinggi. Oleh karena itu, dibutuhkan sistem otomatis yang mampu mendeteksi berita hoaks secara cepat dan akurat.

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi otomatis untuk mendeteksi berita hoaks bertema bencana alam berbahasa Indonesia menggunakan kombinasi model IndoBERT dan algoritma XGBoost. IndoBERT digunakan untuk mengekstraksi representasi semantik dari teks berita, sedangkan XGBoost berperan sebagai algoritma klasifikasi yang mengelompokkan berita ke dalam kategori hoaks atau valid. Sistem ini dikembangkan menggunakan pendekatan *waterfall* dan diimplementasikan sebagai aplikasi web berbasis *React.js* pada sisi antarmuka dan *Flask* pada sisi *backend*. *Dataset* diperoleh melalui *web scraping* dari *turnbackhoax.id* (berita hoaks) dan *detik.com* (berita valid) dengan rentang data dari Mei 2020 hingga Mei 2024. Hasil evaluasi menunjukkan akurasi sebesar 99,11%, dengan *f1-score* sebesar 0,99 pada kedua kelas. Pengujian sistem melalui metode *Blackbox Testing* juga menunjukkan bahwa seluruh fungsi berjalan dengan baik sesuai dengan yang dirancang.

PRAKATA

Puji Syukur dipanjatkan kepada Tuhan Yang Maha Esa, karena atas ridha dan karunianya peneliti dapat menyelesaikan penyusunan laporan penelitian ini. Penulisan ini juga tak lepas dari dukungan pihak yang selalu membantu dalam penulisan penelitian ini. Oleh karenanya peneliti ingin mengucapkan terima kasih kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri.
3. Risa Helilintar, M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri.
4. Ardi Sanjaya, M.Kom. dan Dr. Risky Aswi Ramadhani, M. Kom. selaku Dosen Pembimbing Skripsi yang telah dan mengarahkan kami selama mengerjakan skripsi.
5. Kedua Orang Tua saya dan Keluarga atas doa dan dukungannya.
6. Ucapan terimakasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan penulisan penelitian ini.

Disadari penelitian ini masih banyak kekurangan, maka diharapkan kritik dan saran dari berbagai pihak sangat diharapkan. Semoga penelitian ini bermanfaat bagi semua pihak.

Kediri, 13 Januari 2025



Musthofa Ilmi

NPM. 2113020172

DAFTAR ISI

SAMPUL.....	i
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN.....	v
HALAMAN MOTTO	vi
RINGKASAN	vii
PRAKATA.....	viii
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xi
BAB I PENDAHULUAN	1
A. Latar Belakang.....	1
B. Identifikasi Masalah	2
C. Rumusan Masalah.....	3
D. Batasan Masalah	3
E. Tujuan Penelitian	4
F. Manfaat Penelitian	4
BAB II LANDASAN TEORI	5
A. Teori dan Penelitian Terdahulu	5
1. Landasan Teori	5
2. Kajian Pustaka.....	13
B. Kerangka Berpikir	16
BAB III METODE PENELITIAN	17
A. Desain Penelitian	17
1. Jenis Penelitian	17
2. Variabel Penelitian.....	17
3. Metode Pengumpulan Data	18
B. Instrumen Penelitian	19
1. Perangkat Keras (Hardware).....	19

2.	Perangkat Lunak (Software).....	19
3.	Dataset	21
4.	Analisis Hasil	22
C.	Jadwal Penelitian	24
D.	Objek Penelitian/ Subjek Penelitian.....	24
1.	Analisis Kebutuhan Sistem	24
2.	Objek penelitian.....	25
3.	Subjek penelitian.....	26
E.	Prosedur Penelitian	26
1.	Analisis persyaratan	27
2.	Desain Sistem.....	27
3.	Implementasi.....	28
4.	Pengujian	28
5.	Evaluasi Model.....	29
F.	Teknik analisis data.....	29
1.	Desain Sistem.....	29
2.	Simulasi Penyelesaian Masalah.....	36
BAB IV HASIL DAN PEMBAHASAN.....		44
A.	Hasil Penelitian.....	44
1.	Implementasi Desain Sistem.....	44
2.	Pengujian Fungsional.....	47
2.	Pengujian Non-Fungsional	50
B.	Pembahasan Desain Pengembangan.....	51
1.	Umpan Balik Pengguna terhadap Sistem.....	52
2.	Keunggulan dan kekurangan Sistem yang Dikembangkan	55
3.	Analisis Hasil Pengujian.....	56
BAB V PENUTUP		53
A.	Kesimpulan.....	53
B.	Saran	54
DAFTAR PUSTAKA		55
LAMPIRAN.....		58

DAFTAR TABEL

Tabel 2. 1 Confussion Metrix.....	11
Tabel 3. 1Jadwal Penelitian.....	24
Tabel 3. 2 Atribut Class NewDataset	32
Tabel 3. 3 Method Class NewsDataset	33
Tabel 3. 4 Contoh Dataset.....	37
Tabel 3. 5 Tabel Evaluasi.....	42
Tabel 4. 1 Tabel PraPemrosesan	44
Tabel 4. 2 Hasil Vektor [CLS]	45
Tabel 4. 3 Klasifikasi XGBoost	45
Tabel 4. 4 Pengujian Fungsional (Black Box)	48
Tabel 4. 5 Pengujian Fungsional (White Box).....	49
Tabel 4. 6 Hasil Evaluasi Model	50
Tabel 4. 7 Hasil Confussion Metrix	51
Tabel 4.8 Rencana Pertanyaan Umpan Balik.....	52
Tabel 4. 9 Hasil Rekap Responden	53
Tabel 4. 10 Rata rata penilaian tiap pertanyaan	54

DAFTAR GAMBAR

Gambar 2. 1 Pre-Training dan Fine-Tuning BRT (Devlin dkk., 20).....	9
Gambar 2. 2 Bagan Kerangka Berpikir.....	16
Gambar 3. 1 Diagram Waterfall.....	27
Gambar 3. 2 DFD Level 0.....	29
Gambar 3. 3 DFD Level 1.....	30
Gambar 3. 4 Flowchart Sistem.....	31
Gambar 3. 5 Class Diagram	32
Gambar 3. 6 Desain Wireframe	34
Gambar 3. 7Pohon Keputusan.....	40
Gambar 4. 1 Hasil deteksi sistem untuk berita hoaks	46
Gambar 4. 2 Hasil Deteksi Sistem Untuk Berita Vlaind.....	46

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi informasi telah mengubah secara signifikan cara masyarakat mengakses dan menyebarkan informasi. Di Indonesia, tingginya jumlah pengguna internet yang mencapai lebih dari 220 juta orang pada tahun 2024 (APJII, 2024) menunjukkan bagaimana internet menjadi saluran utama distribusi berita. Namun, pesatnya arus informasi ini juga mendorong maraknya penyebaran informasi yang tidak benar atau berita palsu (hoaks), terutama melalui media sosial. Berita hoaks tidak hanya menyesatkan, tetapi juga dapat menimbulkan keresahan, terutama ketika terjadi pada situasi kritis seperti bencana alam (Biro Humas Kominfo, 2024).

Penyebaran berita palsu terkait bencana alam seperti gempa bumi, banjir, dan letusan gunung api berpotensi menciptakan kepanikan masal dan menghambat penanganan darurat. Informasi palsu semacam ini sering kali dikemas dalam bentuk judul atau isi yang sensasional dan menakutkan misalnya, “tsunami raksasa segera melanda” atau “gempa dahsyat akan terjadi” yang tidak memiliki dasar fakta yang valid. Narasi hiperbolik seperti ini justru memperburuk situasi karena membuat masyarakat mudah terprovokasi dan sulit membedakan informasi *kredibel* (Ginting, 2024). Survei oleh Mastel (2019) mengungkapkan bahwa lebih dari 92% masyarakat mengidentifikasi media sosial sebagai sumber utama hoaks, dan 34% di antaranya menerima hoaks setiap hari. Hal ini menunjukkan bahwa penyaringan informasi secara manual tidak lagi memadai, dan dibutuhkan pendekatan otomatis berbasis teknologi.

Salah satu pendekatan yang banyak digunakan adalah pemrosesan bahasa alami (*Natural Language Processing/NLP*). Model bahasa seperti IndoBERT yang dikembangkan berdasarkan arsitektur BERT (Devlin, 2018), telah menunjukkan performa tinggi dalam menangkap konteks semantik dari teks berita berbahasa Indonesia (Juarto & Yulianto, 2023). Namun, model ini

masih memerlukan algoritma klasifikasi tambahan yang efisien untuk mengolah *vektor* representasi teks yang dihasilkannya agar dapat membedakan antara berita hoaks dan berita valid secara akurat.

Extreme Gradient Boosting (XGBoost) merupakan salah satu algoritma klasifikasi yang terbukti efektif dalam menangani data berukuran besar dan kompleks. Penelitian oleh Ramraj dkk. (2016) menunjukkan bahwa XGBoost dapat mencapai akurasi tinggi pada berbagai skenario klasifikasi, bahkan pada data tidak seimbang. Kombinasi representasi kontekstual dari IndoBERT dan kekuatan klasifikasi dari XGBoost menjadi landasan untuk membangun sistem deteksi berita hoaks yang andal.

Berdasarkan permasalahan dan peluang tersebut, penelitian ini bertujuan untuk mengembangkan sistem deteksi berita hoaks bertema bencana alam menggunakan kombinasi IndoBERT dan XGBoost. Sistem ini diharapkan dapat membantu meminimalkan penyebaran berita palsu dengan meningkatkan akurasi klasifikasi secara otomatis, serta menjadi solusi efektif dalam mendukung literasi digital dan pengambilan keputusan yang cepat dalam situasi krisis.

B. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, berikut merupakan masalah yang telah diidentifikasi:

1. Kurangnya pemahaman masyarakat dalam memverifikasi informasi yang diterima.
2. Kurangnya sistem yang mengoptimalkan kombinasi algoritma IndoBERT dan XGBoost dalam mendeteksi berita hoaks secara spesifik tentang bencana alam.

C. Rumusan Masalah

Berdasarkan masalah yang telah di identifikasikan, rumusan masalah pada penelitian ini adalah :

1. Bagaimana merancang sistem yang efektif untuk membantu masyarakat membedakan informasi palsu dari informasi yang valid?
2. Bagaimana memanfaatkan kombinasi algoritma berbasis NLP seperti IndoBERT dan algoritma klasifikasi seperti XGBoost secara optimal dalam mendeteksi berita palsu berbahasa Indonesia bertema bencana alam berbahasa Indonesia?

D. Batasan Masalah

Untuk memfokuskan permasalahan yang dibahas, penelitian ini menetapkan beberapa batasan masalah, yaitu:

1. Data yang digunakan dalam penelitian ini berupa teks berita berbahasa Indonesia bertema bencana alam, yang diperoleh melalui proses *web scraping* dari situs *turnbackhoax.id* sebagai sumber berita hoaks dan *detik.com* sebagai sumber berita valid, dalam rentang waktu Mei 2020 hingga Mei 2024.
2. Penelitian ini hanya memfokuskan pada dua kelas utama, yaitu berita hoaks (label 0) dan berita valid (label 1), tanpa mempertimbangkan kategori berita lain seperti opini, *satire*, atau *clickbait*.
3. Metode yang digunakan dalam penelitian ini terbatas pada kombinasi IndoBERT sebagai model ekstraksi fitur berbasis *transformer* dan XGBoost sebagai algoritma klasifikasi, tanpa membandingkan algoritma lain secara *implementatif*.
4. Evaluasi kinerja model dilakukan menggunakan metrik kuantitatif yaitu akurasi, presisi, *recall*, dan *F1-score*.
5. Penelitian ini tidak menggunakan data kontekstual atau *historis* lain, seperti pola penyebaran berita, sumber penulis, atau *metadata*, dalam menentukan keaslian sebuah berita.

6. Sistem yang dikembangkan dalam penelitian ini memiliki keterbatasan karena hanya mengandalkan analisis teks isi berita, tanpa mempertimbangkan elemen non-teks seperti gambar dan video.

E. Tujuan Penelitian

Berdasarkan rumusan masalah yang dijabarkan sebelumnya, tujuan penelitian ini adalah :

1. Mengembangkan sistem deteksi berita palsu yang efektif dan dapat membantu masyarakat dalam membedakan informasi yang tidak benar.
2. Implementasi kombinasi algoritma berbasis NLP (*IndoBERT*) dan algoritma *klasifikasi* (*XGBoost*) untuk mendeteksi berita palsu bencana alam secara otomatis.

F. Manfaat Penelitian

Berdasarkan tujuan yang telah dijelaskan, penelitian ini diharapkan:

1. Memberikan solusi dalam mendeteksi berita palsu secara efektif, Serta memberikan kontribusi positif dalam mengurangi penyebaran informasi yang kurang benar di masyarakat.
2. Memberikan referensi yang bermanfaat bagi peneliti dan praktisi dalam mengembangkan sistem deteksi berita palsu bencana alam berbahasa Indonesia.

DAFTAR PUSTAKA

- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>.
- Alpaydin, E. (2020). *Introduction to Machine Learning* (4th ed.). MIT Press
- Banks, A., & Porcello, E. (2017). *Learning React: Functional web development with React and Redux*. O'Reilly Media.
- Biro Humas Kementerian Kominfo. (2024). Siaran Pers No. 02/HM/KOMINFO/01/2024 tentang hingga akhir tahun 2023, Kominfo tangani 12.547 isu hoaks. *Kementerian Komunikasi dan Informatika*. <https://komdigi.kominfo.go.id/>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*. <https://arxiv.org/abs/1810.04805>
- Fawaid, J., Awalina, A., Krisnabayu, R. Y., & Yudistira, N. (2021). Indonesia's fake news detection using transformer network. In *Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology* (pp. 247–251).
- Ginting, I. (2024). Pentingnya daya kritis masyarakat tangkal hoaks. Babelprov.go.id. https://babelprov.go.id/Artikel_detil/Pentingnya-Daya-Kritis-Masyarakat-Tangkal-Hoax

- Grinberg, M. (2018). *Flask Web Development: Developing web applications with Python*. O'Reilly Media.
- Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan metode klasifikasi Support Vector Machine dan Naïve Bayes untuk analisis sentimen pada ulasan tekstual di Google Play Store. *ILKOM: Jurnal Ilmiah*, 12(2), 154–161.
- Juarto, A. R., & Yulianto, F. (2023). Deteksi berita hoaks menggunakan IndoBERT dan deep learning. *Jurnal Teknologi Informasi dan Komputer*, 9(1), 30–37.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing (3rd ed.)*. Pearson.
- Liswara, R. (2023). *Identifikasi Akun Bot Pada Media Sosial Twitter Menggunakan Metode Extreme Gradient Boosting dan Algoritma Cuckoo Search Sebagai Seleksi Fitur* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- Mastel. (2019). *Hasil survey wabah hoax nasional 2019*. <https://mastel.id/hasil-survey-wabah-hoax-nasional-2019>.
- Nugroho, F. M., & Tanuwijaya, F. (2020). Study of fake news dissemination articles on criminal code regulations, law of information and electronic technology, and law of terrorism criminal act eradication. *International Conference in Bangkok, Thailand: Asia Centre*. Digital Repository Universitas Jember.
- Primandari, D. (2020). Pengenalan metode boosting dalam pembelajaran mesin. *Jurnal Informatika & Sistem Informasi*, 6(1), 22–29.
- Ramraj, S., Nishant, U., Sunil, R., & Shatadeep, B. (2016). Experimenting XGBoost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*, 9(40), 651–662.

- Sanjaya, A., Setiawan, A. B., Mahdiyah, U., Farida, I. N., & Prasetyo, A. R. (2023). Pengukuran Kemiripan Makna Menggunakan Cosine Similarity dan Basis Data Sinonim Kata. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(4), 747-752.
- Sharmila, P., & Suresh, S. (2021). *An enhanced K-means clustering algorithm for data mining. Journal of Data Science*, 19(1), 45-60.
- Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y & Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *arXiv preprint arXiv:2009.05387*.