



Deteksi Berita Hoax Berbahasa Indonesia Menggunakan Multilingual-BERT pada Media Berita Online

Syalma Tarissa^{1,*}, Muhamad Akbar², A. Taqwa Martadinata³, Joni Karman³

¹ Fakultas Ilmu Teknik, Program Studi Informatika, Universitas Bina Insan, Lubuklinggau, Indonesia

² Doktor, Universitas Bina Insan, Lubuklinggau, Indonesia

³ Magister Komputer, Universitas Bina Insan, Lubuklinggau, Indonesia

Email: ^{1,*}2202020018@mhs.univbinainsan.ac.id, ²muhamad.akbar@univbinainsan.ac.id, ³taqwa@univbinainsan.ac.id,

⁴joni_karman@univbinainsan.ac.id

Email Penulis Korespondensi: 2202020018@mhs.univbinainsan.ac.id

Abstrak—Perkembangan teknologi informasi semakin meningkat, hal ini dapat memicu adanya penyebaran berita hoax di media berita online mengakibatkan terjadinya misinformasi bagi masyarakat. Deteksi hoax secara manual menjadi sulit karena besarnya volume berita online serta adanya variasi gaya bahasa dan penggunaan bahasa campuran. Perkembangan teknologi kecerdasan buatan, khususnya Natural Language Processing (NLP), membuka peluang besar dalam mendeteksi berita hoax secara otomatis. Model NLP berbasis Transformer, seperti Multilingual-BERT (mBERT). Model berbasis transformer seperti Multilingual-BERT (mBERT) mampu memahami konteks teks secara dua arah dan mendukung berbagai bahasa termasuk bahasa Indonesia. Oleh karena itu, penelitian ini bertujuan untuk menerapkan dan menguji efektivitas mBERT dalam mendeteksi berita hoax di media berita online. Dataset yang digunakan berasal dari sumber TurnbackHoax di situs Kaggle untuk berita hoax, Tempo dan Cable News Network Indonesia untuk dataset berita non-hoax serta tambahan data berasal dari situs Kaggle pada rentan berita tahun 2025. Data diproses menggunakan Tokenizer bawaan mBERT dengan teknik undersampling data yang menghasilkan model mencapai accuracy 0.97 dengan nilai precision 0.96, recall 0.97, dan F1-score 0.97, yang menunjukkan bahwa pendekatan mBERT mampu memberikan performa klasifikasi yang tinggi pada dataset berita hoax berbahasa Indonesia.

Kata Kunci: Berita Hoax; Natural Language Processing; Transformer Model; Multilingual-BERT; Undersampling

Abstract—The development of information technology is increasing, which can trigger the spread of hoaxes in online news media, resulting in misinformation for the public. Manual detection of hoaxes is difficult due to the large volume of online news and variety of language styles and mixed language usage. Advances in artificial intelligence technology, particularly Natural Language Processing (NLP), open up significant opportunities for automatically detecting hoax news. Transformer-based NLP models, such as Multilingual-BERT (mBERT), are emerging. Transformer-based models such as Multilingual-BERT (mBERT) are capable of understanding text context in both directions and support various languages, including Indonesian. Therefore, this study aims to apply and test the effectiveness of mBERT in detecting hoax news in online news media. The dataset used comes from TurnbackHoax on the Kaggle website for hoax news, Tempo and CNN Indonesia for non-hoax news datasets, and additional data from the Kaggle website on news coverage in 2025. The data was processed using the built-in mBERT Tokenizer with data undersampling techniques, resulting in a model with an accuracy of 0.97, a precision value of 0.96, a recall of 0.97, and F1-score of 0.97, which shows that the mBERT approach is able to provide high classification performance on the Indonesian language hoax news dataset.

Keywords: Hoax News; Natural Language Processing; Transformer Model; Multilingual-BERT; Undersampling

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah meningkat dalam komunikasi berita melalui media daring yang menjadi sumber utama masyarakat dalam memperoleh informasi secara cepat dan mudah. Namun, juga menghadirkan tantangan besar, terutama maraknya penyebaran berita *hoax* yang disebarkan secara sengaja maupun tidak sengaja. Berdasarkan Kamus Besar Bahasa Indonesia (KBBI), *hoax* atau berita palsu diartikan sebagai informasi yang tidak akurat, tetapi secara sengaja dibuat dan disebarkan seolah-olah itu adalah kebenaran. Salah satu ciri-ciri berita yang dapat membedakan apakah suatu berita *hoax* atau tidak adalah gaya penulisan yang provokatif dan berlebihan (Amin et al., 2025)

Berita *hoax* dapat menyesatkan opini publik, mengganggu demokrasi, dan mengurangi kepercayaan masyarakat terhadap pemerintah dan institusi resmi. Dampak sosial dan psikologisnya dapat mencakup polarisasi, konflik kelompok, dan tindakan agresif berbasis misinformasi (Amin et al., 2025). Penyebaran berita palsu atau *fake news* di media digital menjadi permasalahan serius karena dapat mempengaruhi opini publik dan menimbulkan misinformasi. Berbagai pendekatan telah dikembangkan untuk mendeteksi berita palsu, termasuk metode berbasis *machine learning* dan *natural language processing* (NLP) yang memanfaatkan analisis teks untuk mengidentifikasi pola informasi yang tidak valid (Zhou & Zafarani, 2020). Metode deteksi konvensional menjadi sulit karena variasi gaya penulisan, penggunaan bahasa campuran, dan istilah teks maupun *slang* yang sering digunakan dalam berita online. Banyak penelitian sebelumnya menggunakan metode klasik seperti *Naïve Bayes*, *Logistic Regression*, atau SVM dengan fitur TF-IDF. Namun, metode ini memiliki keterbatasan dalam menangkap konteks dan hubungan antar kata.

Perkembangan teknologi kecerdasan buatan, khususnya *Natural Language Processing* (NLP), membuka peluang besar dalam mendeteksi berita *hoax* secara otomatis. *Natural Language Processing* (NLP) merupakan bidang penelitian yang bertujuan untuk memungkinkan komputer memahami dan memproses bahasa manusia secara otomatis (Liu et al., 2025). Salah satu metode NLP yang diterapkan adalah klasifikasi teks. Proses ini mencakup pengelompokan dokumen atau kalimat ke dalam kategori tertentu berdasarkan isinya (Fardhina et al., 2025). Model NLP berbasis

Transformer, seperti *Bidirectional Encoder Representations from Transformer* (BERT). Penggunaan model BERT dalam mendeteksi berita *hoax* berbahasa Indonesia menunjukkan peningkatan akurasi jika dibandingkan dengan metode tradisional (Sukmawati et al., 2025). BERT adalah model pre-trained yang dikembangkan oleh *Google Research* yang dibangun dengan arsitektur *Transformer* dan ditujukan untuk berbagai jenis proses pemrosesan bahasa natural (NLP) (Fardhina et al., 2025).

Salah satu varian BERT adalah *Multilingual-BERT* (mBERT). Memiliki kemampuan mengolah lebih dari 100 bahasa, termasuk Bahasa Indonesia (Hutama & Suhartono, 2022). Hal ini sangat relevan untuk diterapkan pada lingkungan media berita online yang sering menggabungkan konten multibahasa, baik secara formal maupun nonformal. Model mBERT merupakan model *deep learning*, *Deep learning* bagian dari *machine learning* yang menitikberatkan pada model menggunakan jaringan saraf berlapis (Zhang et al., 2023). Model mBERT menggunakan arsitektur *transformer* yang sudah dibekali pengetahuan bahasa pada tahap pre-training, menjadikannya bagian dari metode *deep learning* yang lebih canggih.

Berbagai penelitian terdahulu telah membahas deteksi berita *hoax* menggunakan pendekatan pembelajaran mesin dan *deep learning*, khususnya model berbasis *transformer*. Menerapkan model *transformer* multibahasa seperti mBERT, XLM-R, dan BERTopic untuk klasifikasi berita *hoax* berbahasa Indonesia, dimana mBERT menunjukkan performa yang cukup baik (Hutama & Suhartono, 2022). Namun, penelitian tersebut berfokus pada pemodelan topik berita dan belum mengevaluasi kemampuan mBERT secara spesifik pada berita online umum dengan variasi gaya bahasa yang luas. Penelitian lain Fathin et al. (2024) dan Rahmawati et al. (2022) menunjukkan bahwa model IndoBERT mampu mencapai akurasi tinggi dalam mendeteksi berita *hoax*, terutama pada domain tertentu seperti politik. Meskipun demikian, penggunaan model *monolingual* membatasi kemampuan sistem dalam menangani fenomena bahasa campuran yang sering muncul pada media berita online di Indonesia.

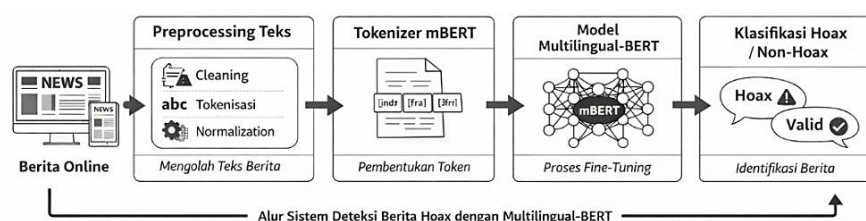
Model mBERT mampu mengungguli BERT dan IndoBERT dalam mendeteksi *hoax Covid-19*, namun ruang lingkup topik yang sempit membatasi generalisasi model (Suadaa et al., 2021). Penelitian Fathin et al. (2024) dan Tobing et al. (2025) membandingkan IndoBERT dan mBERT pada berita *hoax* politik dan menemukan bahwa IndoBERT memiliki performa tinggi, tetapi penelitian tersebut lebih menekankan perbandingan model daripada eksplorasi mendalam mBERT. Peneliti Kamal et al. (2025) dan Brianna et al. (2025) mengkombinasikan BERT dengan arsitektur lain seperti LSTM dan Bi-LSTM, yang menghasilkan akurasi cukup tinggi, namun masih menggunakan model *monolingual* sehingga kurang optimal dalam menangani fenomena bahasa campuran.

Mengkaji kemampuan generalisasi IndoBERT lintas waktu dan memperoleh performa sangat tinggi, tetapi belum melibatkan model multibahasa (Riadi et al., 2025). Di sisi lain, menunjukkan bahwa algoritma *machine learning* klasik seperti *Support Vector Machine* dalam memahami konteks semantik teks (Desriansyah et al., 2025). Menggunakan berbagai model *transformer* seperti BERT, RobBERTa, dan ELECTRA untuk mendeteksi *hoax Covid-19*, namun penelitian tersebut tidak berfokus pada bahasa Indonesia maupun konteks media berita nasional (Jabar et al., 2022).

Berdasarkan berbagai penelitian terdahulu, sebagian besar studi deteksi berita *hoax* berbahasa Indonesia masih menggunakan model *monolingual* seperti IndoBERT atau pendekatan *machine learning* tradisional. Pendekatan tersebut memiliki keterbatasan dalam menangani fenomena bahasa campuran yang sering muncul pada media berita online. Selain itu, beberapa penelitian lebih berfokus pada domain berita tertentu, seperti politik atau Covid-19, sehingga kemampuan generalisasi model terhadap berita online umum masih perlu dievaluasi lebih lanjut. Oleh karena itu, penelitian ini berfokus pada penerapan model *Multilingual-BERT* untuk mendeteksi berita *hoax* pada media berita online Indonesia yang memiliki variasi gaya bahasa dan sumber berita yang beragam.

2. METODOLOGI PENELITIAN

Studi ini menerapkan metode penelitian kuantitatif dengan pendekatan eksperimen komputasional yang bertujuan untuk menguji efisiensi model mBERT dalam mendeteksi berita *hoax* di media berita online berbahasa Indonesia. Pendekatan kuantitatif dipilih karena penelitian ini berorientasi pada pengolahan data teks yang diubah menjadi bentuk numerik untuk selanjutnya dianalisis secara statistik menggunakan model *deep learning*. Metode eksperimen dilakukan dengan merancang, melatih, dan menguji model mBERT dengan menggunakan dataset berita online yang telah dikelompokkan ke dalam dua kategori, yaitu *hoax* dan *non-hoax*. Prosedur eksperimen ini meliputi langkah-langkah *preprocessing* data, penyesuaian model, dan penilaian kinerja model dengan menggunakan berbagai metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*.



Gambar 1. Alur Sistem Deteksi Berita *Hoax* dengan mBERT

2.1 Sumber dan Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan secara daring (online) dengan memanfaatkan data sekunder berupa teks berita yang diperoleh dari berbagai sumber media berita online dan dataset publik. Penggunaan data sekunder dipilih karena penelitian ini berfokus pada eksperimen komputasional dan pemodelan *Natural Language Processing* (NLP) menggunakan *Multilingual-BERT* (mBERT). Tujuan pengumpulan data adalah untuk mendapatkan korpus berita yang representatif, bervariasi, dan relevan untuk melatih dan menguji kemampuan model mBERT untuk mendeteksi berita *hoax*.

Data berita *hoax* berasal dari dataset *TurnBackHoax*, yang dapat diakses di situs *Kaggle* dan mencakup data berita terbaru dari tahun 2025 (*Kaggle*, 2025). Dataset ini dipilih karena telah melalui proses verifikasi dan labeling yang diperlukan untuk menjadi acuan yang dapat diandalkan untuk kategori berita *hoax*. Data berita *non-hoax* atau berita valid diperoleh dari dua sumber utama, yaitu Portal berita online resmi CNN Indonesia dan *Tempo*, yang diambil menggunakan teknik *web scraping* terhadap konten berita publik tahun 2025 dan dataset berita valid tambahan yang bersumber dari *Kaggle* dengan rentang tahun 2025 (*Kaggle*, 2025). Setelah data *hoax* dan *non-hoax* yang telah dikumpulkan digabungkan, proses seleksi awal digunakan untuk menghilangkan data duplikat, tidak lengkap, dan yang tidak relevan dengan tujuan penelitian. Tujuan penggunaan berbagai sumber data adalah untuk meningkatkan variasi teks berita dari segi topik, gaya bahasa, panjang teks, dan struktur kalimat.

Setiap data berita akan diberi label sesuai kategorinya, yaitu “1” untuk berita *hoax* dan “0” untuk berita *non-hoax*. Pada penelitian ini proses pelabelan dilakukan secara otomatis berdasarkan label yang telah tersedia pada dataset *TurnBackHoax* dan dataset berita valid dari *Kaggle*. Untuk mengatasi ketidakseimbangan data antara kelas *hoax* dan *non-hoax*, penelitian ini menerapkan teknik *random undersampling* pada kelas mayoritas sehingga jumlah data pada kedua kelas menjadi seimbang sebelum proses pelatihan model.

2.2 Pra-pemrosesan Data

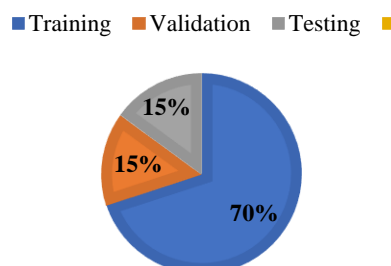
Sebelum teks diproses lebih lanjut oleh model, tahap pra-pemrosesan data teks melibatkan pemrosesan awal untuk menganalisis dan memahami karakteristik data. Termasuk dalam tahapan ini adalah pembersihan teks dari karakter yang tidak relevan, seperti tanda baca berlebih, simbol, angka, spasi ganda, dan pengubahan huruf menjadi huruf kecil. Selain itu, teks dinormalisasi untuk memperbaiki kata yang tidak baku, mengurangi pengulangan huruf yang berlebihan, dan menciptakan bentuk kata yang standar. Untuk kepentingan analisis data, penghapusan *stopword* dilakukan secara selektif dan terbatas. Tujuannya adalah untuk mengurangi kata-kata umum yang tidak memiliki nilai semantik yang signifikan, tetapi tetap mempertahankan konteks penting dari kalimat berita. Analisis data dan eksplorasi teks dibantu oleh tahap pra-pemrosesan ini, bukan sebagai masukan langsung ke model. Pada tahap selanjutnya, *tokenizer* bawaan *Multilingual-BERT* melakukan tokenisasi dan pembentukan representasi input model secara otomatis.

2.3 Tokenizer mBERT

Tokenizer bawaan *Multilingual-BERT* (mBERT), *WordPiece Tokenizer*, digunakan untuk tahap tokenisasi data. Kebutuhan masukan model memungkinkannya mengubah teks menjadi unit token berbasis *subword*. Dengan tokenisasi, model dapat menangani variasi kata, kata tidak baku, dan kosakata yang tidak ada dalam kamus model secara langsung dengan memecah kata menjadi *subword*. *Tokenizer* mBERT memastikan format input data teks yang konsisten sebelum diproses oleh model dengan melakukan proses pemisahan teks, penambahan token khusus, dan penggunaan mekanisme *padding* dan *truncation*. Tokenisasi mBERT berbeda dari pra-pemrosesan klasik karena *tokenizer* mBERT tidak menghapus *stopword* atau proses linguistik manual lainnya sesuai dengan mekanisme tokenisasi dan pemodelan berbasis *transformer* (Wolf et al., 2020).

2.4 Splitting Data

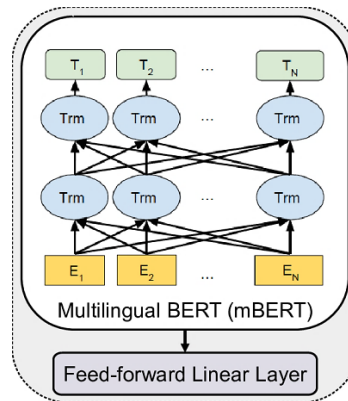
Dataset yang telah melalui tahap praproses dibagi menjadi tiga bagian yaitu, data latih sebesar 70%, data validasi sebesar 15%, dan data uji sebesar 15%. Pembagian ini mengikuti penelitian sebelumnya yang menggunakan rasio serupa untuk klasifikasi teks berbasis deep learning (Tobing et al., 2025). Pengujian dilakukan dengan menggunakan data uji (*testing set*) yang belum pernah dilihat oleh model sebelumnya. Data tersebut dimasukkan ke dalam model yang telah dilatih untuk menghasilkan prediksi berupa kategori “*hoax*” atau “*non-hoax*”.



Gambar 2. Splitting Data

2.5 Model Multilingual-BERT

Multilingual-BERT (mBERT) merupakan varian BERT yang memiliki struktur serupa transformator dan telah dilatih sebelumnya menggunakan teks mentah tanpa anotasi. Tidak seperti BERT yang hanya dilatih dalam satu bahasa, mBERT dilatih dalam 100 lebih bahasa, termasuk bahasa Indonesia, dengan data Wikipedia yang paling besar menggunakan *Masked Language Model* (MLM) (Hutama & Suhartono, 2022). Model Multilingual-BERT dilatih menggunakan arsitektur klasifikasi berbasis transformator dengan penambahan layer fully connected sebagai classification head. Proses pelatihan menggunakan optimizer Adam dengan learning rate sebesar $2e-5$, batch size 16, dan jumlah epoch sebanyak 10. Fungsi loss yang digunakan adalah categorical cross-entropy untuk mengoptimalkan proses klasifikasi antara kelas hoax dan non-hoax.



Gambar 3. mBERT (ResearchGate, 2026)

2.6 Confusion Matrix

Confusion matrix merupakan tabel yang terdiri dari 4 kombinasi berbeda antara nilai prediksi dan nilai yang sebenarnya. Keempat istilah itu adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) (Hanum et al., 2024). Di bawah ini menunjukkan sebuah *confusion matrix* untuk klasifikasi ke dalam dua kategori.

Tabel 1. Visualisasi *Confusion Matrix*

		Aktual	
		Positif	Negatif
Prediksi	Positif	True Positive (TP)	False Positive (FP)
	Negatif	False Negative (FN)	True Negative (TN)

Prediksi yang dihasilkan oleh model, sehingga mempermudah dalam memahami kelebihan dan kelemahan model dalam mengklasifikasikan berita *hoax*. *Confusion matrix* digunakan untuk menghitung berbagai metrik performa guna mengukur kinerja model yang telah dibuat (Hanum et al., 2024). Berikut beberapa *matrix* kinerja yang digunakan:

a. *Precision*

Rasio presisi merupakan ukuran yang menunjukkan seberapa baik model dapat mengenali berita *hoax* secara akurat. Dalam konteks ini, berita *hoax* yang berhasil terdeteksi oleh model adalah *True Positive* (TP), sedangkan berita yang diidentifikasi sebagai *hoax* namun sebenarnya tidak, adalah *False Positive* (FP).

$$P = \frac{TP}{TP+FP} \tag{1}$$

b. *Recall*

Recall merupakan indikator yang menggambarkan sejauh mana model mampu mengidentifikasi semua berita *hoax* yang terdapat dalam dataset. Dalam konteks ini, berita *hoax* yang sebenarnya teridentifikasi oleh model disebut *True Positive* (TP), sedangkan berita *hoax* yang tidak teridentifikasi oleh model disebut *False Negative* (FN).

$$r = \frac{TP}{TP+FN} \tag{2}$$

c. *F1-score*

F1-score merupakan ukuran yang mengintegrasikan presisi dan *recall* menjadi satu nilai tunggal, yang adalah rata-rata harmonis dari kedua *matrix* itu. *F1-score* memberikan representasi yang lebih akurat mengenai performa model secara keseluruhan dibandingkan hanya mempertimbangkan presisi atau *recall* saja.

$$F1 = 2 \times \frac{r \times p}{r+p} \tag{3}$$

d. *Support*

Support digunakan untuk merepresentasikan sampel yang terdapat di setiap kategori set pengujian.

e. *Macro Avg*

Macro avg (average) merupakan metode untuk menghitung nilai rata-rata dari sejumlah matrix evaluasi seperti presisi, recall, atau f1-score untuk semua kelas tanpa memperhitungkan ketidakseimbangan jumlah sampel antara kelas-kelas. Pada tahap ini, matrix evaluasi dihitung terlebih dahulu untuk setiap kelas, setelah itu nilai-nilai tersebut digunakan sebagai dasar untuk menghitung rata-rata macro.

$$Macro\ avg = \frac{\sum \text{Nilai matrik evaluasi semua kelas}}{\text{Jumlah kelas}} \tag{4}$$

f. *Weighted Avg*

Weighted avg merupakan metode untuk menghitung rata-rata dari berbagai matrix evaluasi, seperti akurasi, presisi, recall, dan F1-Score, dengan mempertimbangkan bobot yang berbeda untuk setiap kelas atau kategori dalam matriks evaluasi. Bobot untuk tiap kelas dapat ditentukan oleh beberapa faktor, seperti distribusi kelas dalam dataset atau kepentingan relatif dari setiap kelas dalam evaluasi yang sedang berlangsung.

$$Weighted\ avg = \frac{\sum_{i=1}^n w_i m_i}{\sum_{i=1}^n w_i} \tag{5}$$

3. HASIL DAN PEMBAHASAN

3.1 Scraping Data

Berita *hoax* dan berita *non-hoax* adalah dua kategori data yang dikumpulkan dari beberapa sumber terpercaya guna untuk memastikan keberagaman dan kualitas data. Portal berita nasional dan dataset publik adalah dua sumber data *non-hoax*. Data berita untuk portal berita nasional *Tempo* dan CNN Indonesia dikumpulkan melalui *scraping web* menggunakan metode *RSS parsing*. Proses ini menghasilkan data yang mencakup judul, ringkasan, waktu publikasi, tautan, dan kategori berita.

```
# SIMPAN KE GOOGLE DRIVE
# =====
duplikat = df["title"].duplicated().sum()
print("Jumlah judul duplikat:", duplikat) #hapus duplikat
df = df.drop_duplicates(subset="title").reset_index(drop=True)
df.to_csv(
    "/content/drive/MyDrive/berita/cnn_berita_2025.csv",
    index=False,
    encoding="utf-8-sig"
)

print("Total berita tahun 2025 terkumpul:", len(df))

Drive already mounted at /content/drive; to attempt to forcibly
Parsing RSS: 100% [██████████] 9/9 [13:11<00:00, 87.91s/it]
Jumlah judul duplikat: 0
Total berita tahun 2025 terkumpul: 900

(
  title \
0 BGN Akan Terapkan Sistem Grading Dapur MBG Mul...
1 Cara Daftar Rekrutmen PPPK Tenaga Kependidikan...
2 Surat Prabowo, Gerakan Nurani Bangsa Minta St...
3 Ada Dana On Call Rp 4 Triliun, Banggar DPR: Da...
4 Kemensos Kerahkan Kapal dan Helikopter Salurka...

summary published \
0 Sistem grading dapur MBG akan diasesmen oleh l... 2025-12-04 06:26:46
1 Pemerintah membuka alokasi formasi posisi ini ... 2025-12-04 06:25:45
2 Permintaan penetapan bencana nasional di banjil... 2025-12-04 06:19:01
3 Ketua Banggar DPR menyebut anggaran Rp 4 trili... 2025-12-04 06:13:58
4 Kemensos mengoperasikan distribusi logistik me... 2025-12-04 05:56:35

link category
0 https://nasional.tempo.co/read/2070899/bgn-aka... nasional
1 https://nasional.tempo.co/read/2070897/cara-da... nasional
2 https://nasional.tempo.co/read/2070895/surati... nasional
3 https://nasional.tempo.co/read/2070893/ada-dan... nasional
4 https://nasional.tempo.co/read/2070890/kemenso... nasional
```

Gambar 4. Proses *Scraping Data* Berita *Tempo* dan CNN Indonesia

Penelitian ini juga menggunakan data berita *non-hoax* yang diperoleh dari situs *Kaggle* berjumlah 10.000 Data Berita. Dataset ini terdiri dari kumpulan berita valid tahun 2025 dan diperkaya dengan variasi data dalam hal topik, struktur kalimat, dan gaya bahasa. Data ini digunakan selain data hasil *scraping*. Data berita *hoax* berasal dari dataset *TurnBackHoax*, yang dapat diakses melalui situs *Kaggle* (Kaggle, 2025). Kemudian ditemukan data berita duplikat akan dihapus dan total data bersih menjadi 21.743. Untuk mengimbangi jumlah data di masing-masing kelas, penelitian ini menggunakan teknik *undersampling*. 20.000 data berita (10.000 data *hoax* dan 10.000 data *non-hoax*) dikumpulkan setelah proses penyeimbangan data. Selanjutnya, data ini digunakan untuk *preprocessing*, pelatihan, validasi, dan pengujian model *Multilingual-BERT*.

3.2 Hasil *Preprocessing Data*

Sebelum data digunakan pada tahap pelatihan model *Multilingual-BERT*, proses pra-pemrosesan teks juga dikenal sebagai *preprocessing* dilakukan untuk meningkatkan kualitas data dan membantu analisis karakteristik dataset. Tujuan dari tahap ini adalah untuk menghilangkan elemen-elemen yang tidak relevan dari teks serta menyeragamkan bentuk teks sehingga lebih mudah untuk dianalisis secara linguistik. Perlu ditekankan bahwa pra-pemrosesan yang dilakukan dalam penelitian ini digunakan terutama untuk keperluan eksplorasi data, analisis teks, dan penyajian hasil penelitian, bukan untuk menggantikan proses tokenisasinya.

Tabel 2. *Pre-processing Data*

Original Text	Perubahan iklim tak hanya membebani pelayanan kesehatan, karena membuat kasus semakin naik dan naik, tetapi kami juga menimbang bahwa perubahan iklim akan membebani sistem kesehatan. Sebagai contoh, kekeringan, Penelitian ini lebih jauh memerlukan telaah lanjut untuk menggabungkan indikator iklim ke dalam surveilans berbasis risiko mungkin diperlukan untuk demam berdarah di Indonesia.
Lowercasing	perubahan iklim tak hanya membebani pelayanan kesehatan, karena membuat kasus semakin naik dan



Text	naik, tetapi kami juga menimbang bahwa perubahan iklim akan membebani sistem kesehatan. sebagai contoh, kekeringan, penelitian ini lebih jauh memerlukan telaah lanjut untuk menggabungkan indikator iklim ke dalam surveilans berbasis risiko mungkin diperlukan untuk demam berdarah di indonesia.
Cleaning Text	perubahan iklim tak hanya membebani pelayanan kesehatan karena membuat kasus semakin naik dan naik tetapi kami juga menimbang bahwa perubahan iklim akan membebani sistem kesehatan sebagai contoh kekeringan penelitian ini lebih jauh memerlukan telaah lanjut untuk menggabungkan indikator iklim ke dalam surveilans berbasis risiko mungkin diperlukan untuk demam berdarah di indonesia
Normalisasi Text	perubahan iklim tak hanya membebani pelayanan kesehatan karena membuat kasus semakin naik dan naik tetapi kami juga menimbang bahwa perubahan iklim akan membebani sistem kesehatan sebagai contoh kekeringan penelitian ini lebih jauh memerlukan telaah lanjut untuk menggabungkan indikator iklim ke dalam surveilans berbasis risiko mungkin diperlukan untuk demam berdarah di indonesia
Stopwords Text	perubahan iklim tak hanya membebani pelayanan kesehatan membuat kasus semakin naik naik tetapi kami menimbang bahwa perubahan iklim membebani sistem kesehatan contoh kekeringan penelitian lebih jauh memerlukan telaah lanjut menggabungkan indikator iklim dalam surveilans berbasis risiko mungkin diperlukan demam berdarah indonesia

3.3 Tokenisasi mBERT

Tokenizer ini menggunakan metode *WordPiece*, yang merupakan teknik tokenisasi berbasis sub-kata atau sub-kata. Metode ini memungkinkan model untuk menangani kata tidak baku, kata majemuk, dan variasi bahasa lainnya dengan lebih baik. Setiap kalimat teks dibagi menjadi token yang lebih kecil selama proses tokenisasi. Dalam kasus di mana suatu kata tidak ditemukan secara utuh dalam kosakata model, kata tersebut akan dibagi menjadi beberapa kata terpisah yang masing-masing ditunjukkan dengan awalan "##".

Tabel 3. Tokenisasi mBERT

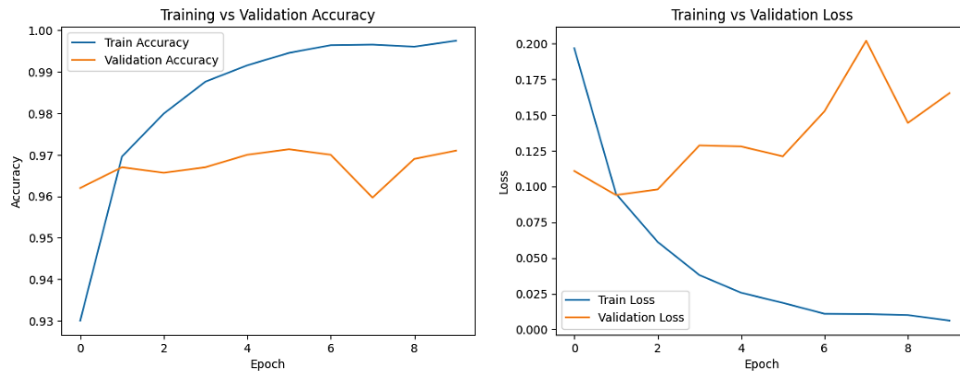
Original Text	Tokenisasi
Perubahan iklim tak hanya membebani pelayanan kesehatan, karena membuat kasus semakin naik dan naik, tetapi kami juga menimbang bahwa perubahan iklim akan membebani sistem kesehatan. Sebagai contoh, kekeringan, Penelitian ini lebih jauh memerlukan telaah lanjut untuk menggabungkan indikator iklim ke dalam surveilans berbasis risiko mungkin diperlukan untuk demam berdarah di Indonesia.	'perubahan', 'iklim', 'tak', 'hanya', 'mem', '##be', '##bani', 'pela', '##yanan', 'kesehatan', ',', 'karena', 'membuat', 'kasus', 'semakin', 'naik', 'dan', 'naik', ',', 'tetapi', 'kami', 'juga', 'meni', '##mbang', 'bahwa', 'perubahan', 'iklim', 'akan', 'mem', '##be', '##bani', 'sistem', 'kesehatan', ',', 'sebagai', 'contoh', ',', 'kek', '##ering', '##an', ',', 'penelitian', 'ini', 'lebih', 'jauh', 'meme', '##rl', '##ukan', 'tela', '##ah', 'lan', '##jut', 'untuk', 'men', '##gga', '##bung', '##kan', 'indi', '##kat', '##or', 'iklim', 'ke', 'dalam', 'sur', '##vei', '##lan', '##s', 'ber', '##basis', 'risiko', 'mungkin', 'dip', '##er', '##lukan', 'untuk', 'dem', '##am', 'ber', '##dar', '##ah', 'di', 'indonesia', ',',

3.4 Pelatihan Model

Tabel 4. Pelatihan Model

Epoch	Loss	Accuracy	Val_loss	Val_acc
Epoch 1	0.1965	0.9300	0.1107	0.9620
Epoch 2	0.0948	0.9696	0.0939	0.9670
Epoch 3	0.0611	0.9799	0.0979	0.9657
Epoch 4	0.0380	0.9876	0.1286	0.9670
Epoch 5	0.0257	0.9916	0.1279	0.9700
Epoch 6	0.0186	0.9946	0.1210	0.9713
Epoch 7	0.0109	0.9964	0.1525	0.9700
Epoch 8	0.0108	0.9966	0.2018	0.9597
Epoch 9	0.0100	0.9961	0.1444	0.9690
Epoch 10	0.0061	0.9975	0.1652	0.9710

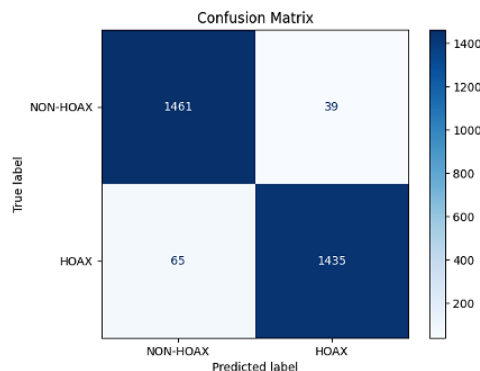
Dengan proses pelatihan selama sepuluh epoch, *model Multilingual-BERT* (mBERT) dilatih menggunakan 70% data latih, 15% data validasi, dan 15% data uji. Berdasarkan hasil pelatihan, akurasi data latih terus meningkat, naik dari 93,00% pada epoch pertama menjadi 99,75% pada epoch ke-10 akurasi validasi berkisar dari 95,97% hingga 97,13%, dengan nilai tertinggi pada epoch ke-6. Pada periode pertama, nilai *loss* pelatihan turun signifikan dari 0,1965 menjadi 0,0061, *Loss validasi* berkisar antara 0,0939 dan 0,2018.



Gambar 5. Grafik perbandingan *training* dan *validation accuracy*, *Training* dan *Validation Loss*

Gambar 5 menunjukkan perbandingan antara ketepatan *training* dan validasi serta *loss* pelatihan dan validasi. Nilai perbedaan antara ketepatan *training* dan validasi berada pada kisaran $\pm 0,02$ yang menunjukkan bahwa perbedaan kinerja antara data pelatihan dan validasi relatif kecil. Grafik kedua kurva menunjukkan tren peningkatan yang relatif konsisten sepanjang proses pelatihan. Namun, pada titik tertentu, terutama ketika keakuratan pelatihan meningkat hingga mendekati nilai maksimumnya, keakuratan validasi tetap berada pada rentang yang stabil. Kondisi ini menunjukkan bahwa ada kecenderungan *overfitting* ringan, yang sering terjadi pada model *deep learning* berbasis *Transformer* seperti *Multilingual-BERT*. Selain itu, fakta bahwa nilai *accuracy*, *precision*, *recall*, dan *F1-score* tetap stabil pada data uji memperkuat kesimpulan bahwa model tidak mengalami *overfitting* yang serius dan masih dapat mengklasifikasikan berita *hoax* dan *non-hoax*.

3.5 Confusion Matrix



Gambar 6. Confusion Matrix

Berdasarkan *confusion matrix* yang dihasilkan, diperoleh hasil sebagai berikut:

- True Negative* (TN) atau berita *non-hoax* yang diprediksi dengan benar sebanyak 1.461 data.
- False Positive* (FP) atau berita *non-hoax* yang salah diklasifikasikan sebagai *hoax* sebanyak 39 data.
- False Negative* (FN) atau berita *hoax* yang salah diklasifikasikan sebagai *non-hoax* sebanyak 65 data.
- True Positive* (TP) atau berita *hoax* yang diprediksi dengan benar sebanyak 1.435 data.

Pada Gambar 6. *confusion matrix*, jumlah *False Negative* (65 data) lebih tinggi dibandingkan *False Positive* (39 data). Hal ini menunjukkan bahwa model cenderung lebih konservatif dalam mengklasifikasikan berita sebagai *hoax*. Dengan kata lain, model lebih berhati-hati agar tidak salah mengklasifikasikan berita valid sebagai *hoax*. Namun, kondisi ini juga menunjukkan bahwa masih terdapat beberapa berita *hoax* yang tidak berhasil terdeteksi oleh model. Fenomena ini kemungkinan disebabkan oleh gaya bahasa *hoax* yang menyerupai berita faktual sehingga sulit dibedakan oleh model.

	precision	recall	f1-score	support
NON-HOAX	0.96	0.97	0.97	1500
HOAX	0.97	0.96	0.97	1500
accuracy			0.97	3000
macro avg	0.97	0.97	0.97	3000
weighted avg	0.97	0.97	0.97	3000

Gambar 7. *Classification Report*



Evaluasi performa model dilakukan menggunakan metrik *precision*, *recall*, dan *F1-score*, hasil evaluasi menunjukkan:

- Kelas *Non-Hoax* memiliki nilai *precision* sebesar 0.96, *recall* 0.97, dan *F1-score* 0.97.
- Kelas *Hoax* memiliki nilai *precision* sebesar 0.97, *recall* 0.96, dan *F1-score* 0.97.
- Nilai *accuracy*, *macro average* dan *weighted average* untuk *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,97, yang menunjukkan bahwa performa model seimbang pada kedua kelas.

3.6 Prediksi Model

Pada tahap ini, pengujian prediksi model dilakukan terhadap beberapa contoh teks berita untuk mengevaluasi kemampuan model mBERT untuk secara langsung membedakan antara berita *hoax* dan *non-hoax*. Pengujian dilakukan menggunakan fungsi *detect_hoax* dengan memasukkan kalimat berita.

<code>detect_hoax("kemdiktisaintek berencana meluncurkan skema student loan agustus atau september 2025")</code>
<code>'NON-HOAX (probabilitas 0.9869)'</code>
<code>detect_hoax("Dishub Jakarta Akan Kaji Usulan UPT Parkir Dibubarkan dan Dialihkan ke Swasta")</code>
<code>'NON-HOAX (probabilitas 0.9820)'</code>
<code>detect_hoax("Muhammad Qodari Usulkan Gibran Jadi Pahlawan Nasional Jasanya Sudah Banyak Buat Bangsa dan Negara Melebihi Soekarno dan Soeharto")</code>
<code>'HOAX (probabilitas 0.9973)'</code>
<code>detect_hoax("Agung Zainal Mutaqin Presiden Prabowo resmi copot Bahilil dari kabinet? Presiden Prabowo merasa aneh ke Jokowi. Kok bisa milih bahilil jadi menteri</code>
<code>'HOAX (probabilitas 0.9994)'</code>

Gambar 8. Prediksi Model

Hasil prediksi menunjukkan bahwa model memiliki kemampuan untuk menyediakan klasifikasi yang sangat kepercayaan (probabilitas). Pada contoh pertama dan kedua, teks berita tentang rencana kebijakan Kemdiktisaintek mengenai pinjaman siswa serta penyelidikan Dishub Jakarta tentang pengelolaan UPT parkir diprediksi sebagai NON-HOAX dengan probabilitas 0.9869. Model sangat percaya pada keputusan klasifikasinya, seperti yang ditunjukkan oleh nilai probabilitas yang mendekati 1. Ini menunjukkan bahwa model memiliki kemampuan untuk mendeteksi pola bahasa berita faktual yang informatif dan tidak provokatif. Namun, dalam contoh ketiga dan keempat, model memprediksi teks sebagai HOAX dengan probabilitas 0.9973 dan 0.9994, masing-masing. Klaim berlebihan, pendapat subjektif, dan narasi yang cenderung provokatif termasuk dalam teks tersebut. Pola seperti ini mudah dikenali karena sering ditemukan dalam berita *hoax*. Nilai probabilitas yang tinggi menunjukkan bahwa model sangat yakin dalam mendeteksi indikasi *hoax* dalam teks. Secara keseluruhan, hasil prediksi ini memperkuat temuan pengujian kuantitatif sebelumnya bahwa model mBERT memiliki kemampuan yang baik dalam mengklasifikasikan berita *hoax* dan *non-hoax*. Model tidak hanya menunjukkan data uji yang sangat akurat, tetapi juga konsisten dalam membuat prediksi yang logis dan meyakinkan pada contoh kasus nyata. Oleh karena itu, model ini dapat digunakan sebagai sistem pendukung untuk proses deteksi *hoax* berita otomatis di Indonesia.

3.7 Pembahasan

Hasil pengujian menunjukkan bahwa model *Multilingual-BERT* (mBERT) sangat baik dalam mengklasifikasikan berita *hoax* dan *non-hoax*. Hasil pengujian menunjukkan model mencapai akurasi 0.97 dengan nilai *precision* 0.96, *recall* 0.97, *F1-score* 0.97 untuk label *non-hoax* dan nilai *precision* 0.97, *recall* 0.96, *F1-score* 0.97 untuk label *hoax*. Hasil ini menunjukkan bahwa model dapat mengenali pola *hoax* tanpa bias terhadap kelas. Kemampuan mBERT untuk memahami konteks bahasa secara dua arah (*bidirectional*) mempengaruhi kinerja tinggi model. Dibandingkan dengan pendekatan klasik berbasis fitur manual atau pengajaran mesin, mekanisme ini memungkinkan model untuk mengidentifikasi hubungan antar kata dalam satu kalimat secara lebih menyeluruh. Selain itu, tokenisasi *WordPiece* pada mBERT telah terbukti efektif dalam menangani variasi kata, kata tidak baku, dan istilah asing yang sering muncul dalam berita online berbahasa Indonesia.

Hasil penelitian ini sejalan dengan penelitian yang menyatakan bahwa model transformer multibahasa seperti mBERT memiliki performa yang kompetitif dalam klasifikasi berita *hoax* berbahasa Indonesia (Hutama & Suhatono, 2022). Namun, penelitian tersebut lebih berfokus pada pemodelan topik, sedangkan penelitian ini secara khusus mengevaluasi kemampuan mBERT dalam mengklasifikasi berita *hoax* dan *non-hoax* lintas topik dan sumber media berita online. Dengan demikian, penelitian ini memperluas temuan sebelumnya dengan menunjukkan bahwa mBERT tidak hanya efektif pada domain tertentu, tetapi juga mampu melakukan generalisasi yang baik pada berita online umum.

Teknik *undersampling* juga membantu meningkatkan kinerja model karena menyeimbangkan distribusi data antara kelas *hoax* dan *non-hoax*. Ketidakseimbangan data sering kali menyebabkan model memprediksi kelas mayoritas, sehingga menurunkan kemampuan deteksi kelas minoritas. Dengan melakukan *undersampling* pada kelas mayoritas, model dilatih menggunakan data yang lebih proporsional, sehingga mampu mempelajari karakteristik kedua kelas dengan lebih baik. Ini ditunjukkan oleh nilai *precision* dan *recall* yang seimbang, yang menunjukkan bahwa model tidak hanya dapat menemukan informasi yang tidak benar tetapi juga dapat menemukan informasi yang benar. Namun, kelemahan penerapan *undersampling* adalah bahwa model tidak dapat menggunakan lebih banyak data



latih. Sebagian informasi penting yang sebenarnya relevan untuk pembelajaran mungkin hilang jika data ini dikurangi. Karena mBERT sebagai model pralatih memiliki pemahaman linguistik yang kuat, dampak tersebut tidak terlihat signifikan dalam penelitian ini. Oleh karena itu, model masih mampu melakukan klasifikasi yang optimal meskipun jumlah data latihan berkurang.

Secara keseluruhan, hasil pengujian menunjukkan bahwa kombinasi model mBERT dan strategi penyeimbangan data mampu mengklasifikasikan berita *hoax* berbahasa Indonesia dengan sangat baik. Hasilnya menunjukkan bahwa pendekatan berbasis *transformer* sangat mungkin diterapkan pada sistem deteksi *hoax* otomatis. Untuk meningkatkan validitas dan generalisasi model, pengujian pada dataset yang lebih beragam dan studi metode penanganan data tidak seimbang lainnya, seperti *oversampling* berbasis sintesis (SMOTE) atau pembelajaran yang sensitif terhadap biaya, dapat membantu meningkatkan kinerja model.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa model *Multilingual-BERT* (mBERT) berhasil mengklasifikasikan *hoax* dan *non-hoax* dalam bahasa Indonesia. Penelitian ini menggunakan data berita dari sejumlah dataset dan media yang dapat dipercaya contohnya, portal berita *Tempo* dan CNN Indonesia menggunakan *scraping web*, serta dataset *Kaggle*, yang mencakup data *non-hoax* dan *hoax* dari *TurnBackHoax*. Tujuan dari penggunaan berbagai sumber data adalah untuk meningkatkan keberagaman dan kualitas dataset yang digunakan dalam proses pelatihan dan pengujian model. Pada tahap pengolahan data, *preprocessing* dan pembersihan data dilakukan untuk menghilangkan duplikasi dan ketidak konsistenan data. Setelah data duplikat dihilangkan, data bersih dikumpulkan dan diseimbangkan menggunakan teknik *undersampling* untuk mengatasi masalah ketidak seimbangan kelas. Proses ini menghasilkan dataset dengan jumlah data *hoax* dan *non hoax* yang sama, yang memungkinkan model untuk dilatih tanpa dengan lebih baik. Hasil pengujian model yang melibatkan 3.000 data uji menunjukkan model mencapai akurasi 0.97 dengan nilai *precision* 0.96, *recall* 0.97, *F1-score* 0.97 untuk label *non-hoax* dan nilai *precision* 0.97, *recall* 0.96, *F1-score* 0.97 untuk label *hoax*, yang menunjukkan bahwa model memiliki kemampuan untuk mengklasifikasikan berita *hoax* dan *non-hoax* dengan tingkat ketepatan yang tinggi. Selain itu, evaluasi lebih lanjut menggunakan *confusion matrix* menunjukkan bahwa jumlah prediksi yang benar untuk kedua jenis berita jauh lebih besar daripada jumlah kesalahan prediksi. Nilai *false positive* dan *false negative* yang rendah menunjukkan bahwa model dapat membedakan antara *hoax* dan valid. Secara keseluruhan, temuan penelitian ini menunjukkan bahwa tujuan penelitian untuk membangun sistem yang berbasis pembelajaran mendalam untuk klasifikasi berita *hoax* dan *non-hoax* telah tercapai, karena model *Multilingual-BERT* mampu mempelajari pola bahasa secara kontekstual dan menghasilkan hasil yang stabil pada data uji. Meskipun hasil penelitian menunjukkan performa yang tinggi, penelitian ini masih memiliki keterbatasan pada jumlah dan keragaman dataset yang digunakan. Selain itu, penelitian ini belum membandingkan performa model mBERT dengan model *transformer* lain seperti IndoBERT atau XLM-R. Penelitian selanjutnya dapat mengembangkan eksperimen dengan dataset yang lebih besar dan melakukan perbandingan dengan berbagai model *transformer* lainnya untuk memperoleh pemahaman yang lebih komprehensif mengenai metode deteksi berita *hoax* berbahasa Indonesia.

REFERENCES

- Amin, M. Z. A., Furqon, M. A., & Wijonarko, D. (2025). Deteksi Berita Hoaks Berbahasa Indonesia Menggunakan One-Dimensional Convolutional Neural Network. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 14(2), 161–169. <https://doi.org/10.22146/jnteti.v14i2.19050>
- Brianna, D. F., Paisal, P., Saputra, M. A., & Hapiz, M. A. (2025). Model Deteksi Berita Palsu Menggunakan BERT dan Bi-LSTM Berbasis Discriminative Approach. *JSAI (Journal Scientific and Applied Informatics)*, 8(3), 626–631. <https://doi.org/10.36085/jsai.v8i3.9384>
- Desriansyah, M. D., Sari, I. U., & Zulfahmi, Z. (2025). Analisis Efektivitas Algoritma Machine Learning dalam Deteksi Hoaks: Pada Berita Digital Berbahasa Indonesia. *Jurnal Sistem Informasi Dan Informatika*, 3(2), 63–69. <https://doi.org/10.47233/jiska.v3i1.2024>
- Deteksi Berita Hoaks Indo—Dataset. (n.d.). Retrieved December 27, 2025, from <https://www.kaggle.com/datasets/mochamadabdulazis/deteksi-berita-hoaks-indo-dataset>
- Fardhina, A., Siregar, R. M., Sibarani, M. R. W. B., Ginting, I. C. B., & Pratama, A. (2025). Sistem Deteksi Berita Hoaks berbasis Algoritma Natural Language Processing (NLP) menggunakan BERT. *Jurnal Manajemen Informatika, Sistem Informasi Dan Teknologi Komputer (JUMISTIK)*, 4(1), 450–461. <https://doi.org/10.70247/jumistik.v4i1.156>
- Fathin, M. A., Sibaroni, Y., & Prasetyowati, S. S. (2024). Handling Imbalance Dataset on Hoax Indonesian Political News Classification using IndoBERT and Random Sampling. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(1), 352–360. <https://doi.org/10.30865/mib.v8i1.7099>
- Figure 2: Multilingual BERT (mBERT) model. (n.d.). ResearchGate. Retrieved January 4, 2026, from https://www.researchgate.net/figure/Multilingual-BERT-mBERT-model_fig2_362263410
- Hanum, A. R., Zetha, I. A., Putri, S. C., Wulandari, R. A., Andina, S. P., Fajrina, J. N., & Yudistira, N. (2024). Analisis Kinerja Algoritma Klasifikasi Teks Bert dalam Mendeteksi Berita Hoaks. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(3), 537–546. <https://doi.org/10.25126/jtiik.938093>



- Hutama, L. B., & Suhartono, D. (2022). Indonesian Hoax News Classification with Multilingual Transformer Model and BERTopic. *Informatica*, 46(8). <https://doi.org/10.31449/inf.v46i8.4336>
- Indonesia News Dataset (2025). (n.d.). Retrieved December 27, 2025, from <https://www.kaggle.com/datasets/sh1zuka/indonesia-news-dataset-2024>
- Jabar, B. A., Seline, S., Bintang, B., Victoria, C. J., & Arifin, R. N. (2022). COVID-19 Fake News Detection With Pre-trained Transformer Models. *Ultimatics: Jurnal Teknik Informatika*, 14(2), 51–56. <https://doi.org/10.31937/ti.v14i2.2776>
- Kamal, A. M., Chrisnanto, Y. H., & Yuniarti, R. (2025). Identifikasi Berita Palsu di Portal Media Online Menggunakan Model IndoBERT dan LSTM. *JURNAL RISET KOMPUTER (JURIKOM)*, 12(3), 287–297. <https://doi.org/10.30865/jurikom.v12i3.8660>
- Rahmawati, A., Alamsyah, A., & Romadhony, A. (2022). Hoax News Detection Analysis using IndoBERT Deep Learning Methodology. *2022 10th International Conference on Information and Communication Technology (ICoICT)*, 368–373. <https://doi.org/10.1109/ICoICT55009.2022.9914902>
- Representation Learning for Natural Language Processing | Springer Nature Link (formerly SpringerLink)*. (n.d.). Retrieved December 27, 2025, from <https://link.springer.com/book/10.1007/978-981-99-1600-9>
- Riadi, A. T., Indriani, F., Mazdadi, M. I., Faisal, M. R., & Herteno, R. (2025). Cross-Temporal Generalization of IndoBERT for Indonesian Hoax News Classification. *Jurnal Teknik Informatika (Jutif)*, 6(5), 5291–5304. <https://doi.org/10.52436/1.jutif.2025.6.5.4757>
- Suadaa, L. H., Santoso, I., & Panjaitan, A. T. B. (2021). Transfer Learning of Pre-trained Transformers for Covid-19 Hoax Detection in Indonesian Language. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), 317–326. <https://doi.org/10.22146/ijccs.66205>
- Sukmawati, E. C., Suryaningrum, L., Angelica, D., & Ramadhan, N. G. (2025). *Klasifikasi Berita Palsu Menggunakan Model Bidirectional Encoder Representations From Transformers (BERT) | SISINFO: Jurnal Sistem Informasi dan Informatika*. https://jurnalunibi.unibi.ac.id/ojs/index.php/SisInfo/article/view/934?utm_source=chatgpt.com
- Tobing, C. J. L., Wijayakusuma, I. L., & Harini, L. P. I. (2025). Perbandingan Kinerja IndoBERT dan MBERT Untuk Deteksi Berita Hoaks Politik dalam Bahasa Indonesia. *JST (Jurnal Sains dan Teknologi)*, 14(1), 114–123. <https://doi.org/10.23887/jstundiksha.v14i1.92126>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. In Q. Liu & D. Schlangen (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
- Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, 53(5), 1–40. <https://doi.org/10.1145/3395046>