SISTEM DETEKSI BERITA PALSU DUA BAHASA MENGGUNAKAN TF-IDF DAN MULTINOMIAL NAIVE BAYES

(Bilingual Fake News Detection Using TF-IDF and Multinomial Naive Bayes)

Rheno Septianto*[1], Yan Rianto[2]
[1,2]Computer Science Faculty, Universitas Nusa Mandiri, Indonesia
Depok,Jawa Bara Indonesia
Email: [1]14240011@nusamandiri.ac.id, [2]yan.yrt@nusamandiri.ac.id

Abstract

The rapid spread of misinformation poses a major threat to public trust and digital literacy. This study develops a bilingual fake news detection system capable of analyzing news content in English and Indonesian. The system uses two separate monolingual models trained independently on the WELFake dataset (English) and the Berita Hoax 2023 dataset (Indonesian). Each model applies text preprocessing techniques such as tokenization, stopword removal, and normalization before transforming the text using TF-IDF. The classification process utilizes the Multinomial Naïve Bayes algorithm, chosen for its efficiency in handling high-dimensional text data. The bilingual system integrates an automatic language detection module that selects the appropriate model based on the detected language. Evaluation results show that the English model achieves an accuracy of 86%, while the Indonesian model achieves an accuracy of 93%. These results indicate that the two-model bilingual approach provides reliable performance for multilingual fake news detection. This study contributes to practical solutions for misinformation mitigation, especially in multilingual environments like Indonesia.

Keywords: *Fake News Detection, Bilingual System,, TF-IDF, NLP, Multinomial Naïve Bayes*

**Corresponding Author*

1. PENDAHULUAN

Penyebaran informasi palsu atau *fake news* semakin menjadi perhatian dunia karena dapat memengaruhi pandangan publik dan menurunkan kepercayaan masyarakat terhadap lembaga resmi. Misinformasi mudah beredar melalui media sosial dan platform berita daring, sehingga opini masyarakat dapat terbentuk oleh informasi yang tidak akurat. Kondisi ini semakin serius ketika konten yang bersifat emosional justru lebih cepat menyebar daripada informasi faktual, terlebih karena penyebarannya kini melampaui batas bahasa dan budaya [1].

Dalam kehidupan sehari-hari, masyarakat tidak hanya bergantung pada satu bahasa untuk memperoleh informasi. Contohnya, pembaca berita di Indonesia sering kali mengakses informasi dari sumber lokal berbahasa Indonesia dan berita internasional berbahasa Inggris. Kondisi ini menjadikan permasalahan misinformasi bersifat multibahasa. Istilah multilingual mengacu pada penggunaan banyak bahasa, sedangkan bilingual terbatas pada dua bahasa. Walaupun lebih sempit, konteks bilingual tetap menghadirkan tantangan tersendiri karena masing-masing bahasa memiliki struktur kalimat, kosakata, dan gaya penulisan yang berbeda.

Menangani misinformasi pada satu bahasa saja sudah cukup menantang, terutama karena variasi makna, gaya penulisan, dan struktur bahasa dapat memengaruhi pola teks dalam berita palsu. Tantangan ini semakin besar ketika sistem harus mampu bekerja pada dua bahasa yang memiliki karakter linguistik berbeda. Sebagian besar penelitian terdahulu lebih banyak berfokus pada deteksi berita palsu dalam bahasa Inggris, atau hanya menggunakan pendekatan multilingual umum tanpa mempertimbangkan perbedaan mendalam antarbahasa [2].

Penelitian sebelumnya juga menekankan pentingnya pemilihan fitur yang tepat dalam membedakan berita asli dan palsu. Namun sebagian pendekatan masih cenderung mengandalkan dataset bahasa Inggris, sehingga variasi budaya dan linguistik pada bahasa lain—khususnya bahasa Indonesia—belum sepenuhnya terakomodasi [3]. Padahal, pola linguistik yang khas pada setiap bahasa memiliki pengaruh besar terhadap keberhasilan deteksi misinformasi.

Untuk menjawab kesenjangan tersebut, penelitian ini mengembangkan sistem deteksi berita palsu dua bahasa (Indonesia dan Inggris) menggunakan pendekatan *Artificial Intelligence* (AI) dan *Natural Language Processing* (NLP). Sistem

dibangun menggunakan dua model terpisah yang masing-masing dilatih dengan dataset sesuai bahasa: WELFake untuk bahasa Inggris dan Berita Hoax 2023 untuk bahasa Indonesia. Pendekatan ini dipilih agar model dapat menangkap pola linguistik yang lebih spesifik pada tiap bahasa. Metode Multinomial Naïve Bayes dipilih karena efisien dan mampu bekerja dengan baik pada data teks berdimensi tinggi.

Sebelum proses klasifikasi, teks terlebih dahulu melalui tahapan praproses seperti tokenisasi, normalisasi huruf kecil, serta penghapusan stopword. Teks yang telah dibersihkan kemudian diubah menjadi representasi numerik menggunakan TF-IDF. Pada tahap implementasi, sistem dilengkapi deteksi bahasa otomatis untuk memilih model yang sesuai sebelum proses prediksi dilakukan.

Secara keseluruhan, penelitian ini berupaya menyediakan solusi yang dapat membantu mendeteksi berita palsu dalam dua bahasa berbeda secara akurat. Sistem ini diharapkan dapat mendukung upaya peningkatan literasi digital dan membantu masyarakat dalam mengidentifikasi informasi yang dapat dipercaya.

2. TINJAUAN PUSTAKA

Penelitian mengenai deteksi berita palsu telah berkembang pesat dalam beberapa tahun terakhir, terutama seiring meningkatnya penggunaan media sosial dan platform digital sebagai sumber informasi utama. Berbagai pendekatan telah digunakan, mulai dari metode berbasis linguistik, algoritma *machine learning* klasik, hingga model *deep learning* dan pendekatan multimodal. Masing-masing metode memiliki keunggulan serta keterbatasan, khususnya ketika diterapkan pada konteks multibahasa seperti bahasa Indonesia dan Inggris.

Untuk memberikan gambaran menyeluruh mengenai perkembangan penelitian sebelumnya, Tabel 1 berikut merangkum berbagai studi terkait deteksi berita palsu beserta teknik, akurasi, dan kontribusi utama yang dilaporkan. Ringkasan ini membantu mengidentifikasi posisi penelitian ini, sekaligus menyoroti celah penelitian (*research gap*) yang menjadi dasar pengembangan sistem deteksi berita palsu dua bahasa.

Beberapa penelitian bahkan melaporkan akurasi di atas 90% pada dataset seperti Twitter dan Weibo. Perkembangan terbaru menunjukkan bahwa penelitian telah berkembang ke arah pendekatan lintas-modal (*cross-modal*), yang menggabungkan mekanisme seperti self-attention, contrastive learning, serta graph neural networks. Untuk meningkatkan kemampuan generalisasi dan ketahanan model dalam menghadapi variasi topik dan bahasa, beberapa studi juga mengombinasikan fitur tekstual, visual, dan pola penyebaran informasi. Pendekatan ini bertujuan menghasilkan model deteksi berita palsu yang lebih adaptif dan akurat pada berbagai konteks.

Tabel I.Tinjauan Pustaka

Studi	Teknik	Akurasi	Insight Utama
Pérez-Rosas et al. [3]	Analisis Fitur Linguistik	78%	Menjadi tolok ukur awal deteksi satu bahasa dan menyoroti tantangan konteks linguistik.
Shu et al. [2]	Rekayasa Fitur	Tidak dilaporkan	Fokus pada profil pengguna, tetapi tidak efektif untuk konteks lintas bahasa.
Kumar Sutradhar et al. [7]	Naïve Bayes, Logistic Regression	56%	Naïve Bayes paling baik di antara metode klasik, namun akurasinya masih rendah.
Kurniawan & Mustikasari [8]	CNN–LSTM	Lebih tinggi pada CNN	CNN lebih unggul untuk teks bahasa Indonesia, menunjukkan kekuatan deep learning dalam konteks multilingual.
Jiang et al. [9]	Multi-Task Learning (Sentimen & Stance)	Tidak dilaporkan	Mengusulkan model multitugas untuk meningkatkan pemahaman konteks berita palsu.
Alghamdi et al. [10]	Hybrid Summarization + mBERT	Tidak dilaporkan	Efektif untuk bahasa dengan sumber data terbatas

Studi	Teknik	Akurasi	Insight Utama
			melalui ringkasan dan model multilingual.
Malik et al. [11]	Ensemble Graph Neural Networks	Tidak dilaporkan	Menggabungkan pola keterlibatan pengguna dan fitur teks untuk deteksi lebih kuat.
Bazmi et al. [12]	Transformer Multi-Domain Berbasis Entitas	72%	Meningkatkan generalisasi model pada domain yang belum pernah dilatih.
Yang et al. [13]	Dual-Stream Fusion + Self-Attention	>90%	Memanfaatkan multimodal learning dan self-attention untuk hasil sangat akurat.
Wang et al. [14]	Miner-UVS + PU Learning	Tidak dilaporkan	Mengatasi konflik verifikasi pada fitur multimodal.
Shan et al. [15]	Cross-Modal Aggregation & Gated Fusion	91.8% (Twitter), 88.7% (Weibo)	Menggabungkan pesan multimodal untuk meningkatkan ketepatan model.
Chen et al. [16]	Contrastive Learning + Propagation Network	Tidak dilaporkan	Integrasi teks, gambar, dan pola penyebaran untuk ketahanan model.
Pawlicka et al. [17]	AI + Analisis Linguistik	Tidak dilaporkan	Mengkaji sinergi antara model AI dan analisis linguistik.
Dhiman et al. [18]	GBERT (Hybrid GPT–BERT)	Tidak tersedia	Menggabungkan kemampuan GPT dan BERT untuk klasifikasi yang lebih kuat.
Wu et al. [19]	Style Clustering + Contrastive Learning	Tidak dilaporkan	Mengatasi perbedaan kategori dan domain dengan pengelompokan gaya.
Yan et al. [20]	Multi-Granularity Fusion + Contrastive Learning	Tidak dilaporkan	Meningkatkan deteksi melalui penyelarasan fitur multimodal.
Han et al. [21]	Multifaceted Reasoning Network	92.9%	Menggunakan graf heterogen untuk deteksi yang dapat dijelaskan (*explainable*).
Yu et al. [22]	Dual Evidence Perception	3.60%	Menggabungkan bukti historis dan eksternal, namun performanya rendah pada dataset tertentu.

Dengan semakin mudahnya masyarakat mengakses informasi di era digital, berbagai penelitian telah dilakukan untuk meningkatkan kemampuan sistem dalam mendeteksi berita palsu. Penelitian-penelitian tersebut mencakup pendekatan monolingual, multilingual, serta penggunaan metode *machine learning* dan *deep learning*, yang secara umum menunjukkan hasil yang menjanjikan untuk berbagai bahasa dan teknik pemrosesan.

Sebagian besar metode deteksi berita palsu pada penelitian terdahulu masih bersifat monolingual dan banyak berfokus pada dataset berbahasa Inggris. Pendekatan yang umum digunakan adalah analisis

fitur linguistik, yang mencapai akurasi sekitar 78% pada salah satu penelitian [7]. Pola sintaksis, analisis sentimen, serta penggunaan kosakata tertentu terbukti efektif dalam membedakan berita asli dan berita palsu, meskipun metode ini kurang dapat diterapkan pada bahasa lain yang memiliki struktur berbeda [3].

Deteksi berita palsu dalam konteks multilingual dan bilingual masih menghadapi tantangan karena perbedaan sintaksis, makna, dan budaya antarbahasa. Studi mengenai transfer pengetahuan dari model bahasa Inggris ke model multibahasa juga masih terbatas. Meskipun model berbahasa Inggris telah berkembang pesat, model tersebut belum mampu menangkap kerumitan deteksi berita palsu pada lingkungan multilingual, sehingga sulit diterapkan secara *cross-lingual* [2].

Metode *machine learning* klasik seperti Naïve Bayes dan Logistic Regression juga banyak digunakan dalam penelitian deteksi berita palsu [23]. Pendekatan ini umumnya memanfaatkan teknik ekstraksi fitur seperti *bag-of-words* (BoW) dan TF-IDF. Dalam sebuah studi, Naïve Bayes mencatat akurasi tertinggi sebesar 56%, namun hasil tersebut masih dianggap rendah karena berita palsu sering kali memiliki ciri fitur yang lebih halus dan kompleks. Meskipun mudah diterapkan, metode tradisional cenderung kurang mampu melakukan generalisasi dan tidak dapat menangkap informasi kontekstual yang lebih mendalam.

Kemajuan teknologi *deep learning* membawa peningkatan signifikan dalam performa deteksi berita palsu. Model seperti Convolutional Neural Networks (CNN) dan Long Short-Term Memory (LSTM) terbukti mampu menangkap struktur dan pola teks yang lebih kompleks. Kombinasi TF-IDF dengan CNN atau DNN bahkan mampu mencapai akurasi sekitar 84,6%. Namun, model-model ini memerlukan sumber daya komputasi besar dan dataset berlabel yang lebih luas, sehingga aplikasinya pada bahasa dengan sumber data terbatas masih menjadi kendala.

Pendekatan hibrida mulai dikembangkan untuk mengatasi kekurangan metode tradisional maupun model *deep learning* murni. Salah satu contohnya adalah model CNN-LSTM yang digunakan dalam penelitian deteksi berita palsu berbahasa Indonesia. Penggabungan CNN dan LSTM memungkinkan model mengekstraksi fitur lokal sekaligus memahami hubungan sekuensial, sehingga memberikan performa lebih baik pada bahasa dengan sumber daya rendah seperti Indonesia.

Pendekatan terbaru seperti WELFake mengombinasikan *word embeddings* dengan fitur linguistik tradisional untuk meningkatkan akurasi deteksi berita palsu. Penggunaan embedding memungkinkan model mempelajari pola bahasa yang lebih halus, sehingga menghasilkan performa yang lebih baik dan lebih stabil dibandingkan metode sebelumnya. Studi menggunakan pendekatan ini menunjukkan peningkatan ketahanan model terhadap variasi bahasa dan jenis berita.

3. METODE PENELITIAN

3.1 Deskripsi Modul

Sistem deteksi berita palsu ini terdiri atas beberapa modul yang saling mendukung dan dijalankan secara berurutan. Modul pertama adalah Modul Pengelolaan Dataset, yang berisi dua dataset monolingual: WELFake sebagai sumber data berbahasa Inggris dan Berita Hoax 2023 sebagai sumber data berbahasa Indonesia. Kedua dataset tersebut menjadi dasar dalam proses pelatihan masing-masing model.

Berikutnya, Modul Deteksi Bahasa menggunakan pustaka *langdetect* untuk mengenali bahasa dari teks yang dimasukkan pengguna. Hasil identifikasi bahasa kemudian menentukan jalur praproses dan model mana yang akan digunakan. Setelah bahasa diketahui, Modul Praproses Teks melakukan tokenisasi, konversi huruf menjadi huruf kecil, serta penghapusan stopword sesuai bahasa agar teks dalam kondisi bersih dan seragam sebelum diekstraksi.

Tahap selanjutnya adalah Modul Ekstraksi Fitur, yang memanfaatkan TF-IDF untuk mengubah teks menjadi vektor numerik yang mencerminkan tingkat kepentingan kata. Hasil ekstraksi ini kemudian digunakan oleh Modul Pembelajaran Mesin, yang melatih dua model Multinomial Naïve Bayes secara terpisah untuk bahasa Indonesia dan bahasa Inggris. Model serta vectorizer yang telah dilatih disimpan dalam format .pkl untuk dapat digunakan kembali pada sistem web. Melalui rangkaian modul ini, sistem mampu mengklasifikasikan berita palsu dalam dua bahasa dengan cara yang efisien dan tetap mempertahankan akurasi yang baik.

3.2 Data Set

Penelitian ini memanfaatkan dua dataset utama untuk membangun sistem deteksi berita palsu dalam dua bahasa. Untuk bahasa Inggris, digunakan dataset WELFake yang berisi lebih dari 72.000 artikel yang telah diklasifikasikan sebagai berita asli maupun palsu. Dataset ini diperoleh dari berbagai sumber seperti Kaggle, McIntire, Reuters, dan BuzzFeed, serta dikenal sebagai salah satu dataset paling besar dan seimbang dalam kajian deteksi berita palsu karena proporsi antara berita asli dan palsu relatif merata. Kondisi tersebut menjadikannya dasar yang kuat untuk melatih dan mengevaluasi model bahasa Inggris.

Sementara itu, untuk bahasa Indonesia digunakan dataset Berita Hoax 2023 yang juga tersedia di Kaggle. Dataset ini terdiri dari artikel berita yang telah diberi label asli atau palsu dan mencerminkan karakter bahasa Indonesia, termasuk

pola hoaks yang umum ditemukan di lingkungan lokal. Kombinasi kedua dataset ini memungkinkan pelatihan dua model monolingual yang disesuaikan dengan karakteristik masing-masing bahasa, sehingga sistem dapat berfungsi secara efektif dalam mendeteksi berita palsu pada teks berbahasa Inggris maupun Indonesia.

Dataset pada masing-masing bahasa kemudian dibagi menggunakan rasio 80% untuk data latih dan 20% untuk data uji. Pembagian ini memastikan bahwa proses evaluasi dilakukan menggunakan data yang benar-benar baru dan tidak pernah digunakan dalam proses pelatihan.

Untuk memberikan gambaran mengenai karakter dataset, Tabel X berikut menyajikan contoh cuplikan berita asli dan palsu dari kedua bahasa.

Tabel II. Dataset Bahasa Inggris

Bahasa	Judul Berita	Cuplikan Teks	Label*
Inggris	**Specter of Trump Loosens Tongues in Silicon Valley – NYT**	"After years of scorning politics, Silicon Valley has leapt into the fray… tech leaders warn Trump's campaign promotes anger and bigotry."	**0 (Real)**
Inggris	**Tim Tebow Will Attempt Another Comeback, This Time in Baseball**	"Tim Tebow, a former NFL quarterback, prepares for a career in Major League Baseball… scouts impressed by his athleticism."	**0 (Real)**
Inggris	**Russian warships ready to strike terrorists near Aleppo**	"Russian aircraft prepare to strike terrorist positions near Aleppo… new missile systems expected to be deployed."	**1 (Fake)**
Inggris	**#NoDAPL: Native American Leaders Vow to Stay All Winter**	"Protesters vow to continue opposition to the Dakota Access Pipeline… police response criticized for excessive force."	**1 (Fake)**

Tabel III. Dataset Bahasa Indonesia

Bahasa	Judul Berita	Cuplikan Teks	Label*
Indonesia	**Gempa M5,7 Guncang Melonguane Sulut**	"BMKG melaporkan gempa Magnitudo 5,7 mengguncang Melonguane, Sulawesi Utara… tidak berpotensi tsunami."	**0 (Asli)**
Indonesia	**Perform a Romelu Lukaku Kian Membaik, Simone Inzaghi Beri Pujian**	"Simone Inzaghi memberi pujian kepada Romelu Lukaku yang performanya kian membaik… perlahan menemukan kondisi terbaiknya."	**0 (Asli)**
Indonesia	**Indonesia Ambil Alih Wilayah China**	"PERJUANGAN INDONESIA TAK SIA-SIA, SUKSES AMBIL ALIH TIGA WILAYAH CHINA…"	**1 (Palsu)**
Indonesia	**Ustadz Maulana Meninggal Tadi Sore Pukul 16:21 WIB**	"INNALILLAHI… Suasana rumah duka Ustadz Maulana disebut ramai, namun informasi ini ternyata tidak memiliki dasar resmi."	**1 (Palsu)**

Contoh ini menunjukkan perbedaan pola bahasa antara berita asli dan palsu pada kedua bahasa, seperti penggunaan judul sensasional, klaim tanpa sumber, atau gaya penulisan emosional yang umum ditemukan pada berita hoaks.

3.3 Data Preprocessing

Setiap dataset diproses melalui serangkaian tahapan praproses untuk memastikan teks berada dalam format yang konsisten sebelum digunakan pada pelatihan model. Tahapan pertama adalah tokenisasi, yaitu memecah teks menjadi unit kata menggunakan pustaka NLTK. Selanjutnya dilakukan penghapusan stopword dalam bahasa Inggris maupun Indonesia untuk mengurangi kata yang tidak memberikan informasi penting, seperti “the”, “and”, “di”, dan “dan”.

Proses normalisasi dilakukan dengan mengonversi seluruh teks menjadi huruf kecil serta menghapus karakter khusus, tanda baca, dan angka yang tidak relevan. Semua tahap ini diterapkan secara terpisah pada masing-masing dataset karena keduanya memiliki struktur linguistik yang berbeda.

Setelah praproses selesai, teks diubah menjadi representasi numerik menggunakan TF-IDF, yang memberikan bobot lebih tinggi pada kata-kata yang dianggap informatif dalam dokumen. Pendekatan ini terbukti lebih efektif dibanding bag-of-words dalam menangkap pola linguistik yang relevan untuk deteksi berita palsu.

Setiap dataset diproses secara terpisah sesuai bahasanya masing-masing, dimulai dari tahap praproses, ekstraksi fitur, hingga pelatihan model. Model bahasa Inggris dilatih menggunakan dataset WELFake, sedangkan model bahasa Indonesia dilatih menggunakan dataset Berita Hoax 2023. Pendekatan pelatihan terpisah ini memberikan hasil yang lebih stabil dan sesuai dengan karakteristik linguistik masing-masing bahasa.

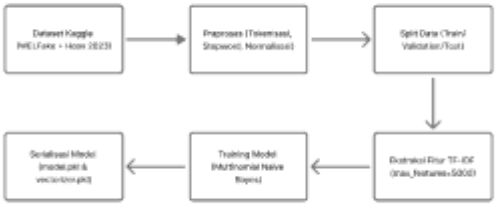
3.4 Pengembangan Model

Sistem deteksi berita palsu dua bahasa ini menggunakan algoritma Multinomial Naïve Bayes, yaitu sebuah klasifikator probabilistik yang sederhana namun efektif, terutama ketika digunakan pada data teks berdimensi tinggi. Naïve Bayes bekerja dengan mengasumsikan independensi antarfitur, sebuah penyederhanaan yang terbukti akurat pada banyak aplikasi pemrosesan bahasa alami.

Untuk merepresentasikan fitur teks, penelitian ini menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), yang mengubah kata menjadi bobot numerik berdasarkan frekuensi kemunculannya dalam suatu dokumen dibandingkan keseluruhan korpus. Jumlah fitur TF-IDF dibatasi hingga 5000 fitur untuk menjaga efisiensi komputasi, menghindari overfitting, serta memastikan model tetap fokus pada kata-kata yang paling informatif tanpa kehilangan konteks penting dari teks.

Implementasi gabungan TF-IDF dan Multinomial Naïve Bayes ini terbukti mampu membedakan berita palsu dan berita asli pada teks bahasa Inggris maupun Indonesia dengan tingkat akurasi yang baik.

Gambar 1 berikut menampilkan alur proses pelatihan model yang digunakan dalam penelitian ini. Diagram ini memberikan gambaran umum mengenai tahapan pembentukan model sebelum diterapkan pada sistem prediksi.



Gambar 1. Flowchart Pelatihan Model dan Pembuatan File .pkl

Gambar 1 menunjukkan alur pelatihan model mulai dari proses ekstraksi fitur TF-IDF, pelatihan menggunakan Multinomial Naïve Bayes, hingga penyimpanan model dan vectorizer ke dalam file .pkl. Flowchart ini merangkum tahapan utama dalam pembentukan model untuk kedua bahasa tanpa mengulangi proses praproses dan pembagian dataset yang telah dijelaskan pada subbab sebelumnya.

3.5 Pelatihan dan Pengujian Model

Proses pelatihan dilakukan secara terpisah untuk dua bahasa menggunakan dua dataset monolingual, yaitu WELFake untuk bahasa Inggris dan Berita Hoax 2023 untuk bahasa Indonesia. Setiap model dilatih untuk mengenali pola linguistik sesuai karakteristik bahasanya masing-masing.

Teks yang telah melalui tahap praproses kemudian dikonversi menjadi representasi numerik menggunakan TF-IDF. Representasi ini menjadi masukan bagi algoritma Multinomial Naïve Bayes, yang mempelajari pola distribusi kata antara kelas berita asli dan berita palsu. Pengaturan nilai α (alpha) serta pembatasan fitur TF-IDF hingga 5000 fitur diterapkan guna menjaga efisiensi komputasi dan meningkatkan kemampuan generalisasi model.

Tahap pengujian dilakukan menggunakan data uji yang telah dipisahkan dari dataset pelatihan. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai performa klasifikasi model. Selain itu, kurva ROC dan nilai AUC digunakan untuk menilai kestabilan model dalam membedakan dua kelas. Pada model bahasa Indonesia, nilai AUC yang tinggi mengindikasikan kecenderungan overfitting akibat ukuran dataset yang kecil dan variasi data yang terbatas dibandingkan dataset bahasa Inggris.

Gambar 2 berikut menyajikan alur lengkap proses pelatihan dan pengujian model yang digunakan dalam penelitian ini. Diagram tersebut merangkum tahapan

utama mulai dari ekstraksi fitur hingga evaluasi performa.

Gambar 2. Flowchart Pelatihan dan Pengujian Model

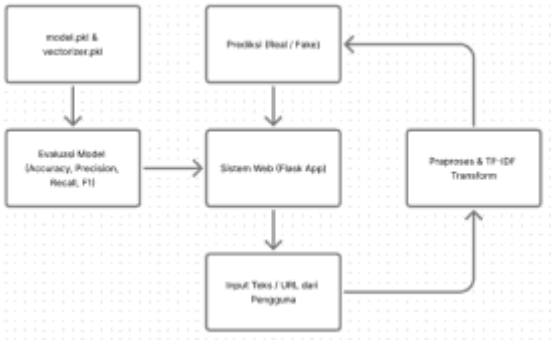
Gambar 2 menunjukkan tahapan pelatihan mulai dari ekstraksi fitur TF-IDF, pelatihan model, hingga proses evaluasi performa dan penyimpanan model dalam bentuk file .pkl.

3.6 Alur Sistem dan Antarmuka Pengguna (UI/UX)

Sistem web yang dikembangkan memiliki alur kerja mulai dari input pengguna hingga keluaran hasil prediksi. Pengguna memasukkan teks atau URL berita, lalu sistem mendeteksi bahasa, melakukan praproses teks, mengekstraksi fitur menggunakan TF-IDF, dan menghasilkan prediksi menggunakan model yang sudah dilatih. Alur lengkap sistem dapat dilihat pada Gambar 6.

Gambar 3. Flowchart Sistem Prediksi Berita Palsu

Antarmuka pengguna dirancang agar mudah



dipahami oleh pengguna berbahasa Indonesia maupun Inggris. Melalui halaman antarmuka, pengguna dapat memasukkan artikel berita atau URL untuk dianalisis, kemudian sistem akan menampilkan label prediksi beserta skor kepercayaan (confidence score). Desain halaman yang sederhana membantu pengguna dalam memahami hasil klasifikasi secara jelas dan informatif.

Gambar 4. Tampilan Antarmuka Prediksi Web

Antarmuka ini merupakan bagian akhir dari alur sistem dan berfungsi sebagai media interaksi utama bagi pengguna dalam melakukan prediksi berita palsu secara real time.

4. HASIL DAN PEMBAHASAN

4.1 Kinerja Model Bahasa Inggris

Kinerja model pada berita berbahasa Inggris dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil evaluasi ditampilkan pada tabel berikut.

Tabel IV. Kinerja Model Bahasa Inggris

Kelas	Precision	Recall	F1-score	Support
0 (Real News)	0.88	0.83	0.86	7081
1 (Fake News)	0.84	0.89	0.87	7227
Akurasi			**0.86**	**14308**
Macro Avg	0.86	0.86	0.86	14308
Weighted Avg	0.86	0.86	0.86	14308

Model mencapai akurasi sebesar **86%**, yang berarti 86% prediksi berhasil diklasifikasikan dengan benar. Namun, akurasi saja tidak cukup merepresentasikan performa model pada dataset yang mungkin memiliki ketidakseimbangan kelas. Oleh karena itu, presisi, recall, dan F1-score digunakan untuk memberikan gambaran yang lebih menyeluruh.

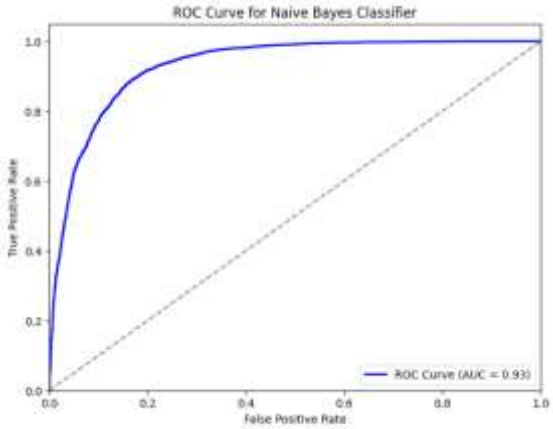
Presisi untuk kelas 0 (berita asli) berada pada angka 0.88, menunjukkan bahwa 88% prediksi "real news" adalah benar. Sementara itu, presisi untuk kelas 1 (berita palsu) sebesar 0.84. Recall kelas 0 adalah 0.83, artinya model berhasil mengenali 83% berita asli. Recall kelas 1 mencapai 0.89, menunjukkan model lebih baik dalam mendeteksi berita palsu.

F1-score kedua kelas relatif seimbang (0.86 dan 0.87), menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall. Secara keseluruhan, model mampu mengidentifikasi berita palsu maupun asli dengan performa yang stabil, dengan kinerja sedikit lebih unggul dalam mendeteksi berita palsu.

4.2 ROC Curve dan AUC untuk Model Bahasa Inggris

ROC Curve digunakan untuk mengevaluasi kemampuan model dalam membedakan berita asli dan palsu pada berbagai nilai ambang (*threshold*). Kurva ini menunjukkan hubungan antara *True Positive Rate* (Recall) dan *False Positive Rate*.

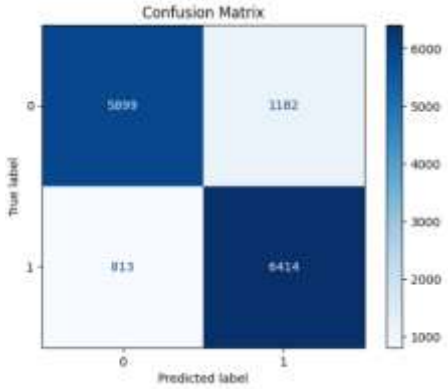
Nilai AUC sebesar 0.93 mengindikasikan bahwa terdapat probabilitas 93% bahwa sampel positif (berita palsu) akan diberi skor lebih tinggi daripada sampel negatif (berita asli). Nilai ini menunjukkan bahwa model memiliki performa sangat baik dalam membedakan kedua kelas.



Gambar 5. ROC Curve Model Bahasa Inggris

4.3 Confusion Matrix untuk Model Bahasa Inggris

Confusion matrix menunjukkan bahwa model memiliki jumlah *true positive* (TP) dan *true negative* (TN) yang tinggi, serta angka *false positive* (FP) dan *false negative* (FN) yang relatif rendah. Hal ini memperkuat bahwa model mampu mengklasifikasikan kedua kelas secara akurat.



Gambar 6. Confusion Matrix Model Bahasa Inggris

4.4 Kinerja Model Bahasa Indonesia

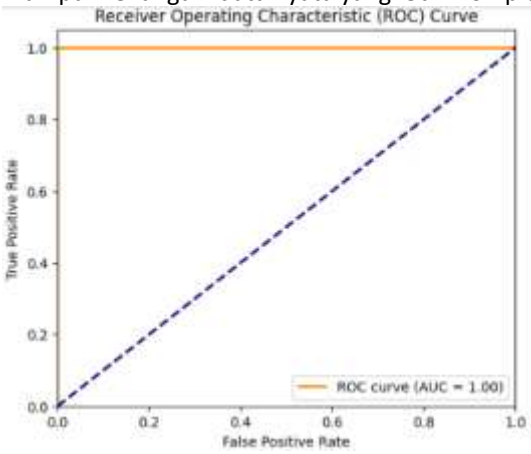
Kinerja model pada dataset berbahasa Indonesia dirangkum pada tabel berikut.

Tabel V. Kinerja Model Bahasa Indonesia

Kelas	Precision	Recall	F1-score	Support
0 (Real News)	0.89	0.96	0.92	305
1 (Fake News)	0.97	0.92	0.94	425
Akurasi			**0.93**	**730**
Macro Avg	0.93	0.94	0.93	730
Weighted Avg	0.94	0.93	0.93	730

4.5 ROC Curve dan AUC untuk Model Bahasa Indonesia

Model bahasa Indonesia menghasilkan nilai AUC mendekati 1.00, yang pada dasarnya mencerminkan performa "sempurna". Namun, hal ini juga menjadi indikasi bahwa terjadi overfitting, karena dataset Indonesia relatif kecil dan tidak beragam. Model cenderung "menghafal" pola data, sehingga hasilnya sangat baik pada data uji internal namun belum tentu mampu menangani data nyata yang lebih kompleks.



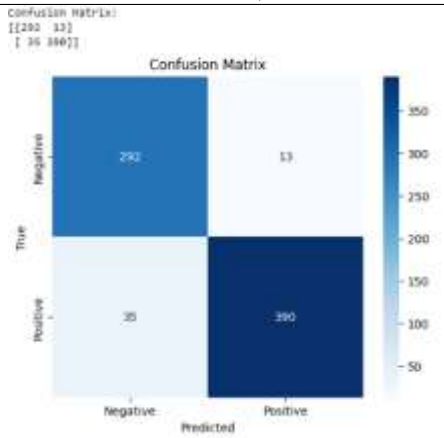
Gambar 7. ROC Curve Model Bahasa Indonesia

Hasil evaluasi menunjukkan bahwa model bahasa Indonesia memiliki kecenderungan overfitting. Hal ini terlihat dari nilai AUC yang sangat tinggi pada data uji internal, meskipun variasi data sebenarnya terbatas. Overfitting terjadi karena model terlalu "menghafal" pola-pola spesifik pada dataset, seperti kemunculan kata-kata sensasional, penggunaan huruf kapital berlebihan, frasa provokatif yang muncul berulang pada banyak sampel hoaks, serta pola kalimat pendek yang umum pada berita palsu lokal. Pola-pola ini sangat dominan di dataset Berita Hoax 2023, sehingga model mampu mengenalinya dengan mudah pada data uji internal, namun belum tentu mampu menggeneralisasi ke berita baru yang memiliki struktur atau gaya penulisan berbeda.

Nilai ROC dan AUC yang sangat tinggi mencerminkan kemampuan model dalam membedakan kelas pada dataset terbatas tersebut, namun tidak menjamin performa serupa pada data nyata yang lebih bervariasi.

4.6 Confusion Matrix untuk Model Bahasa Indonesia

Hasil confusion matrix menunjukkan jumlah TP dan TN yang tinggi, serta nilai FP dan FN yang rendah, sehingga mengonfirmasi kemampuan model dalam mengklasifikasikan berita asli dan berita palsu dengan tingkat kesalahan minimal.



Gambar 8. Confusion Matrix Model Bahasa Indonesia

5. KESIMPULAN DAN SARAN

UCAPAN TERIMA KASIH

Ucapan terima kasih dapat diberikan kepada Bapak Dosen saya Pak Ian serta dataset Welfake https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification dan Dataset Berita HOAX 2023 Indonesia https://www.kaggle.com/datasets/ainunnafiah/dataset-berita-hoax-2023

DAFTAR PUSTAKA

[1] S. Lewandowsky, U. K. H. Ecker, and J. Cook, "Beyond misinformation: Understanding and coping with the 'post-truth' era.," *J Appl Res Mem Cogn*, vol. 6, no. 4, pp. 353–369, Dec. 2017, doi: 10.1016/j.jarmac.2017.07.008.

[2] K. Shu, S. Wang, and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, Apr. 2018, pp. 430–435. doi: 10.1109/MIPR.2018.00092.

[3] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Fake News." [Online]. Available: http://wiki.dbpedia.org/about

[4] X. Zhou and R. Zafarani, "Fake News Detection: An Interdisciplinary Research," in *Companion Proceedings of The 2019 World Wide Web Conference*, New York, NY, USA: ACM, May 2019, pp. 1292–1292. doi: 10.1145/3308560.3316476.

[5] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science (1979)*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.

[6] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov, "Predicting Factuality of Reporting and Bias of News Media Sources," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, pp. 3528–3539. doi: 10.18653/v1/D18-1389.

[7] B. Kumar Sutradhar, M. Zonaid, N. Jahan Ria, S. Rashed Haider Noori, and A. Affiliations, "Machine Learning Technique Based Fake News Detection."

[8] A. A. Kurniawan and M. Mustikasari, "Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 544, Dec. 2021, doi: 10.32493/informatika.v5i4.6760.

[9] S. Jiang, Z. Guo, and J. Ouyang, "What makes sentiment signals work? Sentiment and stance multi-task learning for fake news detection," *Knowl Based Syst*, vol. 303, p. 112395, 2024, doi: https://doi.org/10.1016/j.knosys.2024.112395.

[10] J. Alghamdi, Y. Lin, and S. Luo, "Fake news detection in low-resource languages: A novel hybrid summarization approach," *Knowl Based Syst*, vol. 296, p. 111884, 2024, doi: https://doi.org/10.1016/j.knosys.2024.111884.

[11] A. Malik, D. K. Behera, J. Hota, and A. R. Swain, "Ensemble graph neural networks for fake news detection using user engagement and text features," *Results in Engineering*, vol. 24, p. 103081, 2024, doi: https://doi.org/10.1016/j.rineng.2024.103081.

[12] P. Bazmi, M. Asadpour, A. Shakery, and A. Maazallahi, "Entity-centric multi-domain transformer for improving generalization in fake news detection," *Inf Process Manag*, vol. 61, no. 5, p. 103807, 2024, doi: https://doi.org/10.1016/j.ipm.2024.103807.

[13] Y. Yang, J. Liu, Y. Yang, and L. Cen, "Dual-stream fusion network with multi-head self-attention for multi-modal fake news detection," *Appl Soft Comput*, vol. 167, p. 112358, 2024, doi: https://doi.org/10.1016/j.asoc.2024.112358.

[14] B. Wang, X. Li, C. Li, S. Wang, and W. Gao, "Escaping the neutralization effect of modality features fusion in multimodal Fake News Detection," *Information Fusion*, vol. 111, p. 102500, 2024, doi: https://doi.org/10.1016/j.inffus.2024.102500.

[15] F. Shan, M. Liu, M. Zhang, and Z. Wang, "Fake News Detection Based on Cross-Modal Message Aggregation and Gated Fusion Network," *Computers, Materials and Continua*, vol. 80, no. 1, pp. 1521–1542, 2024, doi: https://doi.org/10.32604/cmc.2024.053937.

[16] H. Chen *et al.*, "Multi-modal Robustness Fake News Detection with Cross-Modal and Propagation Network Contrastive Learning,"

Knowl Based Syst, p. 112800, 2024, doi: https://doi.org/10.1016/j.knosys.2024.112800.

[17] A. Pawlicka, M. Pawlicki, R. Kozik, A. Andrychowicz-Trojanowska, and M. Choraś, "AI vs linguistic-based human judgement: Bridging the gap in pursuit of truth for fake news detection," *Inf Sci (N Y)*, vol. 679, p. 121097, 2024, doi: https://doi.org/10.1016/j.ins.2024.121097.

[18] P. Dhiman, A. Kaur, D. Gupta, S. Juneja, A. Nauman, and G. Muhammad, "GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection," *Heliyon*, vol. 10, no. 16, p. e35865, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e35865.

[19] D. Wu, Z. Tan, H. Zhao, T. Jiang, and N. Geng, "Domain- and category-style clustering for general fake news detection via contrastive learning," *Inf Process Manag*, vol. 61, no. 4, p. 103725, 2024, doi: https://doi.org/10.1016/j.ipm.2024.103725.

[20] F. Yan, M. Zhang, B. Wei, K. Ren, and W. Jiang, "FMC: Multimodal fake news detection based on multi-granularity feature fusion and contrastive learning," *Alexandria Engineering Journal*, vol. 109, pp. 376–393, 2024, doi: https://doi.org/10.1016/j.aej.2024.08.103.

[21] L. Han, X. Zhang, Z. Zhou, and Y. Liu, "A Multifaceted Reasoning Network for Explainable Fake News Detection," *Inf Process Manag*, vol. 61, no. 6, p. 103822, 2024, doi: https://doi.org/10.1016/j.ipm.2024.103822.

[22] W. Yu *et al.*, "Research on Fake News Detection Based on Dual Evidence Perception," *Eng Appl Artif Intell*, vol. 133, p. 108271, 2024, doi: https://doi.org/10.1016/j.engappai.2024.108271.

[23] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "WELFake: Word Embedding over Linguistic Features for Fake News Detection," *IEEE Trans Comput Soc Syst*, vol. 8, no. 4, pp. 881–893, Aug. 2021, doi: 10.1109/TCSS.2021.3068519.