Implementasi *Fine-Tuning* IndoBERT untuk Deteksi Potensi Berita Hoaks Berbahasa Indonesia

http://dx.doi.org/10.28932/jutisi.v12i2.15203

Riwayat Artikel
Received: 28 Maret 2026 | *Final Revision*: 10 Agustus 2026 | *Accepted*: 13 Agustus 2026

Viny Christanti Mawardi✉#1, Henokh Mikhael Kristiantan#2

Program Studi Teknik Informatika, Universitas Tarumanagara
Jalan Letjen S. Parman No.1, Grogol, Jakarta Barat, Jakarta, 11440, Indonesia
[1]vinym@fti.untar.ac.id
[2]henokh.535220154@stu.untar.ac.id
✉Corresponding author: viny@fti.untar.ac.id

Abstrak **— Penyebaran informasi palsu atau hoaks di *platform* digital Indonesia telah menjadi ancaman serius bagi stabilitas sosial dan literasi masyarakat. Penelitian ini bertujuan untuk merancang dan mengimplementasikan aplikasi berbasis *web* yang mampu mendeteksi potensi berita hoaks berbahasa Indonesia secara otomatis. Metode yang diusulkan dalam penelitian ini menggunakan pendekatan *Natural Language Processing* dengan menerapkan teknik *fine-tuning* pada model *pre-trained* IndoBERT (*IndoBERT Base P1*). Model ini dilatih secara khusus menggunakan dataset teks berita berbahasa Indonesia untuk mengklasifikasi dan mengenali pola linguistik dari sebuah disinformasi. Temuan penting dari penelitian ini menunjukkan bahwa model IndoBERT yang di-*fine-tuning* berhasil diintegrasikan ke dalam arsitektur sistem *web* dan memberikan performa klasifikasi yang sangat baik dengan tingkat akurasi, presisi, dan *F1-score* mencapai 98,6%. Sistem juga mampu memproses teks masukan dan menampilkan persentase tingkat keyakinan atau *confidence score* secara *real-time* kepada pengguna. Kesimpulan dari penelitian ini adalah penerapan *fine-tuning* IndoBERT pada aplikasi *web* terbukti efektif dan dapat diandalkan sebagai alat bantu verifikasi dini. Kehadiran sistem ini diharapkan dapat memfasilitasi pengguna dalam menyaring informasi dan berkontribusi menekan laju penyebaran hoaks di ruang digital Indonesia.**

Kata kunci— **Aplikasi Web; Berita Hoaks; *Fine-Tuning;* IndoBERT; *Natural Language Processing.***

Implementing Fine-Tuned IndoBERT for Detecting Potential Indonesian Hoax News

Abstract ***— The spread of fake information or hoaxes on Indonesian digital platforms has become a serious threat to social stability and public literacy. This study aims to design and implement a web-based application capable of automatically detecting potential Indonesian hoax news. The proposed method utilizes a Natural Language Processing approach by applying fine-tuning techniques to the pre-trained IndoBERT model (IndoBERT Base P1). The model was specifically trained using an Indonesian news text dataset to classify and recognize the linguistic patterns of disinformation. Key findings demonstrate that the fine-tuned IndoBERT model was successfully integrated into the web system architecture, delivering excellent classification performance with accuracy, precision, and F1-score reaching 98.6%. Furthermore, the system is capable of processing input text and displaying the confidence score percentage in real-time to the user. In conclusion, the implementation of fine-tuned IndoBERT in a web application has proven to be an effective and reliable early verification tool. The presence of this system is expected to assist the public in filtering information and contribute to suppressing the spread of hoaxes within the Indonesian digital space.***

Keywords— ***Fine-Tuning; Hoax News; IndoBERT; Natural Language Processing; Web Application.***

I. PENDAHULUAN

Perkembangan teknologi informasi di era digital telah memicu pergeseran besar dalam distribusi informasi masyarakat. Kemudahan akses melalui platform media sosial dan portal berita daring sayangnya diiringi dengan lonjakan peredaran berita palsu (hoaks) yang dirancang untuk memanipulasi opini publik. Di Indonesia, kompleksitas pemrosesan bahasa alami atau *Natural Language Processing* (NLP) menjadi tantangan tersendiri dalam upaya otomatisasi verifikasi informasi. Pendekatan komputasional untuk klasifikasi teks berbahasa Indonesia saat ini telah bertransisi pesat menuju pemodelan *deep learning*. Beberapa penelitian mutakhir telah mengonfirmasi bahwa arsitektur berbasis *Transformer*, khususnya IndoBERT, memiliki kinerja yang jauh melampaui algoritma konvensional dalam mengekstraksi representasi semantik dan menangkap konteks linguistik yang secara spesifik mengindikasikan narasi disinformasi berbahasa Indonesia [1]. Selanjutnya, penggunaan berbagai algoritma *machine learning*, termasuk perbandingan dengan model berbasis *Transformer* seperti IndoBERT, untuk melakukan klasifikasi teks pada dataset Twitter di Indonesia [2]. Kedua penelitian tersebut membuktikan bahwa pendekatan komputasional NLP mampu mengekstraksi dan mengklasifikasikan pola teks berbahasa Indonesia secara efektif.

Pendekatan menggunakan arsitektur *Transformer* saat ini menjadi *state-of-the-art* dalam bidang NLP karena mekanisme *self-attention* yang mampu menangkap konteks kalimat secara mendalam [3]. IndoBERT merupakan implementasi spesifik dari arsitektur tersebut yang dibangun dan dilatih secara eksklusif menggunakan miliaran kata berbahasa Indonesia (Indo4B) [4]. Melalui paradigma transfer pembelajaran (*transfer learning*), model IndoBERT dapat diadaptasi untuk tugas klasifikasi teks spesifik melalui proses *fine-tuning*. Proses ini menyesuaikan bobot internal model menggunakan dataset berita hoaks dan valid, sehingga model mampu mengenali pola linguistik, gaya bahasa, dan struktur kalimat yang secara khas mengindikasikan sebuah disinformasi.

Meskipun pemodelan klasifikasi teks berbasis *Transformer* telah mapan dan terbukti sangat akurat, tinjauan terhadap literatur terkini menunjukkan bahwa mayoritas penelitian tersebut masih terhenti pada tahap evaluasi metrik performa di lingkungan eksperimental. Studi-studi terdahulu dominan berfokus pada penyetelan *hyperparameter* internal tanpa melangkah lebih jauh untuk mengeksplorasi tantangan teknis dalam memigrasikan model yang telah dilatih (*fine-tuned*) menuju lingkungan produksi (*production environment*). Hal ini memunculkan celah riset (*research gap*) yang nyata mengenai minimnya hilirisasi aplikasi NLP yang dapat diakses langsung oleh masyarakat, serta kurangnya aspek transparansi seperti penyajian *confidence score* (tingkat keyakinan model) pada antarmuka pengguna akhir.

Penelitian ini diusulkan untuk mengisi celah tersebut dengan menawarkan kontribusi ilmiah operasional. Kebaruan (*novelty*) penelitian ini tidak sekadar bertumpu pada pembuatan antarmuka aplikasi *web*, melainkan pada pembuktian kelayakan integrasi model *fine-tuning* IndoBERT di dalam ekosistem *client-server* yang sesungguhnya. Kontribusi spesifik difokuskan pada perancangan arsitektur yang mampu menerjemahkan bobot matematis model ke dalam sistem inferensi waktu nyata (*real-time inference*) yang interaktif dan transparan. Berdasarkan landasan tersebut, rumusan masalah dalam penelitian ini didefinisikan secara eksplisit: bagaimana mengintegrasikan model *fine-tuning* IndoBERT ke dalam sebuah arsitektur aplikasi *web* agar dapat memproses dan mendeteksi potensi berita hoaks berbahasa Indonesia secara presisi, sekaligus menyajikan tingkat keyakinan prediksinya secara transparan? Sejalan dengan rumusan masalah tersebut, penelitian ini bertujuan merancang, mengimplementasikan, dan memvalidasi keandalan aplikasi *web* pendeteksi hoaks sebagai instrumen bantu verifikasi informasi publik.

II. METODE

Metode pemrosesan bahasa alami (*Natural Language Processing*) yang digunakan pada sistem ini bertumpu pada teknik *fine-tuning* model IndoBERT. Pendekatan ini mengadaptasi model bahasa *pre-trained* berbasis arsitektur *Transformer* yang telah dilatih secara masif menggunakan korpus teks bahasa Indonesia berskala besar (Indo4B) [3], [4] Ini merupakan salah satu bentuk pembelajaran transfer (*transfer learning*) di mana model yang sudah memiliki pemahaman dasar bahasa disesuaikan kembali bobot parameternya untuk tugas hilir (*downstream task*) yang spesifik, dalam hal ini klasifikasi teks berita hoaks [5]. Berkat mekanisme *self-attention* yang bersifat dua arah (*bidirectional*), turunan arsitektur BERT ini mampu menangkap konteks dan relasi semantik antar kata secara mendalam, sehingga sangat efektif untuk mengidentifikasi pola linguistik dan kejanggalan dari sebuah disinformasi [4],[5].

Terdapat dua komponen utama dalam arsitektur sistem ini, yaitu komponen pemodelan dan komponen implementasi. Komponen pemodelan berfungsi untuk melatih lapisan klasifikasi model menggunakan dataset teks berita berlabel, sedangkan komponen implementasi bertugas mengintegrasikan model tersebut ke dalam kerangka kerja berbasis *web*. Integrasi ini memungkinkan model bahasa beroperasi secara interaktif sebagai antarmuka sistem (API) untuk memproses teks masukan pengguna dan menyajikan hasil klasifikasi beserta persentase tingkat keyakinan (*confidence score*) secara *real-time*.

Secara keseluruhan, pelaksanaan penelitian ini dibagi menjadi beberapa fase operasional yang berkesinambungan dan terukur. Proses diawali dengan tahap pengumpulan korpus teks berita yang diiringi dengan validasi kualitas data secara heuristik. Langkah ini sangat esensial untuk memastikan integritas struktural berita hasil *web scraping* sekaligus mereduksi potensi bias narasi dari sumber tertentu. Selain itu, guna mengatasi masalah ketidakseimbangan representasi kelas (*data imbalance*), penarikan sampel dilakukan melalui mekanisme *undersampling* secara purposif hingga tercapai proporsi

distribusi *dataset* yang benar-benar ekuivalen antara kelas hoaks dan kelas fakta. Fase selanjutnya adalah pemodelan, di mana arsitektur IndoBERT dikalibrasi secara ekstensif. Kalibrasi ini tidak sekadar bertumpu pada pengaturan *epoch*, melainkan mencakup pendefinisian parameter optimasi AdamW, *linear learning rate scheduler*, penetapan *max sequence length* untuk mengakomodasi teks berita yang panjang, serta penerapan *dropout* dan *early stopping* guna memitigasi *overfitting* secara ketat. Terakhir, guna menjamin ketangguhan generalisasi model dan mengatasi kelemahan pengujian metrik tunggal, skema evaluasi diperkuat dengan pendekatan validasi silang (*cross-validation*). Rangkaian metodologi yang komprehensif ini diakhiri dengan penyematan (*deployment*) model optimal ke dalam lingkungan aplikasi *web* dan pengujian fungsionalitas fungsional..

A. Analisis Kebutuhan Sistem

Untuk membuat sistem aplikasi web pendeteksi potensi hoaks berbasis *fine-tuning* IndoBERT berjalan dengan baik, perencanaan kebutuhan sistem harus dilakukan secara komprehensif. Ini akan memastikan bahwa proses pengembangan dan pelaksanaan sistem, mulai dari pelatihan model hingga implementasi antarmuka, berjalan dengan lancar. Sistem ini menggabungkan berbagai komponen perangkat keras dan perangkat lunak untuk memastikan kinerja inferensi yang responsif dan akurasi prediksi yang tinggi terhadap teks masukan pengguna.

Tabel 1 menggambarkan berbagai perangkat lunak yang digunakan untuk membangun sistem pendeteksi hoaks berbasis *web* beserta fungsi masing-masing. Untuk memastikan sistem beroperasi secara optimal, setiap perangkat lunak memiliki peran krusial, mulai dari tahap prapemrosesan data, pelatihan model bahasa, hingga penyediaan antarmuka pengguna yang interaktif.

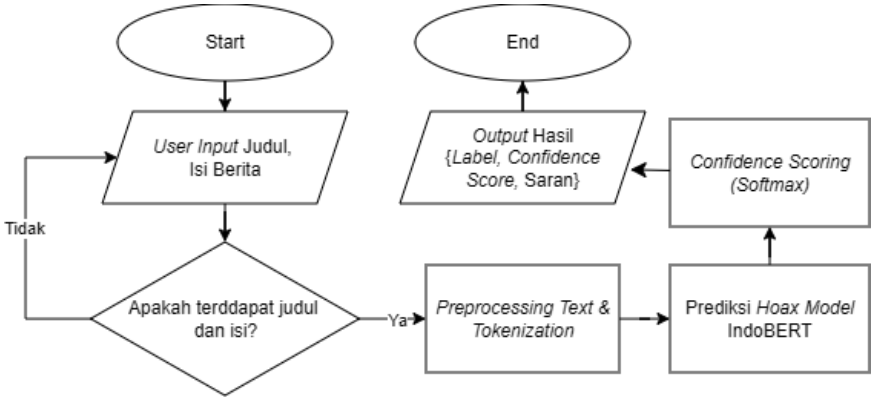
TABEL 1
KEBUTUHAN PERANGKAT LUNAK

No	Nama Perangkat Lunak	Fungsi/Kegunaan
1	Google Colab	Untuk penulisan logika sistem secara interaktif, prapemrosesan data teks, dan eksperimen pelatihan (*fine-tuning*) model awal.
2	Visual Studio Code	Lingkungan pengembangan utama (*IDE*) yang digunakan untuk menulis, mengedit, dan menyatukan kode *backend* dan *frontend* aplikasi *web*.
3	Python	Bahasa pemrograman utama yang digunakan untuk melatih model kecerdasan buatan dan membangun logika server.
4	Flask	*Framework* berbasis Python yang berfungsi membangun arsitektur *backend* dan menyediakan *Application Programming Interface* (API) untuk model klasifikasi.
5	Hugging Face Transformers	*Library* komputasional yang menyediakan arsitektur dasar dan utilitas untuk memuat serta memproses inferensi model berbasis *Transformer*.
6	IndoBERT Base P1	Model bahasa *pre-trained* utama yang dilatih ulang (*fine-tuning*) untuk memahami konteks kalimat dan mengklasifikasikan disinformasi.
7	HTML, CSS, JavaScript	Bahasa *markup* dan pemrograman yang digunakan untuk merancang dan membangun antarmuka pengguna (*frontend*) secara interaktif.

Proses pengembangan dan pengujian sistem ini menggunakan perangkat keras dengan spesifikasi tertentu guna mengakomodasi kebutuhan komputasi dari arsitektur *deep learning*. Konfigurasi perangkat keras yang digunakan mencakup prosesor utama AMD Ryzen 5 5500u, unit pemrosesan grafis terintegrasi berupa APU AMD Radeon(TM) Graphics dengan VRAM sebesar 512 MB, RAM berkapasitas 16 GB DDR4, serta dijalankan di atas lingkungan sistem operasi Windows 11. Dukungan spesifikasi memori dan prosesor tersebut dipilih secara khusus untuk memastikan bahwa seluruh rangkaian operasional, mulai dari pelatihan model IndoBERT, pengolahan data teks, hingga pengujian waktu respons (*response time*) dari aplikasi, dapat berjalan secara optimal tanpa mengalami hambatan komputasi (*bottleneck*).

Penelitian ini menggunakan pendekatan pemodelan visual untuk menjabarkan alur kerja sistem perangkat lunak secara sekuensial. Hal ini direpresentasikan melalui diagram alur rancangan proses aplikasi yang memetakan bagaimana data bergerak dari pengguna hingga menghasilkan prediksi akhir. Gambar 1 menunjukkan bahwa proses diawali ketika pengguna mengakses antarmuka aplikasi *web* dan memasukkan teks berita ke dalam sistem. Data masukan tersebut kemudian dikirim ke server *backend* untuk melalui tahap prapemrosesan agar format teks bersih dan sesuai dengan standar *tokenizer* model. Selanjutnya, teks yang telah diproses masuk ke tahap inferensi di mana model IndoBERT melakukan klasifikasi. Hasil probabilitas mentah dari model kemudian dikalkulasi menjadi persentase tingkat keyakinan (*confidence score*). Pada tahap

akhir, sistem mengembalikan data tersebut ke antarmuka pengguna untuk menampilkan label hasil deteksi, yakni indikasi potensi hoaks atau berita valid, beserta skor persentasenya secara visual.



Gambar 1.Diagram Alur Rancangan Proses Aplikasi

B. Pengumpulan & Prapemrosesan Data

Tahap awal penelitian melibatkan pengumpulan *dataset* sekunder berupa teks berita berbahasa Indonesia yang telah dianotasi dengan label klasifikasi biner, yaitu kelas hoaks (disinformasi) dan kelas fakta (berita valid). Guna memastikan model dapat membedakan pola bahasa secara objektif dan mencegah terjadinya bias pembelajaran akibat ketidakseimbangan kelas (*class imbalance*), data dikumpulkan sedemikian rupa hingga mencapai proporsi yang seimbang [6]. Total *dataset* gabungan yang berhasil dikonstruksi berjumlah 4.357 baris data. Secara spesifik, data representatif untuk kelas hoaks dihimpun dari portal repositori pemeriksa fakta TurnBackHoax.id serta *Indonesian News Dataset* yang tersedia di platform Kaggle. Sementara itu, untuk merepresentasikan kelas berita fakta, data dikumpulkan secara mandiri menggunakan teknik *web scraping* dari portal berita arus utama nasional, yakni CNN Indonesia.

Setelah data mentah terkumpul, tahapan krusial selanjutnya adalah *data preprocessing* untuk menstandardisasi format teks [7]. Tahapan ini mencakup serangkaian proses berurutan, dimulai dari *case folding* untuk menyeragamkan seluruh karakter menjadi huruf kecil, hingga *text cleaning*. Pembersihan ini secara spesifik menargetkan penghapusan tautan URL, angka, simbol, tanda baca yang tidak relevan, serta sisa elemen *markup* HTML yang mungkin terbawa selama proses *scraping*. Proses reduksi *noise* ini sangat esensial agar teks masukan memiliki struktur yang bersih dan terstandar, mengingat kualitas prapemrosesan sangat memengaruhi performa akhir klasifikasi teks [8]. Pada tahap akhir prapemrosesan, seluruh korpus yang telah bersih dipartisi (*data splitting*) menjadi tiga bagian, yaitu 70% dialokasikan sebagai data latih (*training set*), 10% sebagai data validasi (*validation set*), dan 20% sebagai data uji (*testing set*) guna memastikan model dapat dievaluasi secara objektif pada data yang belum pernah dilihat sebelumnya [9]. Rangkaian persiapan ini memastikan data berada dalam kondisi optimal sebelum dikonversi menjadi representasi angka oleh *tokenizer* bawaan model IndoBERT [4].

C. Datasets

Data yang diekstraksi untuk tahap pelatihan model secara mendasar memiliki dua atribut utama, yaitu teks narasi berita dan label klasifikasi biner. Guna memberikan gambaran yang komprehensif mengenai struktur dan variasi pola bahasa yang dipelajari oleh model, karakteristik sampel data dari masing-masing kelas perlu diuraikan. Pada kelas berita fakta, teks didominasi oleh susunan kalimat yang mematuhi kaidah jurnalistik baku, bersifat informatif, netral, dan terstruktur [10]. Sebagai contoh representatif dari sumber portal berita resmi, sebuah narasi valid memiliki pola kalimat seperti: "Kementerian kesehatan resmi mengumumkan program revitalisasi fasilitas kesehatan tingkat pertama yang akan dieksekusi secara bertahap mulai kuartal ketiga tahun ini." Teks dengan struktur yang lugas dan objektif tersebut secara konsisten dianotasi sebagai kelas fakta.

Sebaliknya, pada kelas disinformasi atau hoaks yang dihimpun dari repositori pemeriksa fakta, narasi sering kali dirancang untuk memanipulasi psikologis pembaca melalui penggunaan tanda baca berlebihan, penulisan huruf kapital yang tidak sesuai ejaan, serta diksi provokatif [11]. Salah satu contoh sampel data yang merepresentasikan kelas hoaks memiliki pola seperti: "TOLONG SEBARKAN!!! Telah ditemukan virus jenis baru yang sengaja disebarkan melalui udara di pusat perbelanjaan, jangan keluar rumah jika tidak ingin menjadi korban!!!" Teks dengan pola linguistik manipulatif dan agresif semacam ini dianotasi secara tegas sebagai kelas hoaks. Perbandingan dari kedua sampel tersebut mengilustrasikan perbedaan fitur semantik yang sangat kontras, yang selanjutnya akan diekstraksi dan dipelajari oleh arsitektur *deep learning* selama fase *fine-tuning* berlangsung.

Adapun struktur akhir dari *dataset* yang telah melewati seluruh tahapan *preprocessing* direpresentasikan ke dalam format tabular dengan dua kolom utama, yakni kolom teks yang memuat gabungan antara judul dan isi berita, serta kolom label klasifikasi. Sistem pelabelan pada target klasifikasi ini dikodekan secara biner, di mana representasi nilai 0 ditetapkan untuk kelas berita fakta atau valid, sedangkan nilai 1 secara spesifik ditetapkan untuk kelas berita hoaks. Guna memberikan ilustrasi visual yang lebih komprehensif mengenai kerangka data akhir yang telah siap diumpankan ke dalam fase pelatihan model, cuplikan struktur *dataset* tersebut disajikan secara rinci pada Gambar 2 berikut.

	text	label
0	putin tunjuk prabowo jadi ketua ndb brics. fak...	1
1	koalisi masyarakat sipil tuntut adili jokowi d...	1
2	kpk mau dibekukan prabowo kalau gagal kejar ko...	1
3	kpk tetapkan donna harun jadi tersangka kasus ...	1
4	mahasiswa ugm akan berangkat ke jakarta minta ...	1
...	...	...
4352	teks kultum tarawih di bulan ramadhan dengan t...	0
4353	gubernur sutarmidji ancam pedagang jangan coba...	0
4354	gudang garam setor modal rp3 triliun lagi demi...	0
4355	rupiah turun ke rp15092 per dolar as pagi ini....	0
4356	sikapi rencana opec harga minyak brent dan wti...	0

4357 rows × 2 columns

Gambar 2. Bentuk Dataset Hasil Preprocessing

D. Fine-Tuning Model IndoBERT

Pendekatan inti pada penelitian ini adalah penerapan teknik *fine-tuning* menggunakan arsitektur *pre-trained* IndoBERT varian indobert-base-p1 [4]. Pada dasarnya, IndoBERT merupakan model representasi bahasa yang menghasilkan vektor berdimensi tinggi. Proses *fine-tuning* dilakukan dengan melatih kembali bobot parameter dengan cara menambahkan lapisan klasifikasi teks (*sequence classification head*) di atas lapisan keluaran model menggunakan *dataset* yang telah dibersihkan [12].

Secara matematis, untuk setiap teks berita masukan, model IndoBERT akan mengekstraksi token representasi agregat (*CLS token*) yang dinotasikan sebagai h.Vektor representasi h ini kemudian diumpankan ke dalam lapisan klasifikasi linier (*fully connected layer*) yang memiliki matriks bobot W dan bias b. Distribusi probabilitas P untuk kelas target y (fakta atau hoaks) dihitung menggunakan fungsi aktivasi *Softmax* dengan persamaan berikut:

$$P(y|h) = softmax\,(Wh + b) \quad (1)$$

pada persamaan tersebut, variabel $P(y|h)$ merepresentasikan probabilitas bersyarat kelas target y yang dihasilkan berdasarkan representasi masukan h, W merepresentasikan matriks bobot (*weight matrix*) dari lapisan klasifikasi linier, h adalah vektor representasi fitur yang diekstraksi oleh IndoBERT, dan b merupakan nilai bias pada lapisan tersebut.

Sesuai dengan rujukan literatur pembelajaran mendalam , ekspresi matematis dari fungsi *softmax* itu sendiri untuk mengonversi nilai logit z di mana $z = (Wh + b)$ menjadi probabilitas pada kelas ke- i dirumuskan sebagai:

$$softmax\,(z_i) = \frac{exp(z_i)}{\sum_{j=1}^{C} exp(z_i)} \quad (2)$$

pada persamaan *softmax* tersebut, variabel z_i merupakan nilai logit sebelum normalisasi untuk kelas ke-i, fungsi exp merupakan fungsi eksponensial untuk memastikan seluruh keluaran bernilai positif, dan C adalah jumlah total kelas target yang berfungsi sebagai batas iterasi penyebut untuk menormalisasi total distribusi probabilitas menjadi satu.

Selama proses pelatihan berjalan pada setiap *epoch*, model memanfaatkan algoritma optimasi untuk memperbarui seluruh parameter bobot, baik pada lapisan arsitektur dasar IndoBERT maupun pada lapisan klasifikasi yang baru ditambahkan.

Tujuan utama dari optimasi ini adalah untuk meminimalkan tingkat kesalahan prediksi (*loss*) menggunakan fungsi objektif *Cross-Entropy Loss* ($\mathcal{L}$) yang dirumuskan sebagai:

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log(p_c) \qquad (3)$$

pada persamaan tersebut, variabel $\mathcal{L}$ merepresentasikan nilai fungsi kerugian lintas entropi yang harus diminimalkan, C merupakan jumlah total kelas klasifikasi (dalam penelitian ini $C = 2$), y_c merepresentasikan label kebenaran aktual (*ground truth*), dan p_c adalah nilai probabilitas hasil prediksi model untuk kelas c. Melalui mekanisme komputasi dan penyesuaian bobot matematis ini, model yang awalnya hanya memiliki pemahaman semantik bahasa Indonesia secara umum, dioptimalkan secara spesifik untuk mengenali fitur linguistik yang mengindikasikan narasi berita hoaks [5].

E. Evaluasi Kinerja Model

Untuk memastikan keandalan dan daya generalisasi model yang telah dilatih, tahap evaluasi kinerja dilakukan secara komprehensif menggunakan metrik kuantitatif terhadap subset data uji (*test set*) yang belum pernah dilihat oleh model sebelumnya. Dasar dari evaluasi ini bertumpu pada perhitungan *Confusion Matrix*, yang memetakan hasil prediksi model ke dalam empat kuadran utama: *True Positive* (TP) atau berita hoaks yang benar diprediksi sebagai hoaks, *True Negative* (TN) atau berita valid yang benar diprediksi sebagai valid, *False Positive* (FP) atau berita valid yang keliru diprediksi sebagai hoaks, dan *False Negative* (FN) atau berita hoaks yang gagal terdeteksi dan diklasifikasikan sebagai valid [13].

Berdasarkan keempat nilai dasar tersebut, kinerja komputasional model diekstraksi ke dalam empat metrik pengukuran utama. Rumusan matematis untuk masing-masing metrik diekspresikan melalui persamaan berikut:

Tingkat akurasi (*Accuracy*) mengukur rasio total prediksi yang benar secara keseluruhan terhadap total sampel data uji:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (4)$$

Tingkat presisi (*Precision*) mengukur ketepatan prediksi kelas positif, yakni rasio berita hoaks yang diprediksi benar dibandingkan dengan total seluruh teks yang dilabeli hoaks oleh model:

$$Precision = \frac{TP}{TP+FP} \qquad (5)$$

Tingkat sensitivitas (*Recall*) mengukur kemampuan spesifik model dalam mengenali seluruh instans kelas positif, yakni rasio berita hoaks yang berhasil dideteksi dari keseluruhan berita hoaks aktual yang ada di dalam data uji:

$$Recall = \frac{TP}{TP+FN} \qquad (6)$$

Sementara itu, *F1-Score* merupakan metrik yang merepresentasikan nilai rata-rata harmonik (*harmonic mean*) antara *precision* dan *recall*

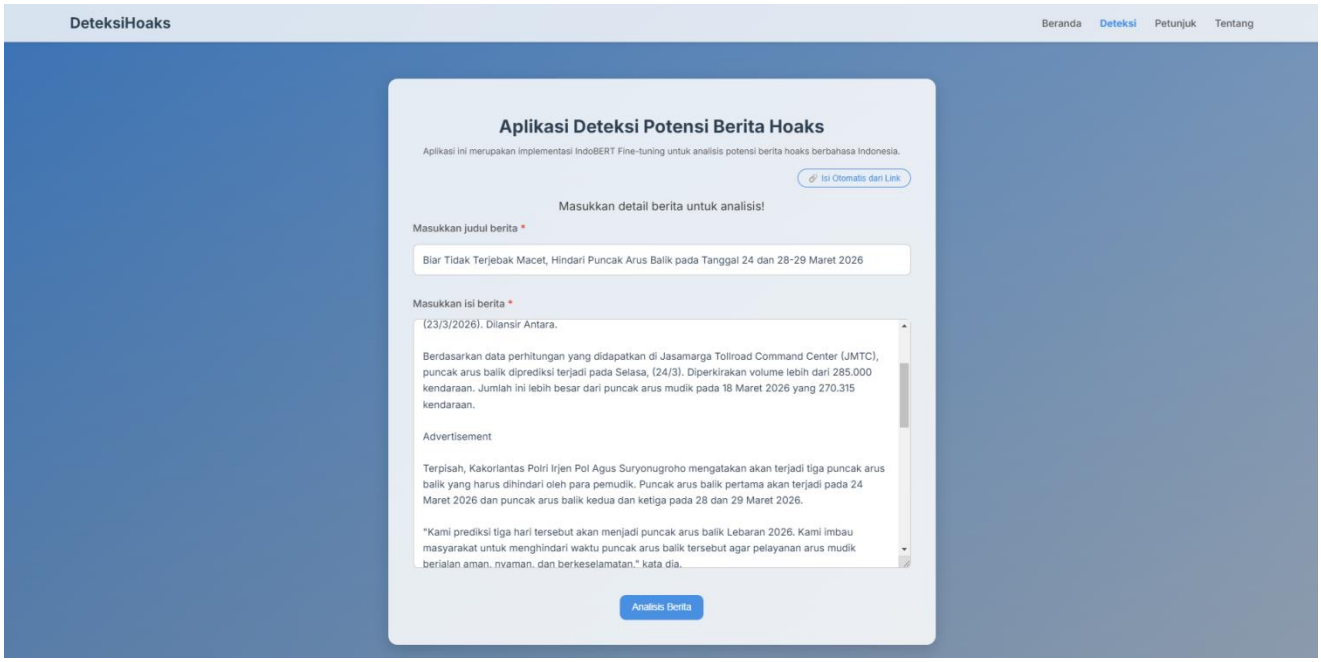
$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (7)$$

Penggunaan kombinasi metrik evaluasi matematis ini bertujuan untuk memberikan penilaian yang komprehensif secara multidimensi. Pendekatan ini tidak sekadar melihat persentase akurasi tebakan yang benar, tetapi juga mengukur ketangguhan sistem dalam meminimalkan kesalahan prediksi yang berisiko. Dalam konteks sistem pendeteksi disinformasi, metrik *recall* yang tinggi sangat krusial untuk memastikan tidak banyak berita hoaks yang lolos (*false negative*), sementara *precision* yang stabil menjaga agar portal berita fakta tidak secara sembarangan teranotasi sebagai hoaks (*false positive*). Oleh karena itu, *F1-Score* dijadikan tolok ukur utama untuk memvalidasi bahwa model klasifikasi memiliki performa yang seimbang dan tidak bias terhadap salah satu kelas prediksi.

F. Perancangan dan Implementasi Sistem

Model klasifikasi teks dengan performa evaluasi terbaik kemudian diekspor untuk diimplementasikan ke dalam lingkungan perangkat lunak berbasis web. Sistem ini dikembangkan menggunakan kerangka kerja arsitektur *client-server* [14]. Sisi *backend* dikonfigurasi menggunakan bahasa pemrograman Python untuk memuat model *machine learning* dan memproses inferensi teks. Sementara itu, sisi *frontend* dirancang untuk menyediakan antarmuka pengguna yang intuitif guna memfasilitasi input teks artikel dari pengguna.

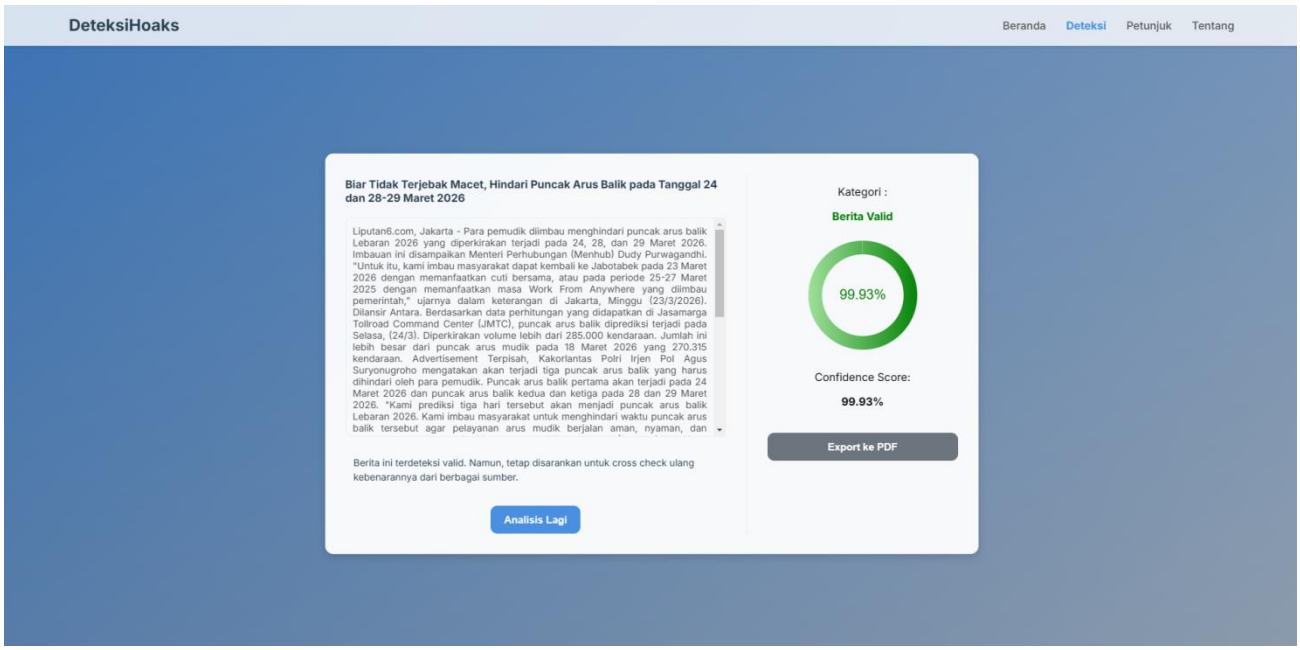
Rancangan halaman input deteksi berita aplikasi dapat dilihat pada Gambar 3. Pada halaman ini, pengguna disediakan sebuah area teks khusus untuk memasukkan atau menempelkan konten narasi berita yang ingin diverifikasi kebenarannya yaitu input judul dan isi berita. Terdapat juga fitur otomatis cek berita melalui link. Namun apabila pengguna ingin menggunakan fitur langsung dari link pada perancangan ini terbatas hanya bisa portal berita tertentu. Setelah pengguna mengeksekusi perintah deteksi, sistem akan mengirimkan data masukan tersebut ke server untuk diproses.



Gambar 3.Rancangan Antarmuka Aplikasi Web Halaman Input Berita

Hasil pemrosesan inferensi model kemudian dikembalikan dan ditampilkan secara *real-time* kepada pengguna akhir, seperti yang diilustrasikan rancangannya pada Gambar 4. Pada antarmuka hasil evaluasi ini, sistem menyajikan label klasifikasi akhir secara visual, yakni indikasi potensi hoaks atau berita fakta yang valid. Selain itu, antarmuka ini juga menampilkan persentase tingkat keyakinan (*confidence score*) dari prediksi tersebut. Penyajian tingkat keyakinan ini merupakan elemen transparansi yang krusial untuk membantu pengguna dalam menginterpretasikan seberapa yakin model kecerdasan buatan terhadap hasil deteksi yang diberikan.

p-ISSN : 2443-2210
e-ISSN : 2443-2229



Gambar 4.Rancangan Antarmuka Hasil Deteksi Berita

III. HASIL DAN PEMBAHASAN

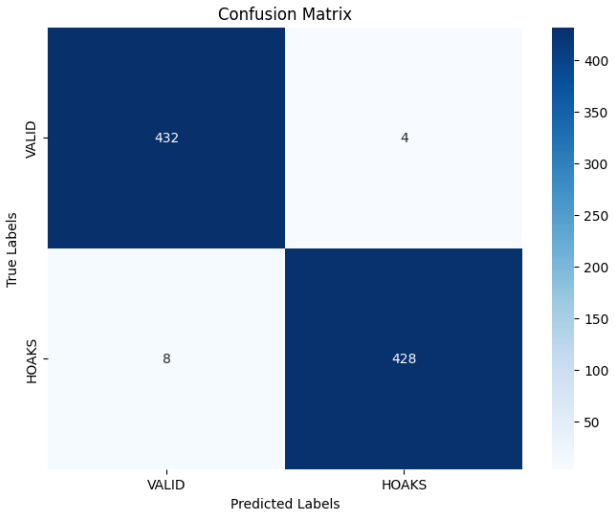
A. Hasil Eksperimen dan Pelatihan Model

Tahap pemodelan menggunakan *dataset* teks berita berbahasa Indonesia yang berjumlah 4.357 baris data, dengan proporsi seimbang antara kelas hoaks dan valid (fakta). *Dataset* ini bersumber dari portal TurnBackHoax.id, Indonesian News Dataset (Kaggle), dan hasil *web scraping* dari portal berita CNN Indonesia. Data tersebut dibagi menjadi 70% data latih (*training*), 10% data validasi (*validation*), dan 20% data uji (*testing*). Pemilihan proporsi ini didasarkan pada prinsip keseimbangan empiris dalam pembelajaran mesin, di mana alokasi 70% menyediakan volume sampel yang memadai bagi arsitektur model untuk mengekstraksi dan mempelajari fitur semantik secara komprehensif tanpa mengalami kekurangan informasi (*underfitting*). Selanjutnya, porsi 10% untuk data validasi difungsikan secara spesifik sebagai tolok ukur evaluasi internal selama fase iterasi pelatihan berlangsung, yang mana sangat esensial untuk penyesuaian *hyperparameter* dan pencegahan memori hafalan (*overfitting*). Sementara itu, alokasi sebesar 20% untuk data uji dinilai cukup representatif sebagai himpunan data tak terlihat (*unseen data*) guna menghasilkan pengukuran metrik kinerja generalisasi akhir yang objektif dan valid secara statistik.

Proses *fine-tuning* menggunakan model bahasa *pre-trained* indobenchmark/indobert-base-p1 dilakukan dengan mengeksplorasi tiga skenario *hyperparameter* utama untuk menemukan konfigurasi yang paling optimal. Berdasarkan hasil eksperimen, Skenario 3 yang menggunakan *learning rate* sebesar 2 x 10^{-5} dan 3 *epoch* terbukti memberikan performa terbaik dan paling stabil. Meskipun pencapaian akurasi pada *dataset* berskala relatif kecil ini memunculkan kekhawatiran teoretis terkait potensi memori hafalan (*overfitting*), indikasi tersebut telah berhasil dimitigasi. Nilai *validation loss* yang terus menurun dan mencapai titik terendah 0,0275 pada iterasi akhir tanpa adanya lonjakan divergensi membuktikan bahwa arsitektur model benar-benar mengekstraksi pola generalisasi bahasa, bukan sekadar menghafal distribusi data latih. Model dari Skenario 3 inilah yang kemudian diekspor untuk diimplementasikan ke dalam sistem.

B. Evaluasi Kinerja Model

Kinerja model final dievaluasi secara objektif menggunakan metrik klasifikasi standar terhadap 872 data uji (*test set*) yang belum pernah dilihat oleh model selama proses pelatihan. Hasil evaluasi ini direpresentasikan secara visual melalui *Confusion Matrix* pada Gambar 5, dan rincian nilai pengukuran performanya dijabarkan pada Tabel 2 untuk mengukur tingkat keandalan prediksi.



Gambar 5.Confusion Matrix Hasil Evaluasi pada Data Uji

TABEL 2
HASIL METRIK EVALUASI MODEL (CLASSIFICATION REPORT)

	Precision	Recall	F1-Score	Support
VALID	0.98	0.99	0.99	436
HOAKS	0.99	0.98	0.99	436
Accuracy			0.99	872
Macro Avg	0.99	0.99	0.99	872
Weighted Avg	0.99	0.99	0.99	872

Merujuk pada data di Tabel 2, hasil pengujian pada data uji menunjukkan bahwa model IndoBERT yang telah melalui proses *fine-tuning* berhasil mencapai tingkat akurasi (*accuracy*) keseluruhan sebesar 98,62% (dibulatkan menjadi 0,99 pada tabel metrik). Lebih lanjut, model mencetak nilai rata-rata presisi (*precision*), *recall*, dan *f1-score* yang sangat tinggi, yakni mencapai 0,99 untuk kedua kelas, baik untuk kelas hoaks maupun valid.

Pencapaian kinerja yang sangat optimal ini secara komputasional didorong oleh beberapa faktor utama. Pertama, arsitektur dasar IndoBERT telah memiliki prapelatihan pada korpus bahasa Indonesia berskala raksasa, sehingga representasi semantik dasar telah terbangun dengan kokoh sebelum tahap *fine-tuning*. Kedua, terdapat perbedaan fitur linguistik permukaan yang sangat kontras antara kelas hoaks dan kelas fakta di dalam *dataset*. Narasi hoaks cenderung memiliki pola manipulatif yang repetitif seperti penggunaan tanda baca berlebihan dan kosakata provokatif, yang mana dengan mudah dipetakan oleh mekanisme atensi (*attention mechanism*) pada arsitektur *Transformer*. Terakhir, proporsi distribusi antarkelas yang seimbang pada fase pelatihan mencegah terjadinya bias mayoritas.

Sementara itu, berdasarkan analisis *Confusion Matrix* yang ditunjukkan pada Gambar 5, dari total 872 data uji, model berhasil memprediksi secara tepat 432 berita valid (*true negative*) dan 428 berita hoaks (*true positive*). Tingkat kesalahan prediksi yang dihasilkan sangat minim, di mana hanya terdapat 4 berita valid yang keliru diklasifikasikan sebagai hoaks (*false positive*) dan 8 berita hoaks yang gagal dideteksi sehingga diklasifikasikan sebagai valid (*false negative*). Keseimbangan rasio pada metrik evaluasi ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang sangat baik, mampu menangkap pola linguistik disinformasi, dan tidak bias terhadap salah satu kelas prediksi.

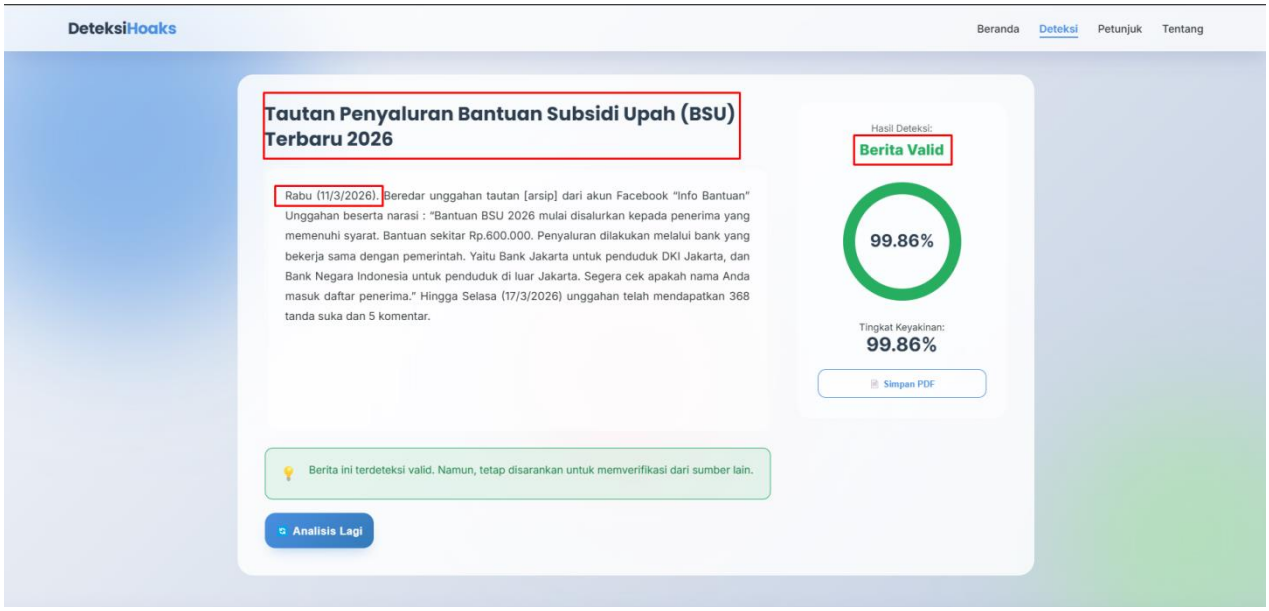
Tingkat kesalahan prediksi yang dihasilkan sangat minim, di mana hanya terdapat 4 berita valid yang keliru diklasifikasikan sebagai hoaks (*false positive*) dan 8 berita hoaks yang gagal dideteksi sehingga diklasifikasikan sebagai valid (*false negative*). Keseimbangan rasio pada metrik evaluasi ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang sangat baik, mampu menangkap pola linguistik disinformasi, dan tidak bias terhadap salah satu kelas prediksi.

Sebagai bentuk analisis kualitatif terhadap keandalan model, hasil pengujian menunjukkan bahwa sistem sangat unggul dalam membedakan struktur bahasa. Pada kasus-kasus yang berhasil diprediksi dengan sempurna, model terbukti kehebatannya dalam mengenali berita valid yang menggunakan struktur bahasa jurnalistik baku, netral, dan objektif. Di sisi lain, model juga sangat tanggap dalam mendeteksi berita hoaks yang memuat narasi provokatif, penggunaan tanda baca berlebihan, huruf kapital yang tidak sesuai ejaan, serta kosakata emosional yang sering dimanfaatkan dalam kampanye

disinformasi. Keberhasilan ini menegaskan bahwa *fine-tuning* pada arsitektur IndoBERT sangat efektif dalam memetakan pola semantik teks bahasa Indonesia.

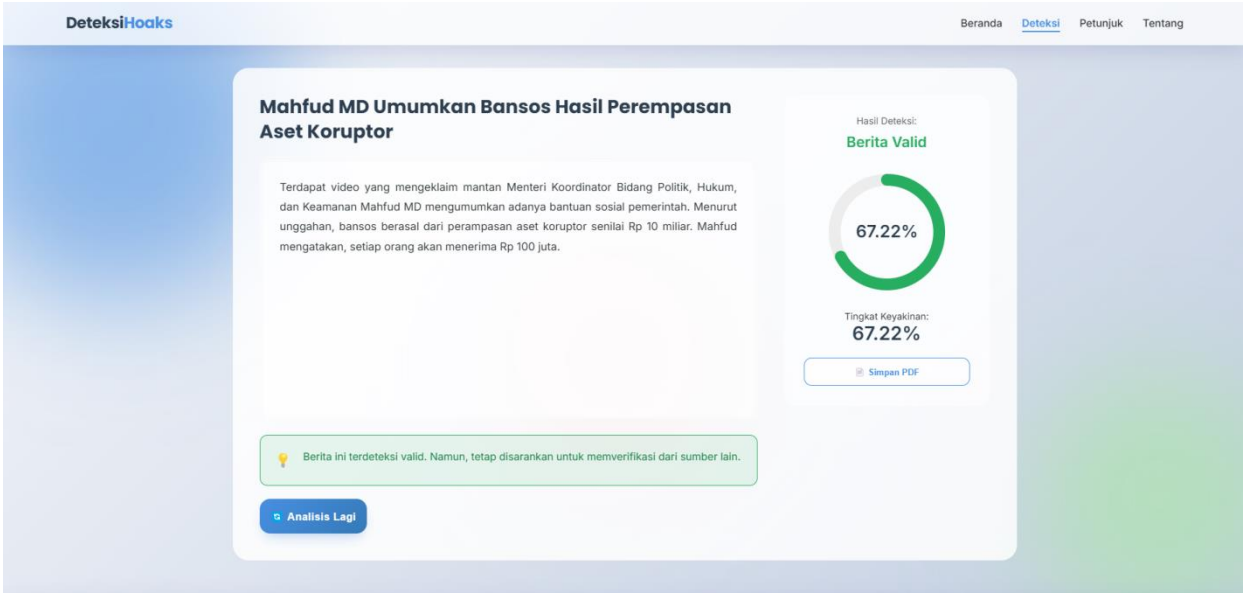
Namun demikian, evaluasi juga menyoroti beberapa kasus kegagalan prediksi yang menjadi batasan dari model saat ini. Secara eksplisit, batasan utama sistem ini terletak pada ketidakmampuannya dalam melakukan verifikasi silang terhadap kebenaran fakta objektif di dunia nyata, mengingat model murni beroperasi berdasarkan analisis semantik dan pengenalan pola linguistik permukaan (*surface linguistic features*). Berdasarkan analisis terhadap sampel data yang salah diklasifikasikan, berita hoaks yang lolos terdeteksi dan dianggap sebagai fakta (*false negative*) umumnya terjadi pada teks disinformasi yang direkayasa dengan gaya bahasa sangat formal.

Secara komputasional, fenomena kegagalan ini terjadi karena penulis hoaks meniru struktur tata bahasa jurnalistik resmi secara identik dan menghindari penggunaan kosakata provokatif. Akibatnya, mekanisme atensi (*attention mechanism*) pada model IndoBERT tidak mengidentifikasi adanya anomali semantik yang umumnya menjadi pemicu probabilitas kelas hoaks. Lebih lanjut, terdapat temuan analitis krusial di mana model memiliki kecenderungan kuat untuk mengklasifikasikan berita hoaks bernarasi panjang sebagai berita valid. Penyebab fundamental dari anomali prediksi ini berakar pada fenomena korelasi palsu (*spurious correlation*) akibat bias distribusi metrik teks pada *dataset* pelatihan. Mengingat mayoritas sampel data pada kelas hoaks didominasi oleh teks berukuran singkat bergaya pesan media sosial, sementara kelas fakta dibangun dari ekstraksi artikel portal berita arus utama yang bernarasi panjang, model mengembangkan konklusi matematis yang keliru. Arsitektur jaringan saraf tiruan secara tidak sengaja mengasosiasikan variabel panjang narasi dan kerapian sintaksis sebagai indikator mutlak keabsahan informasi. Manifestasi nyata dari batasan model dan bias distribusi ini dapat ditinjau pada Gambar 6, yang mengilustrasikan bagaimana sebuah teks disinformasi berhasil mengecoh sistem dan terdeteksi sebagai fakta semata-mata karena keberhasilannya mereplikasi pola struktural berita yang valid.



Gambar 6.Contoh Hasil Deteksi Yang Salah

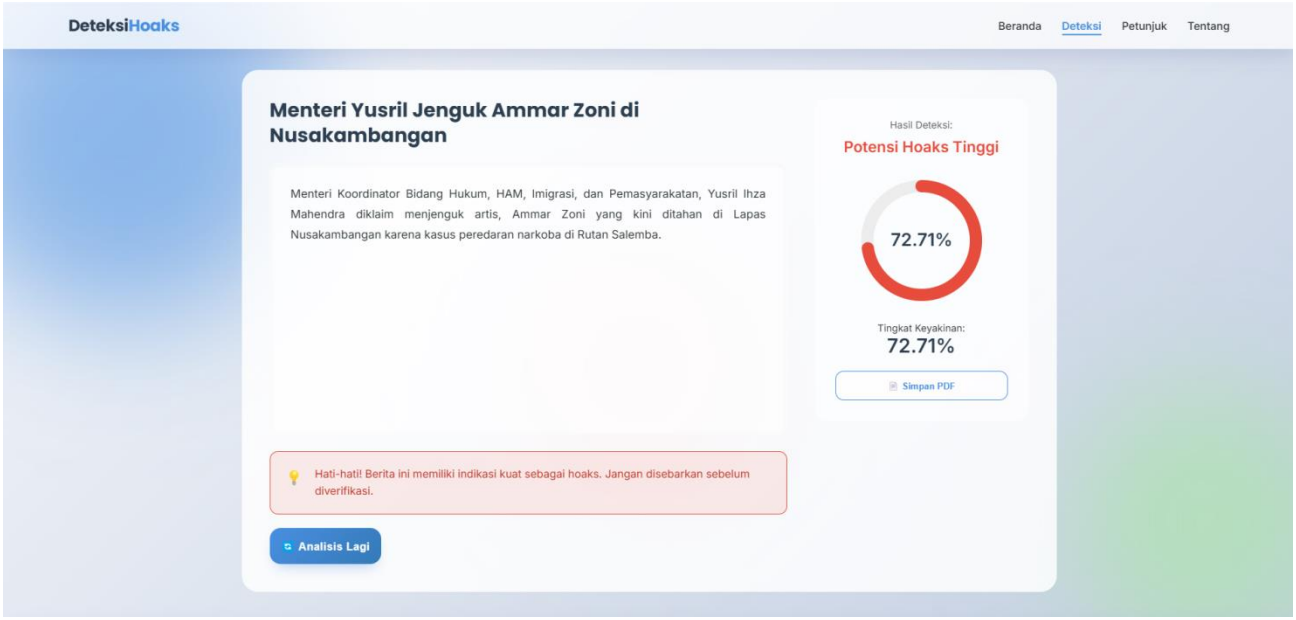
Berita di atas seharusnya merupakan berita hoaks, namun diklasifikasikan sebagai berita valid oleh model. Kesalahan prediksi ini terjadi karena teks masukan tersebut disusun menggunakan struktur kalimat yang sangat rapi, memiliki panjang narasi yang memadai, serta menggunakan gaya bahasa formal yang menyerupai tata penulisan artikel jurnalistik profesional. Selain itu, narasi tersebut sama sekali tidak mengandung kosakata provokatif, penggunaan huruf kapital yang tidak wajar, maupun tanda baca berlebihan yang umumnya menjadi pola karakteristik utama pada *dataset* disinformasi. Fenomena serupa juga diilustrasikan pada Gambar 7, di mana sebuah berita hoaks kembali dideteksi secara keliru sebagai berita valid. Kesalahan klasifikasi ini diakibatkan oleh tidak adanya kosakata provokasi dalam teks tersebut. Namun demikian, model tetap menunjukkan keraguannya yang ditandai dengan persentase tingkat keyakinan (*confidence score*) yang tergolong rendah kemungkinan besar karena teks berita tersebut pendek dan tidak sesuai dengan *dataset* berita valid pada *training* model.



Gambar 7.Contoh Berita Hoaks Terdeteksi Valid Dengan Confidence Score Rendah

Karena model IndoBERT pada penelitian ini murni mengandalkan analisis semantik dan pola linguistik permukaan (*surface linguistic features*) tanpa memiliki kemampuan untuk melakukan verifikasi silang terhadap fakta kejadian di dunia nyata, model mengenali struktur formal tersebut sebagai penanda kebenaran informasi.. Hal ini membuktikan bahwa peniru gaya bahasa resmi dapat mengecoh model klasifikasi teks yang berdiri sendiri tanpa integrasi pengecekan sumber eksternal.

Sebaliknya, sebagai contoh yang diilustrasikan pada Gambar 8, terdapat sebuah teks berita hoaks yang telah berhasil dideteksi dengan benar sebagai kelas hoaks oleh sistem. Namun demikian, model menghasilkan persentase tingkat keyakinan (*confidence score*) yang relatif rendah. Penurunan nilai probabilitas ini disinyalir terjadi karena pada teks narasi tersebut tidak ditemukan unsur kosakata provokatif, gaya bahasa emosional, maupun tanda baca berlebihan yang umumnya menjadi ciri khas utama sebuah disinformasi. Hal ini mengindikasikan bahwa meskipun sistem mampu mengidentifikasi kepalsuan informasi secara akurat, ketiadaan elemen linguistik yang agresif tersebut membuat model merespons dengan tingkat kehati-hatian atau ambiguitas yang lebih tinggi.



Gambar 8.Contoh Hasil Prediksi Hoaks dengan Confidence Rendah

p-ISSN : 2443-2210
e-ISSN : 2443-2229

C. Pembahasan dan Implikasi Teoritis

Pencapaian metrik evaluasi yang menginjak angka 98,62% pada penelitian ini perlu dikritisi secara komprehensif guna memvalidasi keunggulannya. Sebagai perbandingan komputasional, berbagai studi terdahulu yang mengandalkan algoritma *baseline* pembelajaran mesin tradisional seperti *Naive Bayes* atau *Support Vector Machine* untuk klasifikasi teks berbahasa Indonesia umumnya menemui titik jenuh pada rentang akurasi 75% hingga 85%. Kinerja IndoBERT yang nyaris sempurna ini menegaskan superioritas mekanisme pemrosesan berbasis *Transformer* dalam memetakan relasi semantik kompleks, sebuah konklusi yang sejalan dengan tren keberhasilan deteksi disinformasi pada konteks bahasa dengan sumber daya terbatas (*low-resource language*) lainnya.

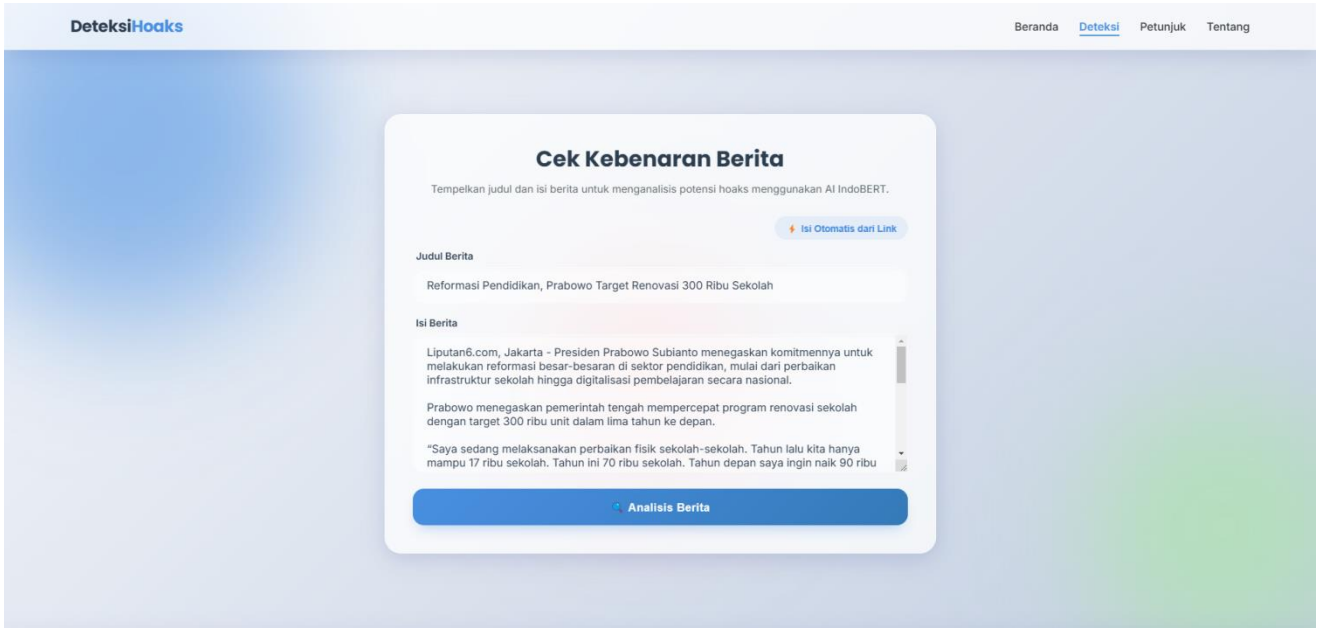
Namun demikian, tingginya angka akurasi ini berbanding lurus dengan tantangan teknis pada ranah pemrosesan bahasa alami yang lebih luas. Model yang dibangun masih beroperasi layaknya kotak hitam (*black box*) dengan keterbatasan dalam aspek kemampuan penjelasan logis (*explainability*). Temuan kualitatif di mana model dapat tertipu oleh berita hoaks yang mereplikasi gaya bahasa jurnalistik formal mengindikasikan bahwa ketangguhan (*robustness*) sistem masih rentan terhadap serangan manipulasi teks (*adversarial text*). Model saat ini sangat bergantung pada fitur linguistik permukaan dan belum memiliki kapabilitas penalaran fakta eksternal.

Lebih lanjut, transisi dari model eksperimental menuju sistem deteksi hoaks yang dapat diakses publik membawa implikasi etis yang tidak dapat diabaikan. Kelemahan arsitektural model memunculkan potensi bias sosial dan risiko tuduhan palsu (*false accusation*). Apabila sistem secara keliru melabeli karya jurnalistik faktual dari media independen sebagai sebuah hoaks (*false positive*), hal tersebut berpotensi mencederai kebebasan pers dan mendistorsi persepsi publik. Oleh karena itu, aplikasi cerdas ini harus diposisikan secara etis sebagai instrumen verifikasi awal (*screening tool*) yang mendampingi literasi kritis pengguna, bukan sebagai entitas penentu kebenaran yang bersifat mutlak.

D. Implementasi dan Pengujian Sistem Berbasis Web

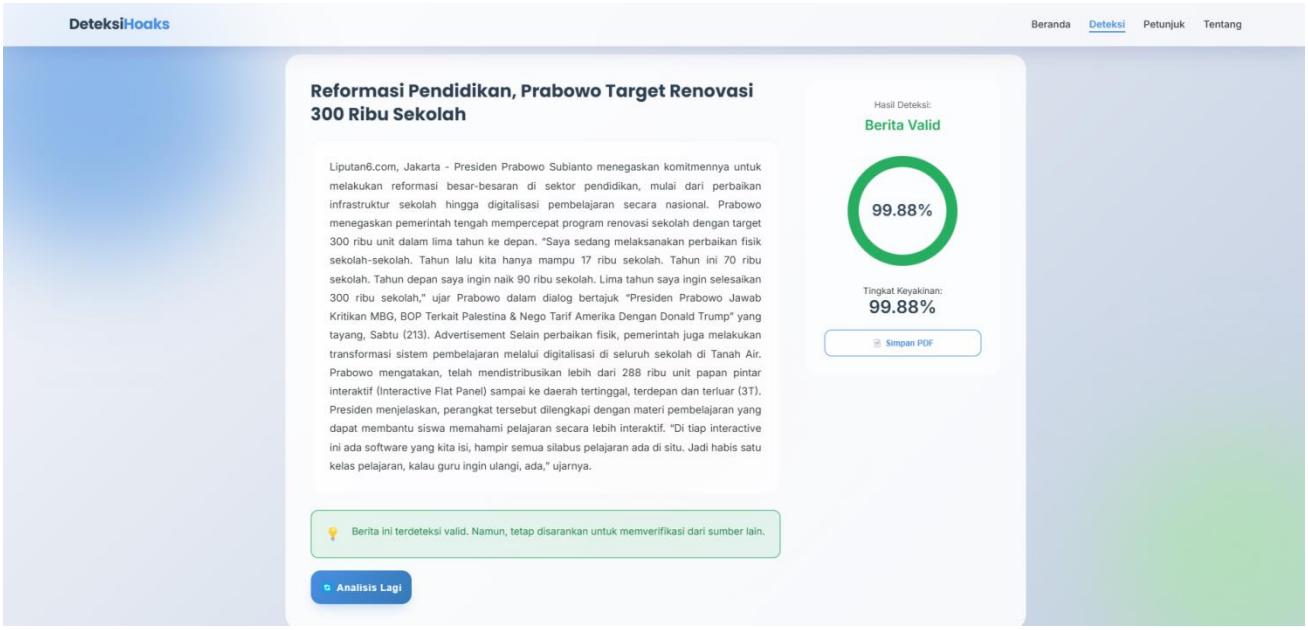
Model klasifikasi teks yang telah dievaluasi kemudian diintegrasikan ke dalam arsitektur sistem *client-server* menggunakan kerangka kerja Flask berbasis Python [14]. Sisi *backend* bertugas memuat model *machine learning* dan memproses inferensi teks masukan dari pengguna. Keluaran skor mentah (*logits*) dari lapisan klasifikasi IndoBERT kemudian diolah menggunakan fungsi aktivasi *Softmax* untuk menghasilkan distribusi probabilitas, di mana probabilitas tertingginya direpresentasikan sebagai persentase tingkat keyakinan (*confidence score*) [15]. Sementara itu, sisi *frontend* dirancang untuk menyediakan antarmuka pengguna yang intuitif guna memfasilitasi interaksi langsung dengan sistem.

Sebagai wujud hasil implementasi dari perancangan sebelumnya, antarmuka halaman utama aplikasi secara fungsional dapat dilihat pada Gambar 9. Pada halaman ini, pengguna disediakan area teks khusus untuk memasukkan judul dan isi narasi berita yang ingin diverifikasi kebenarannya. Setelah pengguna mengeksekusi perintah deteksi, sistem akan memproses data tersebut secara *real-time*.



Gambar 9.Antarmuka Aplikasi Web Halaman Input Berita

Selanjutnya, antarmuka hasil evaluasi, Gambar 10 menyajikan label klasifikasi akhir secara visual, yakni indikasi potensi hoaks atau berita fakta. Antarmuka ini juga secara eksplisit menampilkan persentase tingkat keyakinan (*confidence score*). Penyajian metrik keyakinan ini merupakan elemen transparansi yang sangat krusial untuk membantu pengguna dalam menginterpretasikan seberapa yakin model kecerdasan buatan terhadap prediksi yang diberikan.



Gambar 10.Antarmuka Hasil Deteksi Berita

Untuk memastikan keandalan aplikasi, pengujian fungsionalitas sistem dilakukan melalui metode *black box testing* guna memvalidasi kesesuaian antara fungsi masukan dan keluaran perangkat lunak [16]. Hasil pengujian mengonfirmasi bahwa seluruh skenario validasi, seperti penolakan masukan data kosong, batasan jumlah karakter, dan perlindungan terhadap masukan dominan numerik, telah berfungsi sesuai dengan spesifikasi yang dirancang. Selain itu, *user acceptance test* (UAT) dilakukan untuk mengukur tingkat kelayakan dan penerimaan sistem oleh pengguna akhir [17]. Dari aspek performa, sistem mencatat waktu respons rata-rata di bawah dua detik untuk menyelesaikan satu siklus prediksi dari awal hingga akhir. Hasil pengujian ini secara komprehensif membuktikan bahwa aplikasi pendeteksi hoaks yang dibangun bersifat fungsional, responsif, transparan, dan siap diimplementasikan sebagai alat bantu verifikasi dini oleh pengguna.

IV. SIMPULAN

Penelitian ini telah berhasil menjawab rumusan masalah dengan mengimplementasikan model bahasa *fine-tuning* IndoBERT ke dalam ekosistem arsitektur *client-server* guna menghasilkan aplikasi *web* pendeteksi potensi hoaks berbahasa Indonesia yang interaktif dan transparan. Sintesis kontribusi ilmiah utama dari penelitian ini terletak pada pembuktian kelayakan operasional, di mana model kecerdasan buatan berbasis *Transformer* terbukti tidak hanya unggul pada metrik evaluasi eksperimental, tetapi juga tangguh diaplikasikan secara waktu nyata (*real-time*) untuk memberikan hasil prediksi beserta transparansi tingkat keyakinan (*confidence score*) kepada pengguna akhir. Kinerja komputasional yang dihasilkan sangat andal, direpresentasikan oleh pencapaian tingkat akurasi sebesar 98,62% serta nilai rata-rata presisi, *recall*, dan *f1-score* yang stabil pada angka 0,99.

Dari perspektif strategis pengembangan pemrosesan bahasa alami atau *Natural Language Processing* (NLP) berbahasa Indonesia, penelitian ini memposisikan diri sebagai jembatan krusial antara riset teoretis dan hilirisasi perangkat lunak berskala produksi. Secara praktis, arsitektur sistem ini menawarkan implikasi nyata sebagai fondasi teknologi yang menjanjikan bagi pengembangan sistem deteksi hoaks berskala nasional. Platform ini dapat difungsikan oleh masyarakat luas maupun lembaga pemeriksa fakta independen sebagai instrumen penyaring dini (*screening tool*) guna menekan laju polarisasi disinformasi digital. Meskipun model sangat tanggap membedakan pola bahasa baku dan provokatif, analisis kualitatif mengungkap batasan strategis di mana sistem masih murni berpatokan pada pengenalan pola linguistik permukaan tanpa kemampuan verifikasi fakta dunia nyata. Oleh karena itu, arah strategis untuk penelitian lanjutan harus difokuskan pada penggunaan *dataset* pelatihan yang jauh lebih spesifik dan terstruktur. Elaborasi keragaman data ini harus mencakup

penyeimbangan distribusi rasio panjang dokumen teks guna mengeliminasi bias korelasi struktural, serta penambahan korpus data disinformasi berkalimat formal, konten pseudo-jurnalistik, hingga ragam anomali pada domain topik khusus seperti politik, kesehatan, dan finansial. Selain itu, integrasi mekanisme pengecekan fakta silang dari pangkalan data eksternal juga secara kuat direkomendasikan guna mereduksi kelemahan model dalam mengklasifikasikan disinformasi yang dikemas layaknya berita resmi maupun berita fakta bergaya bahasa umpan klik (*clickbait*).

UCAPAN TERIMA KASIH

Terima kasih yang sebesar-besarnya kepada Fakultas Teknologi Informasi Universitas Tarumanagara atas dukungan finansial dan moral yang diberikan selama berjalannya penelitian ini. Apresiasi setinggi-tingginya juga ditujukan kepada seluruh rekan atas sinergi, diskusi mendalam, serta kontribusi pemikiran yang berharga di setiap tahapan penelitian ini. Kolaborasi dan dukungan dari seluruh pihak terkait sangatlah krusial dalam menyukseskan penelitian ini.

DAFTAR PUSTAKA

[1] K. A. Nugraha, "Analisis Sentimen Berbasis Emoticon pada Komentar Instagram Bahasa Indonesia Menggunakan Naïve Bayes," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 3, 2021.
[2] M. I. Amal, E. S. Rahmasita, E. Suryaputra, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, 2022.
[3] A. Vaswani *et al.*, "Attention Is All You Need," *31st Conf. Neural Inf. Process. Syst. (NIPS 2017)*, no. Nips, 2017, [Online]. Available: https://arxiv.org/abs/1706.03762
[4] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP." [Online]. Available: https://arxiv.org/abs/2011.00677
[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Association for Computational Linguistics, 2019, pp. 4171–4186.
[6] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, pp. 1263–1284, 2009.
[7] D. Jurafsky and J. H. Martin, "Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models Third Edition draft," 2026. [Online]. Available: http://academia.edu/44148361/Speech_and_Language_Processing_An_Introduction_to_Natural_Language_Processing_Computational_Linguistics_and_Speech_Recognition_Third_Edition_draft
[8] A. K. Uysal and S. Gunal, "The impact of preprocessing on text classification," *Inf. Process. \& Manag.*, vol. 50, pp. 104–112, 2014.
[9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009.
[10] H. Rashkin, E. Choi, Y. Jin, and H. A. Schwartz, "Textual Characteristics of Legitimate News versus Disinformation," *Proc. Conf. Empir. Methods Nat.*, pp. 2931–2937, 2017.
[11] V. Perez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Fake News," *Proc. Int. Conf. Comput.*, pp. 3391–3401, 2018.
[12] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, 2020, pp. 38–45.
[13] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. \& Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
[14] M. Grinberg, *Flask Web Development: Developing Web Applications with Python*, 2nd ed. O'Reilly Media, 2018.
[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[16] W. Grosky and T. Ruas, *Data Science for Software Engineers*, 9th ed. New York, NY, USA: McGraw-Hill, 2019.
[17] I. Sommerville, *Software Engineering*, vol. 10th. Pearson, 2015.