

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Fenomena hoaks telah menjadi permasalahan kompleks yang melintasi batas geografis dan bidang sosial. Berbagai penelitian internasional mengungkapkan bahwa penyebaran informasi palsu tidak hanya terbatas pada satu wilayah, melainkan mencakup spektrum global yang luas [1]. Jejak hoaks terlihat jelas dalam berbagai konteks kritis, seperti proses demokrasi dan krisis kesehatan global. Dalam konteks politik, misalnya, pemilihan presiden Amerika Serikat pada 2019 dan 2024 menunjukkan bagaimana hoaks dapat secara signifikan memengaruhi naratif publik dan dinamika elektoral [2], [3]. Demikian pula, pandemi COVID-19 menjadi panggung utama penyebaran misinformasi yang masif, di mana informasi palsu tentang virus dan vaksin menyebar dengan kecepatan yang mengkhawatirkan [4].

Di Indonesia, fenomena hoaks memiliki dimensi khusus yang patut mendapat perhatian serius. Dengan jumlah pengguna internet mencapai 272 juta pada tahun 2023—menempatkan Indonesia di peringkat keempat dunia—risiko penyebaran informasi palsu menjadi semakin tinggi [5]. Survei Katadata Insights Center [6] menunjukkan bahwa 60% responden pernah terpapar berita palsu, dan 11% bahkan pernah secara tidak sengaja menyebarkannya. Menurut Kementerian Komunikasi dan Informatika, terjadi peningkatan eksponensial dalam jumlah hoaks yang terdeteksi, terutama menjelang Pemilu 2024, dengan peningkatan hampir sepuluh kali lipat dibandingkan tahun sebelumnya [7]. Fenomena ini tidak sekadar ancaman informasional, melainkan potensi gangguan sistemik terhadap stabilitas sosial dan integritas proses demokrasi.

Oleh karena itu, diperlukannya pendekatan deteksi berita hoaks ini secara otomatis. Dan untungnya, pendekatan ini telah mengalami transformasi signifikan dalam beberapa dekade terakhir. Pada awalnya, algoritma machine learning klasik seperti Support Vector Machine (SVM), Naive Bayes, dan Decision Trees [8] digunakan untuk melakukan deteksi hoaks. Metode ini bergantung pada teknik ekstraksi fitur seperti *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *CountVectorizer* untuk mengubah teks menjadi representasi numerik yang dapat diproses oleh model. Meskipun metode ini menunjukkan tingkat keberhasilan yang

cukup baik, pendekatan ini memiliki keterbatasan dalam hal kemampuan untuk menangkap makna kontekstual yang lebih dalam dari sebuah teks [9]. Beberapa penelitian lainnya juga telah mencoba efektivitas algoritma machine learning klasik ini. Tajrian et al. melakukan perbandingan berbagai algoritma machine learning untuk klasifikasi berita hoaks dan menemukan bahwa Neural Network mengungguli SVM, Naive Bayes, dan Decision Trees dalam hal akurasi dan skor F_1 [10]. Hal ini menunjukkan adanya keterbatasan algoritma tradisional dalam mendeteksi pola linguistik kompleks yang terkait dengan berita hoaks.

Seiring berkembangnya teknologi, model deep learning muncul sebagai alternatif yang lebih kuat untuk mendeteksi hoaks. Dua arsitektur yang paling menonjol dalam konteks ini adalah Long Short-Term Memory (LSTM) dan Convolutional Neural Networks (CNN) [11]. LSTM memiliki kemampuan khusus dalam menangkap sifat sekuensial dari data teks, sementara CNN efektif dalam mengidentifikasi pola dalam urutan teks, seperti frasa atau kata kunci spesifik yang mengindikasikan hoaks [12]. Model deep learning menawarkan keunggulan dalam kemampuannya untuk mempelajari fitur hierarkis secara otomatis dari teks mentah, meningkatkan performa klasifikasi dibandingkan dengan metode tradisional.

Akhir-akhir ini, pendekatan berbasis *Transfer-based Learning*, terutama *Transformer* mulai mendominasi dunia penelitian deteksi hoaks. Model seperti BERT [13] telah menjadi tulang punggung berbagai riset terkini, dengan variasi seperti IndoBERT [14] khusus dikembangkan untuk konteks bahasa Indonesia [15], [16]. Penelitian oleh Sinapoy et al. juga mendemonstrasikan bahwa model transformer, khususnya IndoBERT, mencapai performa yang lebih unggul dibandingkan LSTM dan pendekatan berbasis neural network lainnya dalam mendeteksi hoaks di platform seperti Twitter [17]. Karakteristik utama model Transformer, terutama kemampuan kontekstual dan representasi semantik yang mendalam, memungkinkan analisis yang lebih komprehensif terhadap struktur dan nuansa bahasa dalam mengidentifikasi potensi hoaks. Akan tetapi, mayoritas penelitian sebelumnya menggunakan dataset terbatas yang tidak sepenuhnya mewakili kompleksitas hoaks di Indonesia. Oleh karena itu, penelitian ini akan menggunakan dataset komprehensif yang mencakup spektrum lebih luas dari konten hoaks, dengan penekanan khusus pada konteks politik. Penelitian ini juga akan mengimplementasikan model IndoBERT untuk mendeteksi berita hoaks dalam konteks politik Indonesia. Pengembangan model ini menjadi sangat relevan mengingat meningkatnya penyebaran hoaks politik menjelang Pemilu 2024, di mana dibutuhkan sistem deteksi yang mampu memahami kompleksitas bahasa dan

konteks sosial-politik Indonesia.

1.2 Rumusan Masalah

Penelitian ini memiliki rumusan masalah sebagai berikut.

1. Bagaimana cara mengimplementasikan model IndoBERT dalam mendeteksi Berita Hoaks berbahasa Indonesia
2. Seberapa akurat model IndoBERT ini dalam mendeteksi berita hoaks berbahasa Indonesia berdasarkan *evaluation metrics*-nya (akurasi, presisi, *recall*, skor F_1)?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, penelitian ini memiliki tujuan sebagai berikut.

1. Melatih model IndoBERT untuk membuktikan kemampuan model ini dalam mendeteksi berita hoaks berbahasa Indonesia.
2. Mengukur *evaluation metrics* (akurasi, presisi, *recall*, skor F_1) dari model IndoBERT ini dalam mendeteksi berita hoaks berbahasa Indonesia.

1.4 Urgensi Penelitian

Penelitian ini dilakukan untuk mengembangkan sistem deteksi hoaks yang efektif menggunakan model IndoBERT, mengingat meningkatnya penyebaran informasi palsu di Indonesia. Dengan jumlah pengguna internet Indonesia yang mencapai 272 juta pada tahun 2023 dan peningkatan eksponensial kasus hoaks menjelang Pemilu 2024, kebutuhan akan sistem deteksi otomatis yang memahami konteks dan nuansa bahasa Indonesia menjadi sangat mendesak.

1.5 Luaran

Luaran dari penelitian ini adalah sebagai berikut.

- Model klasifikasi hoaks berbasis IndoBERT yang sudah di-*finetune*

- Dataset berita hoaks dan berita fakta dari Januari 2024 hingga September 2024.
- Artikel ilmiah di jurnal atau *conference* terakreditasi dan terindeks *Scopus*.

1.6 Manfaat

Penelitian ini memberikan kontribusi metodologis dalam pengembangan teknologi deteksi hoaks. Dengan menggunakan model IndoBERT, penelitian ini menawarkan pendekatan inovatif untuk memahami dan mengklasifikasi informasi palsu dalam konteks kebahasaan dan sosial-politik Indonesia. Model yang dikembangkan dapat digunakan oleh lembaga pemerintah, platform media sosial, dan organisasi yang peduli terhadap integritas informasi untuk secara proaktif mendeteksi dan mencegah penyebaran berita palsu. Selain itu, penelitian ini juga berkontribusi dalam meningkatkan literasi digital masyarakat.

