

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berita palsu menjadi informasi yang disajikan dengan fakta yang tidak benar atau bahkan sepenuhnya bohong. Di era digital saat ini, penyebaran berita semakin cepat dan dapat menjangkau khalayak luas dalam waktu singkat. Jika penyebaran berita palsu ini tidak dikendalikan, dampaknya bisa sangat merugikan, karena menyebabkan kesalahan pemahaman atas informasi yang diterima masyarakat. Ketika informasi yang salah atau menyesatkan terus menyebar, publik berisiko memperoleh pengetahuan yang tidak akurat, yang dapat mempengaruhi persepsi dan pengambilan keputusan mereka (Ramadhan et al., 2022).

Berita palsu sering kali sengaja dibuat dan disebarluaskan untuk memperoleh keuntungan atau mencapai tujuan tertentu, sehingga berita palsu ini dapat menyebar dengan cepat. Informasi yang salah atau menyesatkan dari berita palsu memiliki potensi besar untuk memengaruhi masyarakat karena dapat menimbulkan ketidakpercayaan dan kebingungan atas informasi yang diterima. Berita palsu juga bisa merusak reputasi individu atau kelompok yang menjadi sasarannya. Judul-judul yang digunakan dalam berita palsu biasanya sensasional dan provokatif, sengaja dibuat untuk menarik perhatian dan rasa ingin tahu pembaca. Berdasarkan data, sekitar 30% hingga hampir 60% masyarakat Indonesia terpapar hoaks ketika mengakses atau berkomunikasi di dunia maya, namun hanya sekitar 21% hingga 36% yang mampu mengenali atau mendeteksinya. Berita palsu yang beredar sebagian besar berkaitan dengan isu-isu politik, kesehatan, dan agama (Rakhmat Sani et al., 2022).

Para peneliti telah mengembangkan sejumlah strategi untuk mengatasi masalah ini, termasuk metode *hybrid machine learning*, *text classification*, dan *network analysis*. Penelitian menunjukkan bahwa klasifikasi teks menjadi pendekatan yang paling populer, dan banyak peneliti mengusulkan solusi dalam penelitian mereka dengan merumuskan model *machine learning* dan *deep learning*. (Tandiano & Jollyta, 2024).

Salah satu metode *deep learning* yang juga termasuk dalam arsitektur *transformers* yaitu *Bidirectional Encoder Representations from Transformers* (BERT). Penggunaan BERT dipilih karena metode ini pernah mencatatkan prestasi dengan mencapai *state-of-the-art* pada sebelas tugas utama dalam *Natural Language Processing* (NLP), termasuk di antaranya klasifikasi teks. Keberhasilan BERT dalam berbagai penelitian ini menunjukkan kemampuannya yang sangat efektif dalam memahami konteks teks secara lebih mendalam, baik dari sisi kiri maupun kanan (*bidirectional*), yang membuatnya sangat relevan dalam mengatasi berbagai permasalahan di bidang NLP (Rendragraha et al., 2021).

Seperti pada penelitian yang dilakukan oleh Asep Ripa'I, Firman Santoso, Farihin Lazim (2024), "Deteksi Berita *Hoax* dengan Perbandingan *Website* Menggunakan Pendekatan *Deep Learning* Algoritma BERT". Hasil penelitian tentang deteksi berita *hoax* dengan menguji tiga metode, yaitu BERT, SVM, dan *Random Forest*, menunjukkan bahwa ketiga metode tersebut memiliki performa yang sangat baik. Namun, metode BERT menunjukkan hasil yang paling unggul dengan akurasi mencapai 0,99 dan ROC-AUC sebesar 0,99. Keunggulan ini disebabkan oleh BERT yang merupakan algoritma *deep learning* dengan banyak lapisan, yang memungkinkannya untuk melakukan pembelajaran dan analisis yang lebih mendalam. Dalam analisis teks, ditemukan bahwa berita *hoax* seringkali menggunakan foto dan video yang dimanipulasi, yang tidak sesuai dengan konten asli.

Penelitian yang serupa juga dilakukan oleh Assyfa Rasida Hanum, Ivykaeyla Adriana Zetha, Julia Nur Fajrina, Rafifah Ayu Wulandari, Salwa Cahyani Putri, Sherla Puspa Andina dan Novanto Yudistira (2024), "Analisis Kinerja Algoritma Klasifikasi Teks BERT Dalam Mendeteksi Berita Hoaks". Penelitian ini menunjukkan bahwa algoritma BERT efektif dalam mendeteksi berita hoaks berbahasa Indonesia dengan menggunakan *dataset* gabungan dari tiga sumber publik yang terdiri dari lebih dari 44.000 artikel berita. Model klasifikasi BERT mencapai akurasi 76% pada data validasi dengan *F1-Score* 0.72, presisi 0.87, dan *recall* 0.62 untuk label positif (berita hoaks), menunjukkan kemampuan yang baik dalam mengidentifikasi berita palsu meskipun masih terdapat ruang perbaikan

dalam mendeteksi semua berita hoaks yang ada. Sementara itu, model BERT *Multilingual* menunjukkan performa yang lebih rendah dengan akurasi 63%, namun memiliki *recall* yang lebih tinggi (0.72) untuk mendeteksi berita hoaks. Hasil penelitian membuktikan bahwa pendekatan *fine-tuning* BERT dapat memberikan kontribusi signifikan dalam memerangi penyebaran disinformasi di era digital, meskipun model BERT standar menunjukkan performa superior dibandingkan varian multilingualnya dalam konteks deteksi berita hoaks berbahasa Indonesia. Penelitian ini memberikan fondasi penting untuk pengembangan sistem moderasi konten otomatis yang dapat membantu masyarakat menghindari konsumsi informasi yang menyesatkan.

Urgensi penelitian ini terletak pada dampak signifikan dari penyebaran berita palsu di *platform online*, yang dapat mengakibatkan disinformasi luas di masyarakat, memengaruhi persepsi publik, dan bahkan membahayakan reputasi individu atau kelompok tertentu. Di era digital, berita palsu dapat tersebar sangat cepat melalui media sosial dan *platform* berita, sehingga upaya deteksi otomatis berita palsu menjadi krusial. Berdasarkan laporan, sebagian besar masyarakat Indonesia terpapar hoaks, namun hanya sedikit yang mampu mendeteksinya. Fenomena ini menyoroti lemahnya literasi informasi digital di masyarakat dan perlunya teknologi yang efektif dalam membantu mengidentifikasi kebenaran informasi. Mengingat perkembangan algoritma seperti *Bidirectional Encoder Representations From Transformers* (BERT) yang terbukti efektif dalam klasifikasi berita palsu, penelitian ini penting untuk mengembangkan model deteksi otomatis yang akurat, guna mengurangi penyebaran berita palsu secara masif. Model ini diharapkan dapat berkontribusi secara langsung dalam menyediakan lingkungan digital yang lebih terpercaya, serta mendukung pengambilan keputusan yang tepat berbasis informasi yang valid.

Berdasarkan uraian penjelasan latar belakang diatas, maka judul yang akan diangkat pada penelitian ini yaitu, "**Pengembangan Model Deteksi Berita Palsu Pada Platform Berita Online Menggunakan Metode *Bidirectional Encoder Representations From Transformers* (BERT)**".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, maka permasalahan yang penulis rumuskan adalah :

1. Bagaimana mengimplementasikan algoritma *Bidirectional Encoder Representations From Transformers* (BERT) dalam mendeteksi berita palsu pada *platform* berita *online*?
2. Bagaimana merancang sistem deteksi berita palsu pada *platform* berita *online* menggunakan algoritma *Bidirectional Encoder Representations From Transformers* (BERT)?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang diuraikan di atas, maka tujuan penelitian adalah sebagai berikut :

1. Untuk mengimplementasikan algoritma *Bidirectional Encoder Representations From Transformers* (BERT) dalam mendeteksi berita palsu pada *platform* berita *online*.
2. Untuk merancang sistem deteksi berita palsu pada *platform* berita *online* menggunakan algoritma *Bidirectional Encoder Representations From Transformers* (BERT).

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Penelitian ini memberikan kontribusi dalam membantu masyarakat mengidentifikasi berita palsu secara lebih akurat dan efisien, sehingga masyarakat dapat terhindar dari misinformasi. Dengan adanya model deteksi yang efektif, masyarakat akan lebih terlindungi dari dampak negatif berita palsu, seperti ketidakpastian informasi dan ketidakpercayaan.
2. Penelitian ini memberikan referensi serta landasan ilmiah bagi penelitian-penelitian selanjutnya di bidang deteksi berita palsu. dengan menggunakan metode *Bidirectional Encoder Representations From Transformers*, penelitian ini memperkaya kajian akademis dalam klasifikasi teks, khususnya di ranah *machine learning* dan *natural language processing* (NLP).

3. Penelitian ini memperkaya aplikasi *machine learning* dalam konteks lokal Indonesia, khususnya dengan pengolahan data teks berbahasa Indonesia. Pengembangan model deteksi ini mendorong inovasi di bidang *machine learning*, yang dapat diaplikasikan lebih lanjut dalam berbagai konteks, seperti literasi informasi, keamanan siber, dan moderasi konten *online*.
4. Sebagai sarana untuk menerapkan pengetahuan peneliti yang diperoleh selama menempuh masa studi Strata-1 Teknik Informatika Universitas Malikussaleh.

1.5 Ruang Lingkup dan Batasan Masalah

Adapun batasan masalah yang ditetapkan pada penelitian ini adalah sebagai berikut:

1. Lingkup Data dan Sumber: Data yang digunakan dalam penelitian ini dibatasi pada berita-berita dalam Bahasa Indonesia yang berasal dari *platform* berita *online* terpercaya. *Dataset* penelitian terdiri dari 248 artikel berita yang dikurasi dari empat sumber utama, yaitu Kompas.com, TurnBackHoax.id, CNN Indonesia, dan Antara News. Data yang dipilih telah melalui proses verifikasi dan validasi untuk memastikan keaslian label klasifikasi antara berita asli (*REAL*) dan berita palsu (*FAKE*). Penelitian ini tidak mencakup analisis berita dari media sosial atau *platform* berbagi konten lainnya.
2. Fokus Analisis Tekstual: Penelitian ini membatasi analisis pada konten judul berita (*news titles*) sebagai fitur utama untuk klasifikasi. *Text preprocessing* yang diterapkan meliputi tokenisasi menggunakan IndoBERT *tokenizer*, *truncation* dan *padding* dengan panjang maksimal 128 token, serta konversi ke format *tensor PyTorch*. Penelitian ini tidak melakukan *stemming*, penghilangan *stopwords*, atau transformasi linguistik lainnya, karena model BERT telah dirancang untuk memahami konteks bahasa secara komprehensif termasuk kata-kata umum dan variasi morfologi.
3. Metode Klasifikasi: Penelitian ini hanya menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT), khususnya model IndoBERT (*indolem/indobert-base-uncased*) yang telah dioptimalkan untuk

bahasa Indonesia. Model di-*fine-tune* menggunakan pendekatan *supervised learning* dengan konfigurasi klasifikasi biner.

4. Metrik Evaluasi: Evaluasi performa model dibatasi pada metrik standar klasifikasi biner, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* (baik per kelas maupun *macro* dan *weighted average*). Analisis tambahan menggunakan *confusion matrix* untuk memahami distribusi kesalahan klasifikasi dan tingkat kepercayaan (*confidence score*) untuk setiap prediksi.
5. Konfigurasi Teknis: Penelitian ini menggunakan konfigurasi pelatihan yang dioptimalkan untuk lingkungan komputasi terbatas dengan 3 *epoch*, *batch size* 4, *learning rate* $2e-5$, dan *gradient accumulation steps* 2. Pembagian data menggunakan *stratified train-test split* dengan proporsi 80:20 (156 sampel *training*, 40 sampel *testing*) untuk memastikan distribusi yang seimbang antara kedua kelas pada setiap *subset* data.