

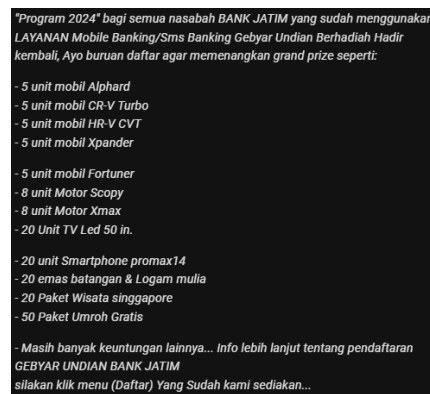
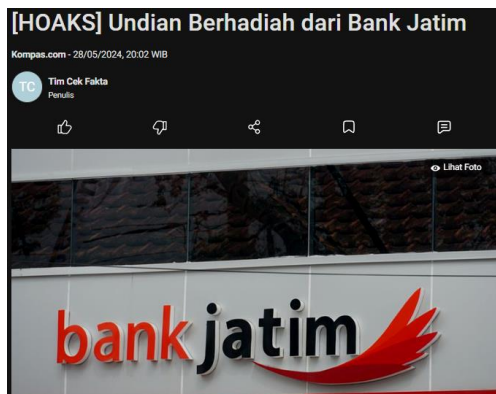
BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Teknologi telah mengalami perkembangan yang pesat selama beberapa dekade terakhir. Perkembangan terjadi di berbagai sektor kehidupan manusia. Internet merupakan salah satu contoh hasil perkembangan teknologi di bidang informasi dan komunikasi. Internet adalah jaringan komputer global yang menghubungkan berbagai *server* dan *client* di seluruh dunia yang memungkinkan semua negara dapat berkomunikasi dan berbagi informasi secara *real time* (Cedillo et al., 2017).

Internet memiliki dampak positif yang besar bagi kehidupan manusia tetapi internet juga memiliki dampak negatif, salah satunya yaitu munculnya berita palsu atau hoaks. Berita palsu atau hoaks adalah suatu informasi yang salah, tidak benar, dan keliru (Wardle et al., 2017). Definisi lain dari hoaks adalah suatu informasi yang tidak dapat dibuktikan kebenarannya dan mempunyai tujuan untuk hal-hal yang tidak baik seperti menyesatkan seseorang dan melakukan propaganda (Shu et al., 2017). Contoh berita yang sudah ditetapkan sebagai hoaks oleh tim cek fakta dari Kompas.com dapat dilihat pada Gambar 1.1 dan Gambar 1.2.



Gambar 1.1 Contoh Hoaks (Kompas, 2024a)

Gambar 1.2 Narasi Hoaks (Kompas, 2024a)

Berita palsu atau hoaks dapat menimbulkan bahaya yang besar bagi masyarakat di berbagai bidang. Di bidang politik, hoaks yang menyerang individu

atau organisasi politik dapat merusak citra mereka di mata publik. Hal ini dapat berakibat hilangnya dukungan dan suara ketika pemilihan umum berlangsung (Sindermann et al., 2020). Di bidang ekonomi, hoaks yang berisi informasi negatif tentang produk atau layanan tertentu dapat merusak reputasi perusahaan dan menurunkan penjualan. Hal ini dapat berakibat pada kerugian finansial bahkan kebangkrutan dari perusahaan (Christov, 2019). Di bidang kesehatan, hoaks tentang bahaya dan ketidakefektifan vaksin dapat menyebabkan masyarakat enggan untuk divaksinasi. Hal ini dapat membahayakan kesehatan masyarakat yang lain dan menghambat upaya penanggulangan serta pencegahan penyakit tersebut (Pulido et al., 2020). Berikut adalah contoh nyata dari bahaya yang ditimbulkan dari hoaks.

Hoax Corona dr Lois Makan Korban Jiwa, Polri Turun Tangan

Adhyasta Dirgantara · detikNews

Senin, 19 Jul 2021 09:34 WIB



Ilustrasi Corona (Edi Wahyono/detik.com)

Jakarta - Hoax seputar Corona atau COVID-19 yang disampaikan oleh dr Lois Owlen diduga menjadi pemicu seorang pria di Tegal wafat saat positif Corona. Polisi bakal turun tangan mengecek hal itu.



Gambar 1.3 Dampak Hoaks COVID-19 (Detik, 2021)

Gambar 1.3 menunjukkan dampak dari hoaks yang memakan korban jiwa. Dalam berita online yang dirilis oleh Detik.com, seorang pria dari Tegal menjadi korban akibat mempercayai pernyataan Dr. Lois yang tersebar di media sosial. Dr. Lois mengklaim bahwa kematian akibat COVID-19 sebenarnya disebabkan oleh interaksi obat, bukan virus itu sendiri. Sehingga pria tersebut menolak untuk mengonsumsi obat apapun selama terinfeksi COVID-19, yang akhirnya berujung pada kematian (Detik, 2021). Kasus ini menyoroti betapa berbahayanya informasi

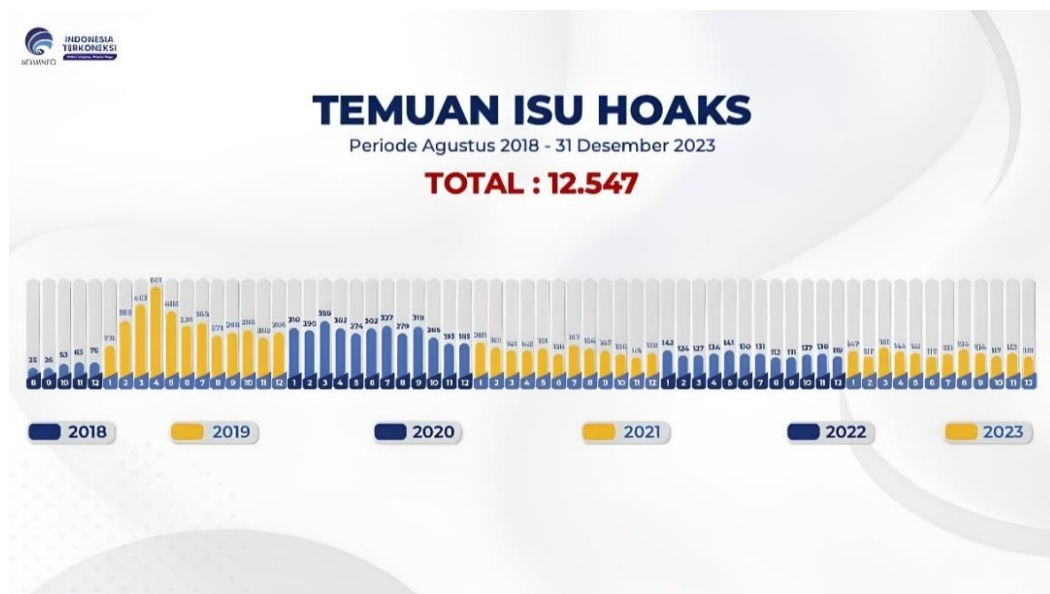
Listia Ningrum, 2024

DETEKSI BERITA PALSU MENGGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

palsu yang beredar di media sosial, karena dapat menyebabkan kesalahpahaman yang fatal dan merugikan nyawa manusia.

Berita palsu tersebar di internet melalui media sosial, beberapa media sosial yang menjadi tempat penyebaran hoaks adalah Facebook, Twitter, Instagram, TikTok, dan YouTube (Meesad, 2021). Banyaknya hoaks yang muncul di media sosial terjadi karena semakin mudahnya akses informasi sehingga siapa pun dapat membuat dan menyebarkan konten di internet. Platform media sosial ini memiliki jangkauan luas dan kecepatan distribusi yang sangat tinggi, sehingga informasi palsu dapat dengan cepat menjangkau ribuan bahkan jutaan orang dalam waktu singkat. Dengan kecepatan dan kemudahan dalam menyebarkan informasi ini, tidak mengherankan hoaks juga dapat menyebar dengan cepat dan luas. Berikut adalah temuan isu hoaks di media sosial pada Agustus 2018 sampai Desember 2023.



Gambar 1.4 Temuan Isu Hoaks Agustus 2018 - 31 Desember 2023 (Kominfo, 2024)

Gambar 1.4 menunjukkan temuan isu hoaks pada periode Agustus 2018 sampai 31 Desember 2023 yang berjumlah 12.547 isu hoaks. Jumlah isu hoaks yang beredar menunjukkan tren yang mengkhawatirkan. Isu hoaks di sektor kesehatan menjadi yang paling banyak ditemukan. Isu yang paling dominan masih berkaitan dengan penyebaran COVID-19. Selain itu, terdapat banyak informasi menyesatkan

Listia Ningrum, 2024

DETEKSI BERITA PALSU MENGGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

tentang obat-obatan dan produk kesehatan yang beredar di masyarakat (Kominfo, 2024).

Banyaknya hoaks yang beredar di masyarakat serta bahaya yang ditimbulkan, perlunya untuk menangani berita palsu secara cepat dan efisien. Untuk mengatasi masalah ini, diperlukan otomatisasi dalam proses deteksi dan verifikasi informasi. Otomatisasi adalah penggantian tenaga manusia dengan tenaga mesin yang secara otomatis melakukan dan mengatur pekerjaan sehingga tidak memerlukan lagi pengawasan manusia (KBBI, 2016c). Otomatisasi dapat dilakukan dengan memanfaatkan teknologi seperti *Natural Language Processing* untuk membangun sistem yang dapat mendeteksi berita palsu (Oshikawa et al., 2018).

Natural Language Processing (NLP) adalah cabang ilmu yang menggabungkan linguistik, ilmu komputer, rekayasa data, dan kecerdasan buatan. NLP berfokus pada interaksi antara manusia dan komputer serta metode untuk memproses dan menganalisis data bahasa alami dalam jumlah besar. Beberapa aplikasi NLP meliputi mesin translasi, pengenalan suara, analisis sentimen, peringkasan, *chatbot*, dan klasifikasi teks (Meesad, 2021). Dalam era digital, NLP semakin penting karena volume informasi yang perlu dianalisis terus meningkat. Kemampuannya untuk memahami dan mengolah bahasa manusia membuatnya sangat berguna dalam berbagai konteks, termasuk deteksi berita palsu.

Salah satu model yang dapat digunakan untuk menyelesaikan berbagai permasalahan terkait NLP khususnya klasifikasi teks adalah transformers. Transformers merupakan sebuah model deep learning yang telah digunakan untuk berbagai macam *task* seperti *Natural Language Processing*, *Computer Vision*, dan *Speech Recognition* (Lin et al., 2022). Keunggulan transformers adalah kemampuannya untuk menangani *long-range dependencies*, yaitu kemampuan model untuk memahami hubungan antara kata yang berada jauh satu sama lain dalam sebuah teks. Transformer juga memiliki keunggulan untuk beradaptasi dengan berbagai ukuran input, kemampuan dalam pemrosesan paralel, dan skalabilitas yang tinggi dalam menangani dataset besar (Maurício, 2023). Arsitektur transformers terdiri dari blok *encoder* dan *decoder*. *Encoder* merupakan sebuah *layer* yang digunakan untuk *text classification*, *decoder* digunakan untuk *text*

generation, dan *encoder-decoder* digunakan untuk *machine translation* (Tunstall et al., 2022). Arsitektur transformers *encoder* sangat cocok digunakan untuk menyelesaikan permasalahan terkait deteksi berita palsu.

Contoh model yang mengadopsi arsitektur dari transformers *encoder* adalah Bidirectional Encoder Representation from Transformers (BERT). BERT merupakan sebuah model yang dirancang untuk mempelajari representasi teks yang tidak berlabel dengan memperhitungkan konteks, baik dari sebelah kiri maupun kanan (*bidirectional*). Keunggulan BERT adalah kemampuannya untuk memahami konteks dari dua arah, yang mampu memahami makna kalimat jauh lebih baik dibandingkan model-model *unidirectional*. Selain itu, BERT juga mendukung proses *pre-training* dan *fine-tuning* sehingga dapat disesuaikan untuk berbagai tugas pengolahan bahasa alami. Model BERT yang telah dilatih dapat disesuaikan kembali dengan menambahkan satu *output layer* tambahan yang dapat disesuaikan dengan masalah yang ingin diselesaikan, contohnya adalah masalah klasifikasi teks untuk mendeteksi berita palsu (Devlin et al., 2018).

Berbagai penelitian sebelumnya telah menunjukkan efektivitas BERT dalam mendeteksi berita palsu. Isa et al., (2022) melakukan penelitian untuk membandingkan model IndoBERT, Support Vector Machine dan Naïve Bayes dalam mendeteksi berita palsu. Hasilnya model IndoBERT mencapai skor akurasi 94.66%, SVM dengan akurasi 90%, dan Naive Bayes dengan akurasi 83%. Hal ini menunjukkan bahwa model IndoBERT memiliki kinerja yang lebih baik untuk mendeteksi berita palsu. Rahmawati et al., (2022) melakukan penelitian untuk membandingkan model IndoBERT, Support Vector Machine dan Naïve Bayes dalam mendeteksi berita palsu. Hasilnya model IndoBERT mencapai skor akurasi 90%, SVM dengan akurasi 73.42%, Naive Bayes dengan akurasi 80.25%, dan menemukan bahwa IndoBERT mampu mendeteksi aspek kontekstual yang terkait dengan pola yang lebih kompleks. Yefferson et al., (2024) melakukan penelitian untuk membandingkan model IndoBERT-LSTM dengan LSTM, dan IndoBERT-BiLSTM dalam mendeteksi berita palsu. Hasilnya model IndoBERT-LSTM berhasil mencapai skor akurasi 93.2%, LSTM dengan akurasi 83.7%, dan IndoBERT-BiLSTM dengan akurasi 93.1%. Model IndoBERT-LSTM ini

mengungguli model LSTM tunggal dan IndoBERT-BiLSTM dalam mendeteksi hoaks. Sinapoy et al., (2023) melakukan penelitian untuk membandingkan akurasi metode LSTM dan IndoBERT dalam mendeteksi hoaks pada Twitter terkait Covid-19. Hasilnya model LSTM memberikan nilai akurasi rata-rata sebesar 87.54%, sementara model IndoBERT sebesar 92.07% dalam mendeteksi hoaks di Twitter terkait COVID-19.

Berdasarkan uraian di atas, topik penelitian skripsi yang akan diangkat adalah “Deteksi Berita Palsu Menggunakan Bidirectional Encoder Representations from Transformers (BERT)”. Penelitian ini akan merancang sebuah sistem deteksi berita palsu bahasa Indonesia dengan menggunakan model BERT. Model Indo-Bidirectional Encoder Representations from Transformers (IndoBERT) dan Multilingual-Bidirectional Encoder Representations from Transformers (MBERT) akan dievaluasi untuk melihat model mana yang memiliki kinerja terbaik dalam mendeteksi berita palsu.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan, peneliti merumuskan beberapa rumusan masalah sebagai berikut.

1. Bagaimana mengembangkan sistem deteksi berita palsu bahasa Indonesia menggunakan model *fine-tuning* IndoBERT dan MBERT?
2. Bagaimana perbandingan hasil evaluasi kinerja sistem deteksi berita palsu bahasa Indonesia menggunakan model *fine-tuning* IndoBERT dan MBERT?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, maka tujuan dari penelitian ini yaitu sebagai berikut.

1. Mengembangkan sistem deteksi berita palsu dengan menggunakan model *fine-tuning* IndoBERT dan MBERT untuk mencapai akurasi deteksi yang optimal.
2. Menganalisis dan membandingkan kinerja model IndoBERT dan MBERT dalam mendeteksi berita palsu untuk menentukan model yang paling efektif.

1.4 Manfaat Penelitian

Berdasarkan tujuan yang telah dijabarkan, penelitian ini diharapkan memberikan berbagai manfaat yang dapat dirinci sebagai berikut.

1. Memberikan kontribusi dalam pengembangan teknologi informasi khususnya dalam deteksi berita palsu di Indonesia, dengan menyediakan model yang dapat diandalkan untuk mengidentifikasi berita palsu secara akurat.
2. Menyediakan referensi dan panduan bagi peneliti dan praktisi dalam pengembangan sistem deteksi berita palsu bahasa Indonesia.

1.5 Batasan Penelitian

Batasan penelitian yang akan ditetapkan dalam penelitian ini guna memfokuskan penelitian yaitu sebagai berikut.

1. Memfokuskan pada berita bahasa Indonesia sebagai objek utama analisis untuk memastikan bahwa penelitian ini relevan dengan konteks lokal dan dapat memberikan hasil yang akurat untuk deteksi berita palsu.
2. Dataset yang digunakan dalam penelitian berasal dari website Detik.com dan Turnbackhoax.id. Kedua sumber ini digunakan karena mewakili berita valid dan berita yang sudah diverifikasi sebagai hoaks. Pemilihan ini bertujuan untuk menyediakan data yang representatif untuk menyelesaikan permasalahan.
3. Penelitian ini hanya mencakup konten berita dalam bentuk teks dan tidak mencakup berita atau informasi dalam bentuk multimedia seperti gambar, video, atau audio, yang memiliki karakteristik berbeda dalam proses pelatihan data.
4. Model yang dikembangkan dalam penelitian ini dirancang berdasarkan dataset yang tersedia saat ini, sehingga efektivitasnya dapat berkurang seiring dengan perubahan pola penyebaran dan struktur kalimat dari berita palsu di masa mendatang. Keterbatasan ini perlu diperhatikan, terutama dalam konteks adaptasi model terhadap perkembangan tren berita dan informasi di masa depan.

1.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun secara terstruktur dengan tujuan memberikan gambaran yang komprehensif dan jelas mengenai isi serta alur penelitian. Penulisan ini dirancang untuk memastikan bahwa setiap aspek dari

penelitian dapat dipahami dengan mudah oleh pembaca dan memberikan informasi yang diperlukan untuk mendalami topik penelitian secara mendalam. Berikut adalah sistematika penulisan yang diterapkan dalam penelitian ini.

BAB I. PENDAHULUAN

Bab ini membahas latar belakang yang melatarbelakangi penelitian mengenai deteksi berita palsu menggunakan model pre-trained Bidirectional Encoder Representations Transformer (BERT), yaitu IndoBERT dan MBERT. Penjelasan mengenai pentingnya topik ini akan diuraikan untuk memberikan konteks dan alasan utama pemilihan topik penelitian. Rumusan masalah akan disusun untuk mengidentifikasi permasalahan spesifik yang akan dipecahkan melalui penelitian ini. Tujuan penelitian akan dijelaskan untuk menetapkan apa yang ingin dicapai dari penelitian ini. Manfaat penelitian akan menguraikan kontribusi yang diharapkan bagi bidang studi terkait serta aplikasinya dalam konteks yang lebih luas. Batasan penelitian akan mengidentifikasi ruang lingkup dan batasan-batasan yang ditetapkan pada penelitian untuk memperjelas fokus penelitian. Terakhir, sistematika penulisan akan memberikan panduan tentang isi dokumen untuk memudahkan pembaca dalam mengikuti alur penelitian.

BAB II. KAJIAN PUSTAKA

Bab ini menyajikan tinjauan pustaka yang meliputi penelitian terdahulu dan landasan teori yang mendasari topik penelitian. Kajian pustaka ini mencakup teori-teori terkait berita palsu, termasuk definisi, karakteristik, dampak, dan teknik untuk mendeteksinya. Selain itu, bab ini juga membahas aspek penting dari machine learning, deep learning, serta teknologi Transformers dan BERT dalam deteksi berita palsu. Fokus utama kajian pustaka adalah untuk memberikan pemahaman mendalam tentang teknologi dan teori yang mendasari penelitian ini serta mendukung pengembangan konsep dan metodologi penelitian.

BAB III. METODOLOGI PENELITIAN

Bab ini menguraikan secara rinci tahapan dan alur penelitian untuk mengembangkan dan mengevaluasi sistem deteksi berita palsu menggunakan model BERT. Penelitian dimulai dengan identifikasi masalah dan penetapan tujuan penelitian. Studi literatur dilakukan untuk memperoleh wawasan tentang metode

yang ada dan menginformasikan pendekatan yang akan diambil. Pengumpulan data melibatkan identifikasi dan pengumpulan berita palsu serta berita asli dari sumber terpercaya. Tahap praproses data mencakup pembersihan dan transformasi data. Eksperimen dilakukan dengan *fine-tuning* model IndoBERT dan MBERT, mengoptimalkan hyperparameter, serta mengevaluasi performa model. Pembuatan antarmuka pengguna yang sederhana juga merupakan bagian dari metodologi untuk memfasilitasi pengujian model dalam kondisi nyata. Akhir dari bab ini menyajikan kesimpulan dari keseluruhan proses penelitian.

BAB IV. HASIL DAN PEMBAHASAN

Bab ini memberikan penjelasan mendalam tentang hasil penelitian. Deskripsi dataset mencakup detail tentang jumlah data, sumber data, dan karakteristik data yang digunakan dalam penelitian. Proses *fine-tuning* model IndoBERT dan MBERT dijelaskan dengan rinci, termasuk konfigurasi hyperparameter yang diterapkan serta hasil pelatihan yang diperoleh. Hasil eksperimen disajikan dalam bentuk grafik, tabel, dan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Perbandingan dengan model lain jika relevan juga dilakukan untuk memberikan perspektif tentang performa model. Bab ini juga mencakup penjelasan mengenai pembuatan dan uji coba antarmuka sistem deteksi berita palsu, termasuk fitur-fitur yang tersedia dan hasil penggunaannya dalam pengujian.

BAB V. PENUTUP

Bab ini menyajikan kesimpulan penelitian, dan rekomendasi untuk penelitian atau pengembangan lebih lanjut. Kesimpulan merangkum hasil utama dan mengukur pencapaian tujuan penelitian. Rekomendasi memberikan saran untuk langkah-langkah lanjutan yang dapat diambil untuk memperbaiki atau memperluas penelitian di masa depan.