
Klasifikasi Berita Palsu Menggunakan Model Bidirectional Encoder Representations From Transformers (BERT)

Enjeli Cistia Sukmawati, Lintang Suryaningrum, Diva Angelica,
Nur Ghaniaviyanto Ramadhan

Program Studi Rekayasa Perangkat Lunak, Fakultas Informatika, Institut Teknologi Telkom Purwokerto
Email: 21104001@ittelkom-pwt.ac.id; 21104013@ittelkom-pwt.ac.id; 21104015@ittelkom-pwt.ac.id;
ghani@ittelkom-pwt.ac.id.

Diterima:
12 Agustus 2024

Diterima Setelah Revisi:
13 Agustus 2024

Dipublikasikan:
16 Agustus 2024

Abstrak

Penyebaran informasi palsu menjadi tantangan serius dalam era digital yang terus berkembang, terutama melalui internet dan platform media sosial. Akses mudah terhadap informasi tidak terverifikasi menciptakan tantangan membedakan antara fakta dan hoaks. Salah satu aspek utama yang perlu diatasi adalah mengklasifikasikan berita palsu dengan tingkat akurasi yang tinggi. Berita sebagai sumber informasi aktual memerlukan pengelompokan untuk memfasilitasi akses, namun tidak semua berita dari berbagai sumber memiliki kredibilitas tinggi, terutama dengan adanya *fake news*. *Fake news* dapat merugikan individu dan berpotensi memanipulasi persepsi masyarakat, terutama melalui media sosial. Mengidentifikasi informasi palsu menjadi tantangan dalam *Natural Language Processing* (NLP) dengan pertumbuhan pesat platform media sosial. Meskipun ada beberapa metode untuk mendeteksi berita palsu, belum ada platform yang secara luas dikenal menerapkan algoritma terfokus pada sumber berita spesifik. Penelitian ini bertujuan untuk mengatasi masalah penyebaran informasi palsu dengan mengelompokkan berita palsu berdasarkan ciri-ciri bahasa, mengadopsi metode klasifikasi menggunakan teknologi *Bidirectional Encoder Representations from Transformers* (BERT). BERT sebagai model bahasa yang dilatih mendalam dapat memahami konteks kata dalam teks dengan baik. Teknologi BERT diadopsi untuk meningkatkan akurasi deteksi berita palsu. Meskipun BERT memiliki kompleksitas, sumber model yang sudah dilatih oleh Google memudahkan penggunaan tanpa perlu membuat model dari awal. Dengan langkah *pretraining* dan *fine tuning*, BERT lebih akurat dalam mendeteksi berita palsu dibandingkan metode lainnya. Penelitian ini memberikan kontribusi dalam menghadapi tantangan penyebaran informasi palsu dengan memanfaatkan keunggulan teknologi BERT dalam mengklasifikasikan berita.

Kata Kunci: Berita Palsu, BERT, Klasifikasi, NLP

Abstract

The spread of false information poses a serious challenge in the evolving digital era, especially through the internet and social media platforms. Easy access to unverified information creates challenges in distinguishing between facts and hoaxes. One key aspect that needs to be addressed is accurately classifying fake news. News, as a source of current information, requires categorization to facilitate access, but not all news from various sources has high credibility, especially with the presence of fake news. Fake news can harm individuals and potentially manipulate public perception, especially through social media. Identifying fake information poses a challenge in Natural Language Processing (NLP) with the rapid growth of social media platforms. Although there are several methods for detecting fake news, there is no widely known platform that applies

algorithms focused on specific news sources. This research aims to address the issue of false information dissemination by classifying fake news based on language characteristics, adopting a classification method using Bidirectional Encoder Representations from Transformers (BERT) technology. BERT, as a deeply trained language model, understands the context of words in the text well. BERT technology adopted to enhance the accuracy of fake news detection. Although BERT has complexity, the pretrained model source provided by Google facilitates usage without the need to create a model from scratch. Through pretraining and fine-tuning steps, BERT is more accurate in detecting fake news compared to other methods. This research contributes to addressing the challenge of false information dissemination by leveraging the advantages of BERT technology in classifying news.

Keywords: BERT, Classification, Fake News, NLP

1 PENDAHULUAN

Di era digital yang semakin berkembang, penyebaran informasi dan berita palsu semakin mudah dan luas serta telah menjadi sebuah permasalahan yang signifikan. Dengan mudahnya akses terhadap internet dan platform media sosial, informasi yang tidak terverifikasi atau palsu dapat menyebar dengan cepat dan luas. Hal ini telah menciptakan tantangan baru dalam membedakan antara fakta dan hoaks serta memerlukan kewaspadaan ekstra dari masyarakat dalam mengonsumsi dan membagikan informasi. Salah satu tantangan yang harus diatasi dalam bidang ini ialah mengklasifikasikan berita palsu dengan tingkat akurasi yang tinggi. Di era penyebaran informasi yang tidak benar semakin meningkat ini, sangat penting untuk mengembangkan teknik yang efisien dalam mengidentifikasi dan membedakan antara artikel berita yang sah dan berita yang tidak benar. Sebelumnya, penelitian belum sepenuhnya menjelajahi potensi teknik pemodelan topik yang menggunakan model bahasa kontekstual seperti *Bidirectional Encoder Representations from Transformers* (BERT) untuk meningkatkan akurasi dalam mengklasifikasikan berita palsu. Integrasi model ini dapat meningkatkan kinerja metode tersebut dan memberikan pemahaman yang lebih mendalam tentang konten berita (Sholikhah et al., 2023).

Berita merupakan salah satu sumber informasi mengenai kejadian terkini yang ada pada media massa seperti surat kabar, televisi, dan media lainnya (Kasanah et al., 2019). Diperlukan pengelompokan atau kategorisasi berita yang tersedia untuk mempermudah akses dan penggunaan informasi yang banyak tersebut (Pramudita et al., 2018). Berita yang diakses dari beberapa sumber belum tentu memiliki kredibilitas yang tinggi terkait suatu peristiwa. Hal ini disebabkan oleh penerbit berita yang mungkin menggunakan informasi yang salah atau menyesatkan untuk memajukan tujuan atau kepentingan tertentu, yang sering kali disebut sebagai *fake news* (Tandijaya et al., 2021). *Fake news* atau berita palsu memiliki dampak yang sangat negatif bagi individu yang cenderung mudah percaya, bahkan dapat merusak karakter mereka. Berita palsu memiliki potensi untuk memanipulasi individu atau kelompok dalam jangka waktu yang panjang, tanpa disadari, membentuk persepsi yang salah dan berdampak pada pemahaman yang keliru dalam masyarakat. Salah satu tempat yang menjadi penyebaran *fake news* yaitu media sosial seperti Twitter, Youtube, Whatsapp, Facebook, Instagram, Line (Riadi Silitonga, 2019). Media sosial berperan strategis untuk transformasi informasi dan sarana komunikasi. Sehingga banyaknya *fake news* atau yang sering kita kenal sebagai *hoax* tersebar luas di media social (Rahadi, 2017).

Mengidentifikasi informasi yang tidak valid adalah sebuah tantangan penting dalam *Natural Language Processing* (NLP). Pertumbuhan cepat dari platform-platform media sosial tidak hanya mengakibatkan lonjakan besar dalam akses terhadap informasi, tetapi juga mempercepat penyebaran informasi yang tidak akurat (Oshikawa et al., 2020). Dikarenakan data terkait informasi yang tidak benar sudah tersedia dalam jumlah besar, menciptakan model dengan menerapkan algoritma yang berbasis pada sumber-sumber berita menjadi suatu tugas yang cukup mudah. Saat ini, belum ada

platform yang secara luas dikenal yang menerapkan algoritma yang terfokus pada sumber berita spesifik. Meskipun ada beberapa metode untuk mendeteksi berita palsu, sampai saat ini belum ada yang berhasil menggabungkan berbagai model ini menjadi satu kesatuan yang komprehensif (Muzakir & Suriani, 2023). Oleh karena itu, diperlukan sebuah platform yang menghimpun berita dari berbagai sumber dan mengkategorikannya untuk mempermudah pengguna dalam mengakses informasi (Yunanto et al., 2021).

Penelitian ini bertujuan untuk menangani masalah yang terkait dengan penyebaran informasi palsu di berbagai kategori. Berita palsu dikelompokkan berdasarkan ciri-ciri bahasa secara eksklusif dalam studi ini. Metode klasifikasi menggunakan teknologi BERT diadopsi agar deteksi berita palsu dapat dilakukan dengan lebih akurat.

2 KAJIAN PUSTAKA

Bidirectional Encoder Representations from Transformers (BERT) adalah model bahasa yang dilatih dengan mendalam dengan data yang besar agar memahami konteks kata dalam teks dengan baik (Sains et al., 2023). BERT merupakan teknik pembelajaran mesin yang dikembangkan oleh Google berbasis Transformer untuk pra-pelatihan pemrosesan Bahasa alami (*Natural Language Processing*). Ini merupakan model penyandingan dua arah yang khas yang dibangun dengan transformator arsitektur transformer (Sholikhah et al., 2023). BERT terdiri dari enam lapisan Transformer yang disusun di atas *encoder* dan *decoder*, masing-masing menambah kompleksitas proses pelatihan, konfigurasi yang tinggi, waktu yang lama untuk melatih, dan biaya yang tinggi. Google telah menyediakan sumber dari model yang sudah dilatih sebelumnya dari BERT, yang dapat digunakan oleh penulis tanpa perlu membuat model dari awal (Devlin et al., 2019). Ada dua langkah untuk kinerja *framework* BERT, yaitu ada *pre-training* dan *fine-tuning* (Tandijaya et al., 2021). Saat melakukan tahap *pre-training*, akan digunakan Masked Language Model (MLM) yang secara acak menyamarkan beberapa token dari kalimat input. Tujuannya adalah untuk memprediksi kata yang telah disamarkan berdasarkan konteks kalimat tersebut. Dalam metode BERT, teks kalimat diubah menjadi urutan token. Ada token [CLS] yang menandakan awal suatu kalimat dan token [SEP] yang memisahkan antara kalimat. Selain itu, token [PAD] digunakan untuk menambahkan *padding* atau token kosong ke dalam kalimat untuk memenuhi panjang tetap dari kalimat yang dimasukkan (Halim et al., 2022). Di samping itu, dalam struktur BERT terdapat penggunaan *transformer encoder* yang berperan dalam memahami keterkaitan antara kalimat-kalimat (Vaswani et al., 2017). Oleh karena itu metode yang tepat untuk penelitian ini adalah *Bidirectional Encoder Representations from Transformers*. Pemilihan algoritma BERT karena dinilai memiliki tingkat akurasi lebih tinggi daripada metode lainnya (Mas et al., 2021).

Beberapa peneliti telah menggunakan metode BERT untuk klasifikasi teks. Beberapa peneliti yang menggunakan metode ini yaitu: pada jurnal FASILKOM Universitas Muhammadiyah Riau dengan judul “Analisis Sentimen terhadap Pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT” peneliti menggunakan metode BERT untuk penelitian mereka. Peneliti memperoleh data melalui *review* pengguna yang ada di *Google Maps* (Ardiansyah et al., 2023). Dalam penelitian yang telah dilakukan oleh Liu dan Lapata untuk menerapkan serta membandingkan metode BERT dengan *Recurrent Neural Network*, hasilnya menunjukkan bahwa metode BERT menghasilkan kinerja yang baik (Liu & Lapata, 2019). Penelitian di jurnal “Analisis Sentimen Terhadap Game Genshin Impact Menggunakan BERT” pada penelitian ini diperlukan sebuah sistem interaktif. Metode yang digunakan penulis adalah metode *CRoss-Industry Standard Process for Data Mining* (Kusnadi et al., 2021).

Selanjutnya ada penelitian yang membandingkan nilai akurasi BERT dan DistilBERT pada *dataset* Twitter. Peneliti menggunakan DistilBERT yang merupakan pengembangan dari BERT. Peneliti menggunakan data twitter yang berhubungan dengan Covid-19 untuk penelitiannya (Fajri et al., 2022). Jurnal yang berjudul “Analisis Sentimen KUHP Baru Pada Data Twitteer Menggunakan

Model BERT” dilakukan beberapa tahapan untuk penelitian jurnal ini, yaitu berawal dari koleksi data, pelabelan data, data preprocessing, kemudian ada analisis sentimen dengan BERT, dan yang terakhir adalah evaluasi. Model BERT efektif untuk penelitian ini karena akurasi prediksi yang di peroleh mencapai 81% pada data uji, hal ini selisih 6 % dengan model SVM (Muhammad Adrinta Abdurrazzaq, 2022). Penelitian yang dilakukan pada jurnal milik Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya dengan judul penelitiannya “Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma BERT” mendapatkan hasil 0,81 dan jika di persentasekan yaitu 81%. Hasil ini adalah hasil dari tingkat akurasi pada analisis sentimen di penelitian ini (Putu et al., 2023).

3 METODE PENELITIAN

Langkah pertama dalam menjalankan penelitian ini adalah melakukan studi pendahuluan yang bertujuan untuk menjadi dasar dalam menyusun penelitian. Langkah berikutnya adalah mengumpulkan data yang dibutuhkan untuk penelitian ini. Tahap ketiga melibatkan pra-pemrosesan data agar siap digunakan. Tahap terakhir adalah training model, tujuannya untuk menghasilkan model yang mampu memahami dan mengekstraksi pola kompleks dalam teks, kemudian menggunakannya untuk memprediksi kategori atau label yang sesuai.



Gambar 1. Diagram Penelitian

3.1 Studi Pendahuluan

Studi Pendahuluan dilakukan dengan studi literatur yaitu mengumpulkan data dan informasi dari berbagai sumber seperti jurnal, buku, internet, dan penelitian terdahulu yang relevan dengan topik yang diteliti. Studi awal ini menjadi fondasi pengetahuan yang diperlukan untuk menyelesaikan penelitian, sehingga peneliti dapat memanfaatkan berbagai informasi dan pemikiran yang diperoleh. Oleh karena itu, berharap dapat berguna dalam melakukan analisis dan siasat dalam proses penelitian

3.2 Pengumpulan Data

Pada penelitian ini data yang akan digunakan adalah dataset *WELFake* yang diunduh pada situs Kaggle yang didapatkan pada *link* berikut <https://www.kaggle.com/code/pkbpkb0055/96-15-news-classification-with-bert-and-tf-idf/notebook>. Data tersebut (*WELFake*) merupakan kumpulan data 72.134 artikel berita dengan 35.028 berita asli dan 37.106 berita palsu berdasarkan data berita populer (yaitu Kaggle, McIntire, Reuters, BuzzFeed Political) untuk mencegah penyesuaian pengklasifikasi yang berlebihan dan untuk menyediakan lebih banyak data teks. Kumpulan data berisi empat kolom: Nomor seri (mulai dari 0); Judul (tentang teks judul berita); Teks (tentang isi berita); dan Label (0 = palsu dan 1 = asli). Ada 78.098 entri data dalam file csv dan hanya 72.134 entri yang diakses sesuai bingkai data (Putu et al., 2023).

Tabel 1. Contoh Data Berita Fake atau Real

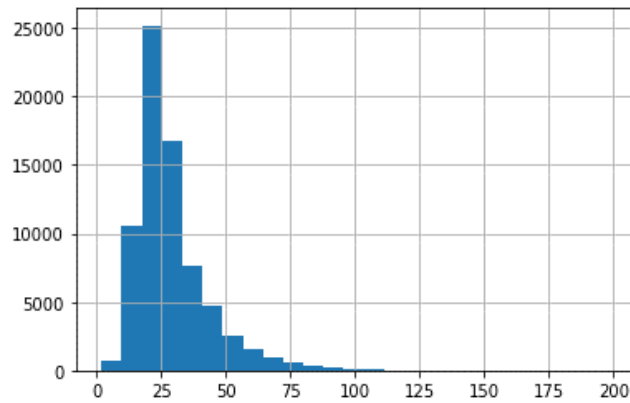
Label	Title	Test
1	LAWENFORCEMENT ON HIGH ALERT Following Threat	No comment is expected from Barack Obama Membe...
1	NaN	Did they post their votes for Hillary already?
0	Bobby Jindal, raised Hindu, uses story of Chri..	A dozen politically active pastors came here t..
1	SATAN 2: Russia unvelisan image of its..	The RS-28 Sarmat missile, dubbed Satan 2, will..

3.3 Pra-Pemrosesan

Pra-Pemrosesan atau *Preprocessing* merupakan suatu tahapan penting dalam klasifikasi data teks yang gunanya untuk pembersihan teks, penghapusan tanda baca atau karakter khusus, konversi ke huruf kecil, dan lainnya untuk mempersiapkan data teks sebelum dilakukan analisis lebih lanjut. Tahap preprocessing dilakukan dengan tujuan untuk menyortir data dari kata yang tidak relevan atau kurang tepat (Kholifatullah & Prihanto, 2023). Transformator dasar terdiri dari *encoder* untuk membaca input teks dan *decoder* untuk menghasilkan prediksi tugas. Memasukkan token pertama ke *encoder* akan mengubah menjadi *vector*, kemudian *neural network* akan memprosesnya.

Selanjutnya menambahkan token, *segmentation*, dan *position embeddings* untuk membentuk representasi input. Tokenisasi adalah proses pemotongan setiap kata menjadi token dalam kalimat untuk mempersiapkan input yang membatasi panjang teks dan mengubah token menjadi input IDs dan attention masks yang dibutuhkan. Proses *stemming* yaitu untuk penghilangan imbuhan awalan dan akhiran dari setiap kata. Berikut adalah *input code* tokenitazion.

```
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
idx = np.random.choice(range(len(texts)))
print(texts[idx])
# Tokenized
print(tokenizer.tokenize(texts[idx]))
# Token to Integral
print(tokenizer.convert_tokens_to_ids(tokenizer.tokenize(texts[idx])))
```



Gambar 2. Hasil Tokenisasi

3.4 Training Model



Gambar 3. Proses Training

Pada Gambar 3 terdapat 57.707 sampel dalam dataset pelatihan dan setiap 3.200 batch status pelatihan dicetak. Setelah satu epoch selesai maka output akan mencakup informasi tentang epoch tersebut, termasuk dengan nilai rata-rata dari total loss selama satu epoch. Contohnya yaitu Epoch: 1, Loss: 0,1824 artinya model telah menyelesaikan satu putaran lengkap melalui seluruh dataset pelatihan, nilai rata-rata dari fungsi kerugian (loss) selama satu epoch yaitu 0,1824.

4 HASIL DAN PEMBAHASAN

Dokumen yang telah diubah menjadi *word vector* selanjutnya akan dihitung menggunakan rumus TF-IDF, dengan menggunakan rumus ini maka akan menghasilkan *word vector*. TF atau *Term Frequency* itu sendiri adalah banyaknya frekuensi kemunculan kata dari suatu *term* dalam dokumen bersangkutan, sedangkan IDF atau *Inverse Document Frequency* adalah perhitungan dari bagaimana term disebar atau didistribusikan secara luas dalam koleksi dokumen yang bersangkutan. Proses perhitungan bobot kata dilakukan dengan proses awal menghitung TF atau *Term Frequency* terlebih dahulu. Berikut ini adalah *input code* dalam prosesnya.

```

port = PorterStemmer()
def stemming(text):
    stem_text = re.sub('[^a-zA-Z]', '', text)
    stem_text = stem_text.lower()
    stem_text = stem_text.split()
    stem_text = [port.stem(word) for word in stem_text if not word in stopwords.words('english')]
    stem_text = ' '.join(stem_text)
    return stem_text

```

Tabel 2. Hasil Stemming

Label	Title	Test	Refined
1	LAWENFORCEMENT ON HIGH ALERT Following Threat	No comment is expected from Barack Obama Membe...	said
1		Did they post their votes for Hillary already?	Already hillari post vote
0	Bobby Jindal, raised Hindu, uses story of Chri..	A dozen politically active pastors came here t..	jindal
1	SATAN 2: Russia unvelisan image of its..	The RS-28 Sarmat missile, dubbed Satan 2, will..	Design dub km mile missil rs sarmat satan weapon

Pengujian model dilakukan dengan cara mendemokan program yang tujuannya untuk menguji model klasifikasi berita pada dataset. Dilakukan input iterasi sebanyak 5 kali, dimana setiap iterasinya memilih satu data berita secara acak, mencetak judul dan teks, melakukan prediksi dengan model, dan menampilkan hasil prediksi bersama dengan kategori sebenarnya.

Tahap selanjutnya adalah melakukan analisis pada *pre-training* BERT. Model yang sudah didapatkan lalu dilakukan *fine-tuning* dengan penambahan layer untuk proses klasifikasi (Saputra, 2022). Tahapannya sebagai berikut:

1. Input IDs dan attention mask yang didapat setelah proses tokenisasi
2. Input IDs dan attention mask yang diproses pada BERT layer. Output berupa vector
3. Vektor lalu diproses pada neural network menggunakan feed forward fully connected layer
4. Hasilnya yang berupa bobot selanjutnya dilakukan proses klasifikasi.
5. Hasil proses klasifikasi yaitu label data

Model evaluasi yang digunakan yaitu Confusion matrix. Confusion matrix untuk mengetahui keakuratan proses pengklasifikasian teks berita *hoax*. Confusion matrix memiliki nilai atau persamaan untuk mendapatkan nilai keakuratan sistem tersebut. Maka proses dilakukan Confusion matrix sangat penting, sehingga sistem yang telah dibuat memiliki sistem yang baik untuk proses pengklasifikasian. Tujuan dilakukannya evaluasi yaitu untuk mengukur performansi sistem yang telah dibangun. Nilai akurasi yang didapatkan sebesar 0,9615 menunjukkan bahwa model klasifikasi yang dievaluasi mempunyai tingkat keakuratan sebesar 96,15%. Akurasi merupakan metrik umum yang dapat mengukur sejauh mana model dapat mengklasifikasi data dengan benar dari jumlah data yang dievaluasi.

4.1 Pengujian BERT

Pada pengujian ini dapat dilihat apakah penggunaan model klasifikasi BERT sesuai untuk klasifikasi berita palsu. Berikut adalah hasil dari uji coba dengan menggunakan BERT.

Tabel 1. Hasil Uji BERT

Judul	Prediksi	GT
STUNNER! FLORIDA TRUMP EVENT: Former Haitian Senate President Drops CLINTON BOMBSHELL Exposing Unbelievable Corruption [VIDEO]	real	real
Fact: Superdelegates Have Never Determined The Democratic Nominee	real	real
Trump's attacks could leave him friendless if impeachment comes	fake	fake
BLACK WOMAN IN CHARLESTON WARNS "There's gonna be a race war against 'Cracka's"	real	real
Spain reluctantly forced to act in Catalan vote, official says	real	fake

Pada Tabel 3 menjelaskan tentang hasil pengujian klasifikasi BERT. Hasil menunjukkan terdapat 3 berita yang mengandung pesan "Predict: real" dan "GT: real". Pesan tersebut menunjukkan hasil dari prediksi dan ground truth (kategori sebenarnya) untuk suatu sampel data berita. Dalam artian:

- "Predict: real" artinya model klasifikasi telah melakukan prediksi bahwa kategori atau label yang paling mungkin untuk sampel tersebut adalah "real"
- "GT: real" artinya kategori sebenarnya (ground truth) dari sampel data berita tersebut adalah "real".

Jika demikian maka prediksi model cocok dengan kategori sebenarnya dari data tersebut, yang menunjukkan bahwa model memberikan hasil yang benar untuk sampel tertentu. Selain 3 berita tersebut juga ditemukan 1 berita yang mengandung pesan "Predict: fake" dan "GT: fake". Dalam arti:

- "Predict: fake" artinya model klasifikasi memprediksi bahwa contoh data tersebut termasuk dalam kategori "fake" berdasarkan fitur atau informasi yang dipelajari selama pelatihan.
- "GT: fake" dari berita tersebut adalah "fake".

Jadi dalam berita ini prediksi model "fake" sesuai dengan GT data tersebut yaitu "fake". Ini merupakan hasil yang diinginkan dalam tugas klasifikasi, hal ini menunjukkan bahwa model memberikan prediksi yang benar untuk berita tersebut.

Terdapat juga 1 berita dengan pesan "Predict: real" dan "GT: fake". Artinya:

- "Predict: real" artinya model klasifikasi memprediksi bahwa berita tersebut masuk ke dalam kategori "real".
- "GT: fake" artinya berita tersebut adalah "fake".

Dalam berita ini model memberikan prediksi yang tidak benar (prediksi "real") karena GT (kategori sebenarnya) dari berita tersebut adalah "fake". Hal ini menunjukkan bahwa model perlu diperbaiki atau disesuaikan untuk meningkatkan performa pada jenis data tertentu.

5 SIMPULAN

Penelitian ini menggunakan model klasifikasi BERT yang dapat meningkatkan kemampuan sistem untuk mengidentifikasi dan mengklasifikasi berita palsu, mendeteksi dan mengatasi penyebaran berita palsu. Keuntungan dari menggunakan model BERT yaitu kemampuannya dalam

transfer learning. Model BERT juga memberikan tingkat akurasi yang tinggi pada klasifikasi berita palsu karena kemampuannya dalam memahami konteks dan bahasa yang kompleks. Model dapat mengenali kata-kata yang sering diubah yang bertujuan untuk menyembunyikan informasi. Evaluasi harus dilakukan secara terus menerus dan pemeliharaan model perlu dilakukan untuk memastikan kinerja model tetap optimal.

DAFTAR PUSTAKA

- Sholikhah, I. I., Harjanta, A. T. J., & Latifah, K. (2023). Machine Learning Untuk Deteksi Berita Hoax Menggunakan BERT. *Prosiding Seminar Nasional Informatika*, 1(1), 524–531.
- Kasanah, A. N., Muladi, M., & Pujianto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 196–201. <https://doi.org/10.29207/resti.v3i2.945>.
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 269–276. <https://doi.org/10.25126/jtiik.201853810>.
- Tandijaya, J. H., Liliana, & Sugiarto, I. (2021). Klasifikasi dalam Pembuatan Portal Berita Online dengan Menggunakan Metode BERT. *Jurnal Infra*, 09(02), 320–325.
- Ramadhan, Nur Ghaniaviyanto, et al. (2022). Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 251–256.
- Rahadi, D. R. (2017). Perilaku Pengguna Dan Informasi Hoax Di Media Sosial. *Jurnal Manajemen Dan Kewirausahaan*, 5(1), 58–70. <https://doi.org/10.26905/jmdk.v5i1.1342>.
- Oshikawa, R., Qian, J., & Wang, W. Y. (2020). A survey on natural language processing for fake news detection. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, 6086–6093.
- Muzakir, A., & Suriani, U. (2023). Model Deteksi Berita Palsu Menggunakan Pendekatan Bidirectional Long Short-Term Memory (Bilstm). *Journal of Computer and Information Systems Ampera*, 4(2), 93–105.
- Yunanto, R., Purfini, A. P., & Prabuwisesa, A. (2021). Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning. *Jurnal Manajemen Informatika (JAMIKA)*, 11(2), 118–130. <https://doi.org/10.34010/jamika.v11i2.5362>.
- Sains, F., Teknologi, D., Islam, U., & Sunan, N. (2023). *PEMODELAN TOPIK DALAM AL-QUR'AN MENGGUNAKAN LIBRARY*. 14(2), 319–327.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- Halim, F., Liliana, & Gunadi, K. (2022). Ringkasan Ekstraktif Otomatis pada Berita Berbahasa Indonesia Menggunakan Metode BERT. *Jurnal Infra*, 10(1), 162–168.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 5999–6009.
- Mas, R., Panca, R. W., Atmaja1, K., & Yustanti2, W. (2021). Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *Jeisbi*, 02(3), 55–62.
- Ardiansyah, Adika Sri Widagdo, Krisna Nuresa Qodri, Saputro, F. E. N., & Nisrina Akbar Rizky P. (2023). Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT. *Jurnal Fasilkom*, 13(02), 326–333. <https://doi.org/10.37859/jf.v13i02.5170>.

- Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3730–3740. <https://doi.org/10.18653/v1/d19-1387>.
- Kusnadi, R., Yusuf, Y., Andriantony, A., Ardian Yaputra, R., & Caintan, M. (2021). Analisis Sentimen Terhadap Game Genshin Impact Menggunakan Bert. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 6(2), 122–129. <https://doi.org/10.36341/rabit.v6i2.1765>.
- Fajri, F., Tutuko, B., & Sukemi, S. (2022). Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter. *JUSIFO (Jurnal Sistem Informasi)*, 8(2), 71–80. <https://doi.org/10.19109/jusifo.v8i2.13885>.
- Muhammad Adrinta Abdurrazzaq, E. L. T. (2022). Analisis Sentimen KUHP Baru Pada Data Twitter Menggunakan Model BERT. *Jurnal Komunikasi, Sains Dan Teknologi*, 1(2), 38–43.
- Putu, N., Saraswati, V. D., Yudistira, N., & Adikara, P. P. (2023). Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(2), 909–916.
- Ramadhan, Nur Ghaniaviyanto. (2021). Indonesian online news topics classification using Word2Vec and K-nearest neighbor. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1083-1089.
- Saputra, R. (2022). Implementasi Bidirectional Encoder Representations From Transformers (BERT) Untuk Mendeteksi Hatespeech Dan Abusive Language Pada Twitter Bahasa Indonesia. *Uin Suska Riau*.