BAB 2
LANDASAN TEORI

2.1 Konflik Terjadi Perang Antara Rusia dan Ukraina

Perang antara Rusia dan Ukraina yang dimulai pada awal tahun 2022 memberikan dampak yang signifikan tidak hanya bagi kedua negara yang terlibat secara langsung, tetapi juga terhadap stabilitas geopolitik dan perekonomian global. Konflik tersebut memicu ketidakpastian dalam berbagai aktivitas ekonomi internasional, mengingat Rusia dan Ukraina merupakan aktor penting dalam produksi dan distribusi energi serta komoditas strategis dunia. Gangguan pada sektor perdagangan, energi, dan rantai pasok global menyebabkan ketidakstabilan pasar internasional serta memperlambat proses pemulihan ekonomi global pascapandemi COVID-19 [10]. Selain itu, penerapan sanksi ekonomi terhadap Rusia dan terganggunya distribusi komoditas utama turut mendorong terjadinya fluktuasi harga minyak, gandum, dan energi di pasar global, yang pada akhirnya meningkatkan tekanan inflasi serta memengaruhi stabilitas kebijakan fiskal di berbagai negara, termasuk Indonesia [11, 12].

Dari perspektif sektor pertanian dan ketahanan pangan, konflik ini mengakibatkan terganggunya distribusi komoditas utama seperti gandum, jagung, dan minyak bunga matahari yang selama ini dipasok Ukraina ke berbagai negara. Hambatan logistik dan penutupan akses pelabuhan selama berlangsungnya konflik menyebabkan terjadinya gangguan rantai pasok pangan global [13]. Kondisi tersebut berimplikasi pada peningkatan harga komoditas pangan dunia serta menimbulkan ancaman terhadap ketahanan pangan, khususnya bagi negara-negara yang memiliki ketergantungan tinggi terhadap impor dari kawasan tersebut [13]. Hal ini menunjukkan bahwa konflik geopolitik memiliki dampak multidimensional yang tidak hanya terbatas pada aspek militer dan politik, tetapi juga berpengaruh secara langsung terhadap stabilitas ekonomi global dan kesejahteraan masyarakat internasional.

2.2 *YouTube*

YouTube merupakan salah satu platform berbagi video terbesar di dunia yang menyediakan layanan distribusi dan konsumsi konten audiovisual secara daring. Platform ini memungkinkan pengguna untuk mengakses informasi dalam bentuk

video tanpa harus membaca artikel tertulis [14]. Di Indonesia, jumlah kunjungan terhadap situs *YouTube* mencapai sekitar 856,7 juta kunjungan. Tingginya tingkat kunjungan tersebut menunjukkan bahwa *YouTube* menjadi salah satu media digital yang banyak digunakan oleh masyarakat untuk mengakses informasi dan hiburan. Selain sebagai media untuk menonton video, *YouTube* juga menyediakan fitur kolom komentar yang memungkinkan pengguna untuk berinteraksi, menyampaikan pendapat, serta mendiskusikan isi konten yang mereka tonton [15].

Dalam konteks penelitian, khususnya pada bidang analisis sentimen, *YouTube* menjadi salah satu sumber data yang relevan untuk mengidentifikasi opini publik terhadap suatu isu tertentu. Komentar yang dituliskan oleh pengguna pada kolom komentar video dapat merepresentasikan pandangan, persepsi, maupun sikap masyarakat terhadap topik yang sedang dibahas. Oleh karena itu, data komentar pada *YouTube* sering dimanfaatkan sebagai sumber data dalam penelitian analisis sentimen untuk mengungkap kecenderungan opini masyarakat terhadap suatu peristiwa atau fenomena tertentu [16].

2.3 Analisis Sentimen

Analisis sentimen adalah teknik dalam bidang *Natural Language Processing* (NLP) yang digunakan untuk mengekstraksi dan mengidentifikasi informasi subjektif dari data teks. Teknik ini bertujuan untuk mengelompokkan opini atau tanggapan yang disampaikan oleh pengguna ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral [17]. Pada platform media sosial, analisis sentimen dimanfaatkan untuk mengkaji berbagai komentar yang dihasilkan pengguna sehingga dapat diperoleh pemahaman mengenai kecenderungan opini publik terhadap suatu topik yang sedang dibahas [18].

Dalam konteks penelitian, analisis sentimen mengalami perkembangan yang pesat seiring dengan meluasnya penggunaan web dan media sosial sebagai sumber utama data opini publik. Platform seperti *YouTube* menyediakan data dalam jumlah besar yang bersifat *real-time*, sehingga memungkinkan peneliti untuk menganalisis persepsi masyarakat terhadap berbagai isu, seperti politik, ekonomi, hingga teknologi [19]. Hal ini menjadikan analisis sentimen sebagai salah satu metode yang penting dalam pengambilan keputusan berbasis data serta pemahaman perilaku sosial masyarakat [20].

2.4 *Lexicon Labelling*

Metode *Lexicon* merupakan cara menganalisis sentimen pada media sosial dengan menggunakan kamus *Inset Lexicon* yang berisi kumpulan kata positif dan negatif dalam bahasa Indonesia untuk menentukan kategori sentimen pada setiap komentar berdasarkan skor sentimen yang dihasilkan [21]. Setelah perhitungan skor masing masing kata, maka selanjutnya seluruh skor kata dalam satu kalimat komentar dijumlahkan untuk menghasilkan nilai polarisasi yang menunjukkan kecenderungan sentimen komentar tersebut.

Rumus skor polaritas adalah sebagai berikut:

$$S = \sum_{i=1}^{n} (w_i) \tag{2.1}$$

Dengan:

1. S = Skor Sentimen Total (*Polarization Score*).
2. w_i = Kata ke-i dalam dokumen.
3. n = jumlah kata dalam komentar.

Berdasarkan hasil nilai polarisasi yang sudah ditentukan, penentuan label dilakukan dengan aturan sebagai berikut:

1. Jika $S > 0$, maka diklasifikasikan sebagai sentimen Positif [22].
2. Jika $S < 0$, maka diklasifikasikan sebagai sentimen Negatif [22].
3. Jika $S = 0$, maka diklasifikasikan sebagai sentimen Netral [22].

2.5 BERT

Bidirectional Encoder Representations from Transformers (BERT) adalah model berbasis arsitektur *transformer* yang dikembangkan untuk memahami representasi bahasa secara dua arah (*bidirectional*). Melalui pendekatan ini, model dapat menangkap makna suatu kata dengan mempertimbangkan konteks sebelum dan sesudahnya secara bersamaan pada tahap *pre-training* [18]. Arsitektur *transformer* yang mendasari BERT memungkinkan pemodelan hubungan antar kata

dalam teks secara efektif, sehingga model mampu memahami informasi kontekstual secara lebih mendalam [8, 23].

Adapun terdapat 2 komponen arsitektur *transformer* yang terdiri dari *encoder* dan *decoder*, yang masing-masing memiliki peran sebagai berikut:

1. *Encoder*

   Komponen *encoder* berfungsi untuk memproses dan memahami keseluruhan informasi yang terkandung dalam teks masukan [23]. Arsitektur *encoder* tersusun atas beberapa lapisan identik yang disusun secara berurutan, di mana setiap lapisan terdiri dari dua sublapisan utama, yaitu *self-attention* dan *feed-forward neural network* [23, 24]. Mekanisme *self-attention* memungkinkan model untuk mengidentifikasi hubungan antar kata dalam suatu kalimat dengan memberikan bobot perhatian berdasarkan tingkat relevansi masing-masing kata terhadap kata lainnya, sehingga konteks semantik dapat dipahami secara lebih mendalam. Adapun persamaan yang digunakan pada mekanisme *self-attention* ditunjukkan sebagai berikut [25].

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{2.2}$$

   Dimana:

   (a) Q: *Query*

   (b) K: *Key*

   (c) V: *Value*

   (d) dk: Dimensi dari *key*

   (e) T: Transposisi Matriks

   Dengan mekanisme ini, setiap posisi dalam *encoder* dapat mengakses informasi dari seluruh posisi lainnya, baik sebelum maupun sesudahnya [23].

2. *Decoder*

   Komponen *decoder* bertugas untuk menghasilkan urutan keluaran berdasarkan representasi yang telah dihasilkan oleh *encoder* [23]. Struktur *decoder* tersusun atas beberapa lapisan yang serupa dengan *encoder*, namun dilengkapi dengan lapisan *attention* tambahan yang berfungsi untuk menghubungkan informasi dari *encoder* ke dalam proses pembentukan

keluaran melalui mekanisme *multi-head attention*. Persamaan rumus pada mekanisme *multi-head attention* ditunjukkan pada berikut ini [25].

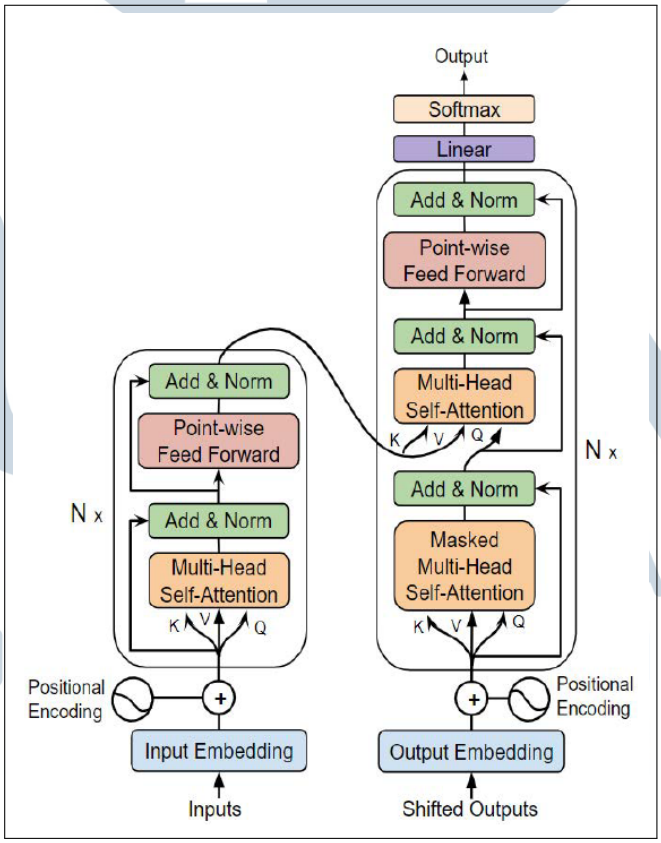
$$MultiHead(Q,K,V) = Concat(head_1,....,head_h)W^O \tag{2.3}$$

Dengan:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{2.4}$$

Dimana: W_i^Q, W_i^K, W_i^V, dan W^O adalah matriks bobot ang akan disesuaikan atau diperbarui nilainya selama proses *fine-tuning* [25].

Dengan demikian, *decoder* dapat memfokuskan perhatian pada informasi penting dari hasil pemrosesan *encoder* dalam menghasilkan prediksi keluaran [23, 24].



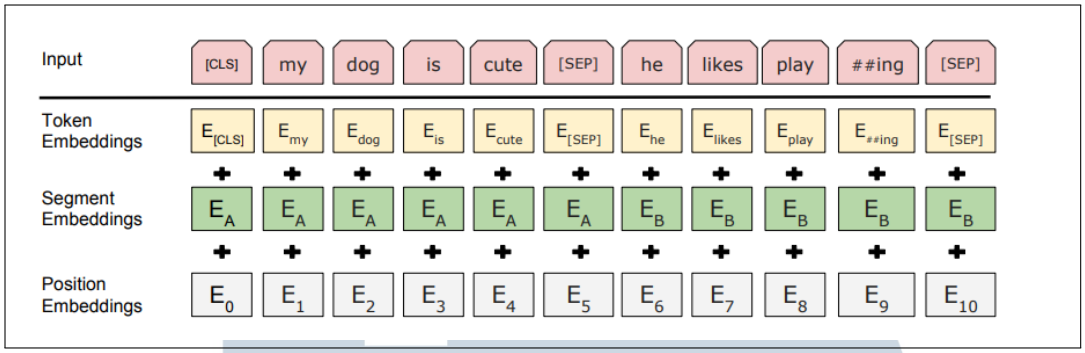
Gambar 2.1. Arsitektur model *Transformer*

Berdasarkan ilustrasi pada Gambar 2.1, alur kerja arsitektur *transformer* dapat dijelaskan sebagai berikut:

1. Setiap kata pada input teks terlebih dahulu direpresentasikan dalam bentuk vektor melalui proses *embedding* [8].
2. Vektor input kemudian diproses melalui dua sublapisan utama pada setiap *encoder*, yaitu *self-attention* dan *feed-forward neural network*, untuk menangkap hubungan antar kata dalam konteks kalimat [8].
3. Setelah proses pada *encoder* selesai, keluaran berupa representasi vektor (termasuk *key* dan *value*) diteruskan ke *decoder* sebagai dasar dalam menghasilkan output [8].

Berdasarkan arsitektur tersebut, model *BERT* hanya memanfaatkan bagian *encoder* dalam proses transformer [23]. Proses awal dimulai dengan mengubah setiap kalimat teks menjadi token dengan menggunakan BERT tokenizer [26]. Pada proses tersebut ditambahkan token `CLS` di awal kalimat yang berfungsi sebagai representasi keseluruhan kalimat untuk tugas klasifikasi, serta token `[SEP]` yang berfungsi sebagai penanda pemisah antar kalimat [23, 27]. Selanjutnya, setiap token diubah menjadi representasi vektor melalui proses *embedding* yang terdiri dari tiga komponen, yaitu sebagai berikut.

1. *Token Embedding*
   Token Embedding berfungsi sebagai representasi bentuk vektor di setiap token dalam kosa kata tertentu [23, 26].
2. *Segment Embeddings*
   Segment embeddings digunakan untuk membedakan bagian-bagian kalimat dalam satu input, khususnya pada kasus yang melibatkan dua kalimat, seperti kalimat pertama dan kalimat kedua [23, 26].
3. *Position Embeddings*
   Position embeddings berfungsi untuk memberikan informasi mengenai posisi setiap token dalam urutan kalimat. Hal ini diperlukan karena arsitektur *transformer* tidak memiliki mekanisme bawaan untuk memahami urutan kata [23, 26].



Gambar 2.2. Proses Representasi input pada model BERT

Setelah setiap kata direpresentasikan dalam bentuk vektor melalui proses *embedding*, tahapan selanjutnya adalah menuju ke tahap *training* yang terdiri dari *pre-training* dan *fine tuning*. Pada tahap *pre-training*, model BERT dilakukan secara bersamaan menggunakan dua tugas, yaitu *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP) [23]. *Masked Language Modeling* (MLM) dilakukan dengan cara menyembunyikan sekitar 15% token dari setiap kata diganti dengan token `[MASK]` untuk memprediksi nilai asli kata berdasarkan kata-kata lain yang tidak dipilih oleh token `[MASK]` di setiap kalimat [23, 28].

Selain itu, ada *Next Sentence Prediction* (NSP) yang tujuannya melatih untuk memprediksi apakah kalimat kedua dalam pasangan kalimat adalah kalimat lanjutan dalam dokumen aslinya [23, 28]. Dalam proses pelatihan, 50% dari input merupakan pasangan kalimat, di mana kalimat ke dua merupakan kalimat selanjutnya dari dokumen asli. Sisanya 50% lainnya merupakan kalimat yang diambil secara acak [23].

Setelah melakukan proses pre-training, tahap selanjutnya adalah fine-tuning yang bertujuan untuk menyesuaikan model agar dapat menyelesaikan tugas spesifik menggunakan data berlabel. Pada tahap ini, model dilatih kembali menggunakan pendekatan supervised learning sehingga parameter pada model BERT dapat diperbarui sesuai dengan karakteristik tugas yang dihadapi. Dalam tugas klasifikasi teks, representasi dari token `[CLS]` yang dihasilkan pada layer terakhir digunakan sebagai representasi keseluruhan kalimat, yang kemudian dihubungkan dengan lapisan klasifikasi untuk menghasilkan prediksi kelas [29]. Lapisan klasifikasi yang digunakan yaitu fungsi aktivasi *softmax* yang dirumuskan sebagai berikut [30]:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2.5}$$

Dimana:

1. $P(y_i)$ = Probabilitas kelas ke-i
2. z_i = Nilai logits
3. K = jumlah kelas

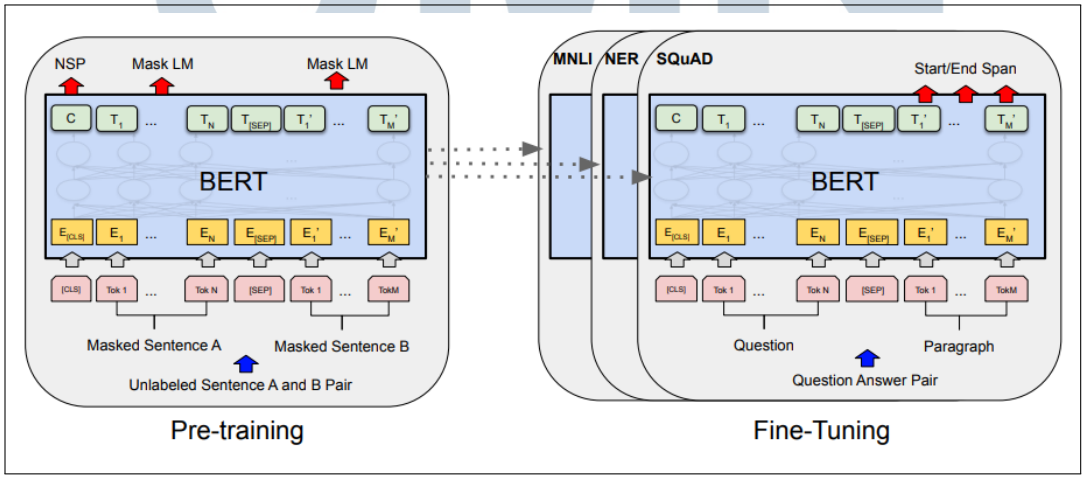
Untuk mengukur perbedaan antara distribusi probabilitas hasil prediksi model dengan label sebenarnya, digunakan rumus *Cross Entropy Loss* yang dirumuskan sebagai berikut [30]:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{2.6}$$

Dimana:

1. L = Nilai loss
2. C = Jumlah kelas
3. y_i = Label sebenarnya
4. $\hat{y}_i$ = Probabilitas hasil prediksi model

Dengan dilakukan proses *fine-tuning*, model BERT dapat menyesuaikan representasi bahasa yang telah dipelajari pada tahap *pre-training* agar lebih sesuai dengan karakteristik data dan tugas spesifik yang dikerjakan.



Gambar 2.3. Proses Pre-Training dan Fine-Tuning dalam model BERT

2.6 IndoBERT

IndoBERT adalah model bahasa berbasis arsitektur BERT yang dirancang khusus untuk memproses teks berbahasa Indonesia. Model ini dilatih menggunakan korpus bahasa Indonesia yang berasal dari berbagai sumber, seperti portal berita daring, situs web, dan dokumen teks lainnya, sehingga mampu memahami struktur bahasa, konteks, dan makna semantik teks berbahasa Indonesia dengan lebih baik [31]. Dalam penggunaannya, IndoBERT umumnya diterapkan melalui proses *fine-tuning*, di mana teks diubah menjadi token sub-kata (*subword*), diproses oleh model, dan selanjutnya digunakan untuk menghasilkan prediksi sesuai dengan tugas yang diinginkan [9].

2.7 Evaluasi

Evaluasi dilakukan menggunakan empat metrik yang terdiri dari *Precision*, *Recall*, *F1-Score*, dan *Accuracy* yang bertujuan untuk mengukur performa model. Metrik tersebut dititung dari nilai yang dihasilkan di *Confusion Matrix* yang terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), serta *False Negative* (FN) [32].

Accuracy adalah ukuran seberapa banyak prediksi model yang benar dibandingkan dengan seluruh data yang diuji.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.7}$$

Precision menunjukkan seberapa banyak prediksi positif yang benar dari semua data yang diprediksi positif oleh model.

$$Precision = \frac{TP}{TP + FP} \tag{2.8}$$

Recall mengukur seberapa banyak data positif yang berhasil ditemukan oleh model dari seluruh data positif yang sebenarnya.

$$Recall = \frac{TP}{TP + FN} \tag{2.9}$$

F1-Score adalah rata-rata harmonis antara *Precision* dan *Recall* yang digunakan untuk menyeimbangkan keduanya.

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{2.10}$$

2.8 Studi Literatur

Terdapat beberapa studi literatur yang membahas analisis sentimen terkait konflik Rusia dan Ukraina, penggunaan model BERT dan IndoBERT, serta penerapan analisis sentimen pada komentar di berbagai platform media sosial dengan beragam permasalahan dan pendekatan. Tinjauan terhadap beberapa studi literatur disajikan pada tabel berikut.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 2.1. Perbandingan Studi Terkait model BERT

Tahun	Penulis	Dataset	Model	Domain	Akurasi	Rangkuman penelitian
2022	Julianto, M. F., Malau, Y., & Hidayat, W. F.	Tweet perang RusiaUkraina	BERT	Konflik Rusia & Ukraina	97%	Analisis sentimen opini Twitter masyarakat terkait perang RusiaUkraina menggunakan model BERT.
2023	Hasan, M	Komentar sosial media RusiaUkraina (Bangla)	Transformers (BanglaBERT)	Sentimen RUUK Bangla	86%	NLP dan sentiment analysis pada komentar Bangla tentang perang RusiaUkraina menggunakan model transformer.
2023	Vidya Chandradev, I. M. A. D. S., & I. P. A. Bayupati	Review hotel	BERT	Sentimen review hotel Indonesia	91%	Analisis sentimen review hotel Indonesia menggunakan model BERT.
2023	Putri, N. A. R. & Ardiansyah	Teks kemajuan kecerdasan buatan di Indonesia	BERT + RoBERTa	Sentimen tentang AI di Indonesia	83%-84%	Analisis sentimen kemajuan AI di Indonesia menggunakan BERT dan RoBERTa.

Tahun	Penulis	Dataset	Model	Domain	Akurasi	Rangkuman penelitian
2023	Manthena, S. P.	Twitter bencana	BERT - BiLSTM - CNN	Respon Bencana (Disaster response)	90-93%	Penelitian memanfaatkan tweet untuk respons bencana cepat menggunakan model BERT-BiLSTM-CNN.
2024	Kumar, B. V. P., & Sadanandam, M.	Social media platforms	BERT + RoBERTa	Sentimen umum social media	94,3%	Penelitian bertujuan meningkatkan performa analisis sentimen social media dengan arsitektur fusi BERT + RoBERTa.
2024	Khadapi, M. & Pakpahan, V. M.	Diskusi Twitter tentang Pemilu 2024	BERT + LSTM	Sentimen politik Pemilu 2024	LSTM 85% dan BERT 64,53%	Analisis sentimen terhadap Pemilu 2024 menggunakan BERT dan LSTM.
2024	Pavita, R. & Hasan, F. N.	Data hate speech calon wakil presiden Indonesia	BERT	Hate speech politik Indonesia	98,17%	Analisis sentimen ujaran kebencian terkait calon wakil presiden Indonesia menggunakan model BERT.
2025	Ningrum, D. Y. A. Ayu; Daniati, E.; Muzaki, M. N.	Ulasan Aplikasi BRI Mobile	BERT vs RNN-LSTM	Sentimen aplikasi perbankan	BERT 73%, LSTM 69%	Model BERT dan RNN-LSTM pada sentimen BRI Mobile.

Tahun	Penulis	Dataset	Model	Domain	Akurasi	Rangkuman penelitian
2026	Yuswira, J. D. & Ginting, J. A.	Komentar X/Twitter tentang isu makanan bergizi gratis	IndoBERT + SNA	Sentimen & social network analysis	75%	Analisis sentimen dan jaringan sosial terhadap isu MBG di X/Twitter dengan model IndoBERT.

Dari beberapa studi literatur yang tertera, terbukti bahwa BERT menunjukkan performa yang sangat baik dalam berbagai tugas analisis sentimen. Pada isu konflik Rusia dan Ukraina, model ini secara konsisten mampu mencapai akurasi sekitar 86% hingga 97% melalui analisis data dari berbagai platform media sosial [8, 30]. Hasil tersebut menunjukkan bahwa BERT dan variannya memiliki kemampuan dalam memahami konteks bahasa secara mendalam, bahkan pada topik global dan multibahasa.

Selain itu, penerapan model ini pada berbagai domain, seperti ulasan hotel, isu kecerdasan buatan, serta respons terhadap bencana alam, juga menunjukkan hasil akurasi di antara 83% sampai dengan 93% [33, 17, 26]. Pengembangan lebih lanjut dengan mengombinasikan BERT dengan model lain seperti BiLSTM, CNN, maupun RoBERTa terbukti mampu meningkatkan performa analisis sentimen [34, 35]. Hal ini menunjukkan bahwa integrasi beberapa arsitektur dapat memperkuat kemampuan model dalam menangkap pola kompleks pada data teks.

Pada konteks yang lebih spesifik, seperti analisis sentimen politik dan hate speech, model BERT tetap menunjukkan performa yang kompetitif [36]. Selain itu, BERT juga terbukti lebih unggul dibandingkan metode tradisional seperti RNN-LSTM dalam beberapa studi perbandingan [18]. Penggunaan IndoBERT dalam analisis sentimen media sosial juga menunjukkan hasil yang baik [9].

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA