Buletin Sistem Informasi dan Teknologi Islam
Vol 7, No 3 (2026), pp 265-276
ISSN: 2721-0901
DOI: https://dx.doi.org/10.33096/busiti.v7i3.3367

Perbandingan Model IndoBERT dan Bi-LSTM untuk Analisis Sentimen Publik di Platform X terhadap Isu Bergabungnya Indonesia dalam *Board of Peace*

Comparative Study of IndoBERT and Bi-LSTM Models for Public Sentiment Analysis on Platform X Regarding the Issue of Indonesia Joining the Board of Peace

Rahmah Muthmainah[a,1,*]

[a]*Jurusan Ilmu Komputer, Universitas Riau, Pekanbaru, Indonesia*
[1]*rahmah.muthmainah3281@student.unri.ac.id*
**corresponding author*

Informasi Artikel

Diserahkan : 28 Februari 2026
Diterima : 17 Mei 2026
Direvisi : 4 Juni 2026
Diterbitkan : 26 Agustus 2026

Kata Kunci:
Analisis Sentimen
Bi-LSTM
IndoBERT
CRISP-DM
Platform X

ABSTRAK

Perkembangan diplomasi digital dan meningkatnya diskursus publik di media sosial menjadikan analisis sentimen sebagai metode yang penting untuk memetakan opini masyarakat terhadap kebijakan luar negeri Indonesia. Salah satu isu yang menarik perhatian publik adalah keputusan Indonesia bergabung dalam *Board of Peace*. Namun, penelitian yang membandingkan kinerja IndoBERT dan Bi-LSTM pada isu geopolitik di media sosial Indonesia masih terbatas. Penelitian ini membandingkan kedua model dalam menganalisis sentimen publik di Platform X terkait isu tersebut menggunakan kerangka CRISP-DM yang meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, dan evaluasi. *Dataset* terdiri atas 1.627 unggahan yang diperoleh melalui penyaringan 2.314 *tweet* melalui teknik *scraping* berbasis kata kunci dan dilabeli secara manual ke dalam tiga kelas sentimen, yaitu 98 Negatif, 1.383 Netral, dan 146 Positif. Ketidakseimbangan kelas ditangani dengan SMOTE, sedangkan evaluasi dilakukan pada 245 sampel uji menggunakan akurasi, presisi, *recall*, dan *F1-score*. Hasil menunjukkan IndoBERT mencapai akurasi 95,51% dan *weighted F1-score* 0,9514, lebih tinggi dibandingkan Bi-LSTM dengan akurasi 94,29% dan *F1-score* 0,9409. IndoBERT unggul pada klasifikasi kelas minoritas, sedangkan Bi-LSTM lebih efisien secara komputasi. Temuan ini menunjukkan bahwa model berbasis *transformer* lebih efektif dalam menangkap konteks semantik wacana geopolitik digital untuk analisis sentimen kebijakan luar negeri Indonesia.

***Keywords*:**
Sentiment Analysis
Bi-LSTM
IndoBERT
CRISP-DM
X Platform

ABSTRACT

The development of digital diplomacy and the increasing volume of public discourse on social media have made sentiment analysis an important method for mapping public opinion on Indonesia's foreign policy. One issue that has attracted public attention is Indonesia's decision to join the Board of Peace. However, studies directly comparing the performance of IndoBERT and Bi-LSTM on geopolitical issues in Indonesian social media remain limited. This study compares the two models in analyzing public sentiment on Platform X regarding this issue using the CRISP-DM framework, which includes business understanding, data understanding, data preparation, modeling, and evaluation. The dataset consisted of 1,627 posts filtered from 2,314 tweets collected through keyword-based scraping and manually labeled into three sentiment classes: 98 Negative, 1,383 Neutral, and 146 Positive. Class imbalance was addressed using SMOTE, while model performance was evaluated on 245 test samples using accuracy, precision, recall, and F1-score. The results show that IndoBERT achieved 95.51% accuracy and a weighted F1-score of 0.9514, outperforming Bi-LSTM, which obtained 94.29% accuracy and an F1-score of 0.9409. IndoBERT performed better in minority-class classification, whereas Bi-LSTM was more computationally efficient. These findings indicate that transformer-based models are more effective in capturing semantic context within digital geopolitical discourse for foreign policy sentiment analysis.

I. Pendahuluan

Perkembangan kecerdasan buatan, khususnya pemrosesan bahasa alami (*natural language processing*), dalam satu dekade terakhir mendorong pemanfaatan analisis sentimen untuk mengekstraksi opini publik dari media sosial secara otomatis dan berskala besar [1]. Platform global seperti X menjadi ruang utama artikulasi sikap politik, respons terhadap kebijakan, dan ekspresi emosi kolektif yang dapat dimodelkan melalui teknik pembelajaran mesin dan pembelajaran mendalam [2]. Peningkatan volume data teks yang masif dan bersifat *real-time* menjadikan analisis sentimen sebagai instrumen penting dalam memahami dinamika opini publik terhadap isu-isu politik internasional dan kebijakan luar negeri [3].

Transformasi diplomasi tradisional menuju diplomasi digital menunjukkan bahwa negara dan aktor resmi semakin mengandalkan platform digital untuk menyampaikan posisi politik, membangun citra, dan memengaruhi persepsi global [4]. Media sosial tidak hanya menjadi kanal komunikasi satu arah, melainkan arena interaksi timbal balik yang membentuk relasi baru antara negara dan publik domestik maupun internasional [5]. Dalam konteks ini, pemetaan sentimen publik atas kebijakan luar negeri menjadi krusial untuk mengelola legitimasi, risiko reputasi, dan efektivitas diplomasi publik di ranah digital [6].

Di Indonesia, penetrasi internet dan penggunaan media sosial yang tinggi menjadikan X sebagai ruang diskursif penting dalam politik domestik dan kebijakan luar negeri, termasuk isu konflik Israel–Palestina yang selama ini menjadi salah satu pilar utama politik luar negeri bebas-aktif Indonesia. Komitmen historis Indonesia terhadap kemerdekaan Palestina inilah yang secara langsung mendasari keterlibatannya dalam berbagai inisiatif perdamaian multilateral; *Board of Peace* merupakan wadah yang secara eksplisit menekankan solusi dua negara sebagai landasan perdamaian global, sehingga keputusan Indonesia bergabung di dalamnya merupakan kelanjutan logis sekaligus manifestasi konkret dari posisi geopolitik jangka panjang tersebut [7]. Penelitian mutakhir menunjukkan bahwa akun resmi pemimpin negara di X memainkan peran sentral dalam diplomasi digital Indonesia, meskipun optimalisasi pelibatan aktor tingkat menengah dan akar rumput masih terbatas [8]. Dalam konteks kebijakan luar negeri Indonesia terhadap isu Israel–Palestina, respons publik di X menjadi indikator penting yang perlu dimonitor secara sistematis melalui pendekatan komputasional, terutama ketika pemerintah mengambil langkah konkret yang mempertegas posisi tradisionalnya.

Fenomena aktual yang menjadi fokus penelitian adalah keputusan Indonesia bergabung dalam *Board of Peace*, sebuah inisiatif multilateral yang secara resmi ditandai dengan penandatanganan piagam oleh Presiden pada 22 Januari 2026. Langkah ini ditegaskan sebagai bentuk peran aktif Indonesia dalam menjaga implementasi solusi dua negara dan diplomasi perdamaian global [9]. Dengan demikian, kebijakan ini merupakan kelanjutan logis dari komitmen lama Indonesia terhadap penyelesaian konflik Israel–Palestina. Kebijakan luar negeri dengan sensitivitas tinggi seperti ini, yang secara langsung menyentuh posisi ideologis dan solidaritas historis Indonesia terhadap perjuangan Palestina, berpotensi memicu polarisasi opini di media sosial, baik dalam bentuk dukungan, kritik, maupun misinformasi. Kondisi ini menuntut kehadiran sistem analisis sentimen yang andal untuk memetakan respons publik secara terukur [10].

Secara metodologis, analisis sentimen politik di media sosial menghadapi tantangan berupa bahasa informal, sarkasme, ambiguitas, dan ketidakseimbangan kelas, sehingga menuntut model NLP yang mampu menangkap konteks dan dependensi sekuensial secara lebih mendalam [11]. Model berbasis *transformer* seperti BERT telah terbukti melampaui metode klasik maupun *word embedding* statis dalam tugas klasifikasi sentimen pada data media sosial yang kompleks [12]. Di sisi lain, arsitektur *Long Short-Term Memory* (LSTM) dan variannya, termasuk Bi-LSTM, tetap relevan karena kemampuannya memodelkan ketergantungan jangka panjang dan struktur sekuensial pada data teks [13]

Sejumlah studi menunjukkan bahwa model BERT dan turunannya mampu memberikan akurasi sangat tinggi dalam analisis sentimen dan emosi pada komentar media sosial, sementara penelitian lain melaporkan keunggulan model LSTM atau kombinasi CNN–Bi-LSTM untuk menangkap pola lokal dan temporal pada teks [14]. Di Indonesia, IndoBERT dan IndoBERTweet mulai banyak digunakan untuk tugas klasifikasi sentimen berbahasa Indonesia, namun hasil empirisnya bervariasi bergantung pada domain, kualitas data, dan skema pelabelan [15]. Penelitian lain mengimplementasikan LSTM atau Bi-LSTM pada data X berbahasa Indonesia bertopik politik, sosial, dan ekonomi dengan akurasi tinggi, tetapi belum secara sistematis dibandingkan dengan model *transformer* dalam konteks kebijakan luar negeri yang sensitif [16].

Meskipun demikian, terdapat beberapa keterbatasan penting dari penelitian terdahulu. Pertama, banyak studi analisis sentimen politik di Indonesia berfokus pada pemilu, kebijakan domestik, atau opini terhadap institusi negara, misalnya penelitian tentang kebijakan Ibu Kota Negara [17], sehingga belum menyentuh secara spesifik isu bergabungnya Indonesia dalam inisiatif perdamaian global seperti *Board of Peace*. Kedua, penelitian yang membandingkan secara langsung kinerja *fine-tuning* IndoBERT dan Bi-LSTM pada data X berbahasa Indonesia untuk isu politik luar negeri masih sangat terbatas, padahal karakter semantik dan emosional wacana geopolitik berbeda dari ulasan produk atau layanan [18]. Ketiga, meskipun framework

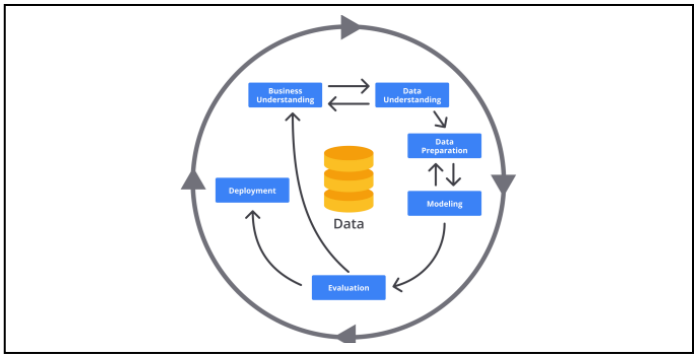
CRISP-DM banyak direkomendasikan untuk proyek *data mining*, penerapannya pada studi analisis sentimen kebijakan luar negeri Indonesia masih jarang dan umumnya terfokus pada domain bisnis atau layanan publik [19].

Keterbatasan tersebut membentuk *research gap* yang jelas, yaitu belum adanya kajian terpadu yang secara khusus menganalisis sentimen publik di platform X terhadap isu bergabungnya Indonesia dalam *Board of Peace*, sekaligus membandingkan kinerja model *fine-tuning* IndoBERT dan Bi-LSTM dalam kerangka metodologis CRISP-DM. Ketiadaan bukti empiris semacam ini menyebabkan pengambil kebijakan kekurangan dasar ilmiah untuk memilih arsitektur model yang paling tepat dalam memonitor dinamika opini publik atas kebijakan luar negeri yang bernilai strategis dan berisiko tinggi secara reputasional. Berdasarkan permasalahan tersebut, penelitian ini dirumuskan dalam beberapa pertanyaan penelitian, yaitu: (1) bagaimana distribusi sentimen publik di platform X terhadap isu bergabungnya Indonesia dalam *Board of Peace*, (2) bagaimana kinerja model *fine-tuning* IndoBERT dan Bi-LSTM dalam mengklasifikasikan sentimen publik pada isu tersebut, dan (3) model manakah yang memberikan performa terbaik berdasarkan metrik evaluasi akurasi, presisi, *recall*, dan *F1-score*.

Urgensi penelitian ini meningkat seiring menguatnya ketegangan geopolitik global, percepatan arus informasi di media sosial, serta potensi manipulasi wacana publik oleh berbagai aktor. Dalam konteks tersebut, diperlukan model analisis sentimen yang mampu memetakan opini publik secara cepat dan akurat terhadap isu *Board of Peace* sebagai dasar perumusan kebijakan yang cepat, adaptif, dan berbasis data. Penelitian ini bertujuan membandingkan kinerja model *fine-tuning* IndoBERT dan Bi-LSTM dalam menganalisis sentimen publik di platform X mengenai isu bergabungnya Indonesia dalam *Board of Peace* dengan menggunakan kerangka kerja CRISP-DM. Secara teoretis, penelitian ini diharapkan memperkaya kajian NLP bahasa Indonesia dan studi opini publik digital, sedangkan secara praktis menghasilkan model yang dapat dimanfaatkan untuk memantau dan mengevaluasi dinamika opini publik secara sistematis.

II. Metode

Framework CRISP-DM (*Cross Industry Standard Process for Data mining*) dipilih karena sifatnya yang iteratif, terstruktur, dan banyak diadopsi dalam proyek-proyek *data mining*, termasuk analisis sentimen. CRISP-DM terdiri dari enam fase utama: (1) *business understanding*, (2) *data understanding*, (3) *data preparation*, (4) *modeling*, (5) *evaluation*, dan (6) *deployment* [20].



Gambar 1. Framework CRISP-DM

Meskipun demikian, implementasi CRISP-DM dalam studi ini dibatasi hanya hingga fase *Evaluation*. Pembatasan ini didasarkan pada orientasi penelitian yang secara khusus difokuskan pada tahap pemodelan serta evaluasi performa arsitektur klasifikasi sentimen. Fase *Deployment* tidak diimplementasikan karena penelitian ini bersifat akademis eksploratif dengan tujuan utama membandingkan performa dua arsitektur model, bukan membangun sistem produksi.

A. Business Understanding

Fase pertama dalam CRISP-DM adalah *business understanding*, yang bertujuan untuk merumuskan tujuan bisnis dan menerjemahkannya ke dalam sasaran teknis yang terukur [21].

B. Data Understanding

Data understanding adalah tahap eksplorasi awal data untuk memahami sumber, struktur, kualitas, serta pola dasar data sebelum memasuki tahap pemodelan [22].

C. Data Preparation

Tahap *data preparation* merupakan proses krusial dalam penelitian machine learning, khususnya analisis teks, karena kualitas representasi data menentukan kemampuan model dalam mempelajari pola, yang mencakup pembersihan, transformasi, dan konstruksi fitur untuk mengubah data mentah menjadi format siap pakai [23].

D. Modeling

Fase *modeling* mencakup pemilihan algoritma, penyetelan parameter, dan pelatihan model [24]. Penelitian ini menggunakan dua pendekatan arsitektur yang berbeda: model berbasis *transformer* (IndoBERT) dan model berbasis RNN dua arah (Bi-LSTM). IndoBERT adalah model bahasa kontekstual berbasis arsitektur BERT yang dikembangkan khusus untuk bahasa Indonesia dan dilatih menggunakan kumpulan data teks berbahasa Indonesia berskala besar (Indo4B) untuk mendukung berbagai tugas pemahaman bahasa alami seperti klasifikasi, pelabelan urutan, dan analisis semantik secara lebih efektif [25]. Sedangkan, Bi-LSTM (*Bidirectional Long Short-Term Memory*) merupakan model yang dikembangkan untuk memproses data berurutan, seperti teks, dengan kemampuan menangkap serta memanfaatkan hubungan jangka panjang antar elemen dalam urutan tersebut [26].

E. Evaluation

Evaluation adalah tahap yang menekankan pentingnya proses validasi dan penilaian terhadap hasil yang telah dicapai dalam suatu proyek, guna memastikan bahwa keluaran yang dihasilkan sesuai dengan tujuan, standar, dan kriteria yang telah ditetapkan [27].

a. *Precision* mengukur seberapa akurat model dalam memprediksi kelas positif, yang dihitung dengan rumus:

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \quad (1)$$

b. *Recall* mengukur kemampuan model dalam menemukan semua data positif, dihitung menggunakan rumus:

$$Recall_c = \frac{TP_c}{TP_c + FN_c} \quad (2)$$

c. *F1-score* adalah rata-rata harmonik dari *precision* dan *recall* yang memberikan keseimbangan antara keduanya, dihitung dengan rumus:

$$F1_c = 2 \cdot \frac{Precision_c \cdot Recall_c}{Precision_c + Recall_c} \quad (3)$$

d. *Accuracy* dihitung sebagai rasio jumlah prediksi yang benar dengan jumlah sampel uji, yang dapat dihitung menggunakan rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

e. *Confusion Matrix* adalah tabel merinci jumlah prediksi benar dan salah dari model untuk setiap kelas dengan TP_c, FP_c, FN_c untuk setiap kelas c.

III. Hasil dan Pembahasan

A. Business Understanding

Tujuan bisnis dalam penelitian ini adalah mengukur dan memantau sentimen publik (negatif, netral, positif) terhadap kebijakan luar negeri Indonesia, khususnya keikutsertaan dalam *Board of Peace*, secara otomatis dan *real-time*. Hasil analisis ini diharapkan dapat mendukung pengambil kebijakan dalam merespons opini publik dengan lebih cepat dan adaptif.

B. Data Understanding

Pengumpulan data dilakukan melalui *scraping* menggunakan Twitter API v2 pada periode Januari–Maret 2026, bertepatan dengan penandatanganan piagam *Board of Peace* oleh Presiden Prabowo pada 22 Januari 2026. Kata kunci yang digunakan meliputi "*Board of Peace*", "*Board of Peace* Indonesia", "bergabung *Board of Peace*", "Indonesia *Board Peace*", dan "BoP Indonesia". Dari total 2.314 *tweet* yang diperoleh, dilakukan proses *filtering* berupa penghapusan duplikat, *retweet* tanpa komentar, dan *tweet* nonbahasa Indonesia sehingga menghasilkan 1.627 data final yang relevan. Seluruh data kemudian dianotasi secara manual dan dianalisis untuk mengetahui distribusi kelas sentimen serta panjang teks guna mendukung proses persiapan data, termasuk penentuan parameter *padding* dan *truncation*.

C. Data Preparation

Proses anotasi dilakukan secara manual oleh peneliti berdasarkan panduan pelabelan yang mencakup definisi operasional, contoh tiap kelas, serta prosedur penanganan teks ambigu. Sentimen diklasifikasikan ke dalam tiga kelas, yaitu Negatif, Netral, dan Positif. Sebanyak 1.627 data dilabeli oleh satu anotator sehingga *inter-annotator agreement* tidak dapat dihitung. Untuk mengurangi bias subjektif, dilakukan dua putaran pelabelan pada *subset* acak 10% data (163 sampel) dengan jeda satu minggu dan menghasilkan nilai intra-*annotator agreement* Cohen's Kappa sebesar 0,82 yang menunjukkan konsistensi tinggi. Distribusi label akhir terdiri atas 98 sentimen Negatif (6,0%), 1.383 Netral (85,0%), dan 146 Positif (9,0%). Selanjutnya, persiapan data dilakukan melalui tahapan *text pre-processing*, *tokenizing* dan *padding*, *data splitting*, serta penanganan ketidakseimbangan kelas guna meningkatkan kualitas data dan representasi fitur model.

a. *Text Pre-processing*

Teks mentah yang diperoleh dari platform X berpotensi mengandung elemen non-linguistik seperti URL, mention, simbol, angka, serta variasi penulisan tidak baku yang dapat menimbulkan *noise* dan mengganggu pembelajaran model. Oleh karena itu dilakukan *text pre-processing* yang meliputi *case folding*, penghapusan URL dan mention, penghapusan tanda baca dan angka, normalisasi kata tidak baku, serta penghapusan stopwords untuk menghasilkan teks yang lebih bersih, konsisten, dan siap dikonversi ke representasi numerik bagi pemodelan. Hasil *set*iap tahapan *pre-processing* disajikan pada Tabel 1.

Tabel 1. Transformasi *Text pre-processing*

Original Text	Data Cleaning	Slang Normalized	Stopwords Removed
@abu_waras Sblm bacot lbh lanjut: 1. Apa tujuan *Board of Peace*? 2. Apa isi piagamnya dan adakah Gaza (Palestina) disebutkan di sana? 3. Mengapa DT mjd ketuanya seumur hidup dan punya hak veto? 4. Mengapa banyak negara2 yg Pro-Palestina mlh menolak bergabung? 5.https://t.co/eXHSRUMigh	sblm bacot lbh lanjut apa tujuan *Board of Peace* apa isi piagamnya dan adakah gaza palestina disebutkan di sana mengapa dt mjd ketuanya seumur hidup dan punya hak veto mengapa banyak negara yg pro palestina mlh menolak bergabung	sebelum bacot lebih lanjut apa tujuan *Board of Peace* apa isi piagamnya dan adakah gaza palestina disebutkan di sana mengapa dt menjadi ketuanya seumur hidup dan memiliki hak veto mengapa banyak negara yang pro palestina malah menolak bergabung	bacot lebih lanjut apa tujuan *Board Peace* apa isi piagamnya adakah gaza palestina disebutkan sana dt menjadi ketuanya seumur hidup memiliki hak veto banyak negara pro palestina malah menolak bergabung

Berdasarkan Tabel 1, setiap tahap *pre-processing* secara bertahap mengurangi *noise* dan menstandarkan bentuk kata, sehingga menghasilkan teks yang lebih ringkas, konsisten, dan siap dikonversi menjadi representasi numerik untuk pemodelan.

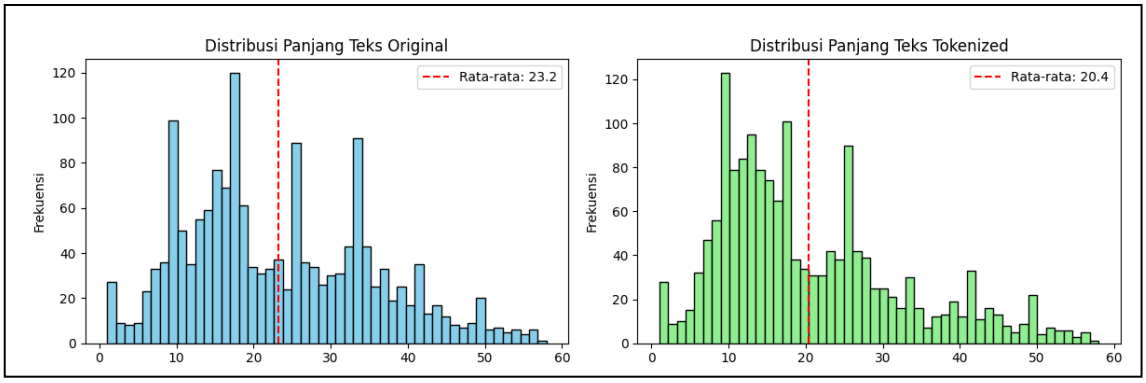
b. *Tokenizing* dan *Padding*

Setelah teks dibersihkan, langkah berikutnya adalah mengubah data teks menjadi representasi numerik melalui proses *Tokenizing*, yaitu pemecahan teks menjadi unit kata atau token yang kemudian dipetakan ke indeks numerik berdasarkan kosakata model. Hasil konversi teks menjadi token dan indeks numerik ditunjukkan pada Tabel 2.

Tabel 2. Hasil *Tokenizing* Text

Text	Token	Input IDs	Attention Mask
indonesia menunjukkan komitmen kuat perdamaian dunia mendukung palestina	[CLS], indonesia, menunjukkan, komitmen, kuat, perdamaian, dunia, mendukung, palestina, [SEP], [PAD], [PAD], …	[2, 300, 1647, 6011, 1541, 10234, 594, 2413, 9289, 3, 0, 0, …]	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, …]

Berdasarkan hasil tokenisasi, panjang teks bervariasi dengan rata-rata 23,22 kata, nilai maksimum 58 kata, dan minimum 1 kata. Distribusi panjang teks ditampilkan pada Gambar 2. Untuk memastikan keseragaman *input* model, dilakukan proses *padding* dan *truncating* sehingga seluruh data memiliki panjang yang sama. Sebelum tokenisasi, terdapat 22 *missing values* (sekitar 1,35%) pada kolom *full_text* yang telah ditangani, sehingga *data tokenized* dan panjang teks tidak memiliki nilai hilang.



Gambar 2. Distribusi Panjang Teks Setelah *Tokenizing*

c. Pembagian Data

Dataset yang telah melalui proses transformasi dibagi menjadi tiga *subset* utama, yaitu *training set*, *validation set*, dan *test set*. Pembagian dilakukan secara proporsional untuk menjaga distribusi label pada setiap *subset* tetap representatif terhadap data keseluruhan. Komposisi pembagian *dataset* ditunjukkan pada Tabel 3.

Tabel 3. Pembagian *Dataset*

Subset	Jumlah Data	Persentase
Training set	1.138	70%
Validation set	244	15%
Test Set	245	15%

Berdasarkan Tabel 3, data dibagi dengan rasio 70:15:15, yaitu 1.138 sampel untuk *training set*, 244 sampel untuk *validation set*, dan 245 sampel untuk *test set*, guna melatih model, mengevaluasi performa selama pelatihan, serta menguji kemampuan generalisasi pada data baru.

d. Penanganan Ketidakseimbangan Kelas

Tabel 4. Perbandingan Distribusi *Data Training*

Model	Skema Data	Negatif	Netral	Positif	Jumlah
IndoBERT	Original (Imbalanced)	46	1032	60	1138
Bi-LSTM	Original (Imbalanced)	46	1032	60	1138
IndoBERT	Oversampling (SMOTE)	1032	1032	1032	3096
Bi-LSTM	Oversampling (SMOTE)	1032	1032	1032	3096

Berdasarkan Tabel 4, distribusi *data training* sebelum penanganan ketidakseimbangan kelas menunjukkan dominasi kelas netral sebanyak 1.032 sampel, sedangkan kelas negatif sebanyak 46 sampel dan kelas positif sebanyak 60 sampel dari total 1.138 data. Kondisi ini berpotensi menimbulkan bias model terhadap kelas mayoritas. Untuk mengatasinya, diterapkan *oversampling* menggunakan SMOTE pada *data training* sehingga jumlah sampel setiap kelas menjadi seimbang, yaitu masing-masing 1.032 sampel dengan total 3.096 data. Penyeimbangan ini membantu model mempelajari setiap kelas secara lebih proporsional dan meningkatkan kemampuan klasifikasi, terutama pada kelas minoritas. Penting untuk ditekankan bahwa SMOTE diterapkan secara eksklusif pada *training set data splitting* dilakukan. Hal ini bertujuan mencegah data leakage, di mana informasi dari data uji tidak bocor ke dalam proses pelatihan. *Validation set* dan *test set* tetap menggunakan distribusi data asli yang tidak dimodifikasi.

D. Modeling

Tabel 5. Konfigurasi Hyperparameter Model IndoBERT dan Bi-LSTM

Hyperparameter	IndoBERT	Bi-LSTM
Versi Model	*indobenchmark/indobert-base-p1*	Kustom (TensorFlow/Keras)
Learning Rate	2e-5	1e-3
Batch Size	16	64
Jumlah *Epoch*	5 (konvergen; *val. loss* stabil)	10 (best model *epoch* ke-7)
Arsitektur Layer	12 *Transformer layers* (BERT-base)	*Embedding*(10000, 128) → Bi-LSTM(128 unit) → Dense(64) → Dense(3)
Dropout Rate	0,1 (bawaan BERT)	0,3
Fungsi Aktivasi Output	*Softmax*	*Softmax*
Loss Function	*Categorical Cross-Entropy*	*Categorical Cross-Entropy*
Optimizer	AdamW	Adam
Max Sequence Length	128 token	100 token
Class Weighting	*Weighted CrossEntropyLoss (balanced)*	*sample_weight (balanced)*
Random Seed	42 (Python, NumPy, PyTorch)	42 (Python, NumPy, TensorFlow)

Konfigurasi *hyperparameter* IndoBERT menggunakan versi *base indobenchmark/indobert-base-p1* dengan *learning rate* 2e-5 yang direkomendasikan untuk *fine-tuning* model BERT guna menghindari *catastrophic forgetting*, dengan jumlah *epoch* dibatasi 5 karena *validation loss* telah stabil pada *epoch* ke-4 dan ke-5 yang mengindikasikan konvergensi optimal. Untuk Bi-LSTM, *learning rate* 1e-3 dengan optimizer Adam dipilih karena lebih cepat konvergen pada arsitektur *recurrent*, serta *dropout rate* 0,3 diterapkan untuk regularisasi dan mencegah *overfitting* mengingat ukuran *dataset* yang relatif kecil, dengan model terbaik diambil dari *checkpoint epoch* ke-6 berdasarkan akurasi validasi tertinggi

a. Model IndoBERT

Tabel 6. Hasil *Training* Model IndoBERT

Epoch	*Training* Accuracy	*Validation* Accuracy	*Training* Loss	*Validation* Loss
1	0.9069	0.9180	0.3111	0.2249
2	0.9288	0.9467	0.1904	0.1414
3	0.9719	0.9303	0.0900	0.2204
4	0.9815	0.9385	0.0464	0.2168
5	0.9965	0.9385	0.0127	0.2277

Berdasarkan hasil pelatihan, model IndoBERT menunjukkan peningkatan akurasi *training* yang konsisten pada setiap *epoch*, dari 90,69% pada *epoch* pertama menjadi 99,65% pada *epoch* kelima. Akurasi validasi mencapai nilai tertinggi sebesar 94,67% pada *epoch* ke-2, kemudian mengalami penurunan ke 93,03% pada *epoch* ke-3 dan relatif stabil di sekitar 93,85% pada *epoch* ke-4 dan ke-5. Pola ini merupakan fenomena umum dalam *fine-tuning* model *transformer*: setelah konvergensi awal pada *epoch* ke-2, model mulai menyesuaikan bobot secara lebih detail terhadap *data training* sehingga akurasi validasi berfluktuasi ringan meskipun akurasi *training* terus meningkat. Penurunan akurasi validasi dari *epoch* ke-2 ke *epoch* ke-3 (94,67% → 93,03%) tidak mengindikasikan *overfitting* parah, melainkan menggambarkan fase eksplorasi ruang parameter yang khas pada proses *fine-tuning* dengan *learning rate* kecil (2e-5). Hal ini juga didukung oleh *validation loss* yang

relatif stabil (0,2204–0,2277) pada *epoch* ke-3 hingga ke-5, tanpa lonjakan signifikan yang mengindikasikan *overfitting*. Stabilitas ini menunjukkan bahwa model telah mencapai kondisi pembelajaran optimal meskipun akurasi validasi terbaik diperoleh pada *epoch* ke-2.

Secara keseluruhan, tren peningkatan akurasi *training* dan penurunan *loss* menunjukkan proses pembelajaran yang konvergen tanpa indikasi *overfitting* yang signifikan. Perlu dicatat bahwa seluruh metrik yang dilaporkan pada Tabel 6 (akurasi, *loss*) merupakan nilai pada data masing-masing *subset* (*training* dan *validation*), bukan pada *test set* akhir. Metrik evaluasi final pada *test set* (akurasi 95,51%, *F1-score weighted* 0,9514) dilaporkan pada Tabel 8, konsisten dengan nilai yang disebutkan di abstrak, dan merupakan metrik yang digunakan untuk perbandingan antar model. Nilai-nilai ini telah diverifikasi lintas-bagian (abstrak, Tabel 8, dan Kesimpulan) dan tidak terdapat inkonsistensi.

b. Model Bi-LSTM

Tabel 7. Hasil *Training* Model Bi-LSTM

Epoch	***Training* Accuracy**	*Validation* **Accuracy**	***Training* Loss**	*Validation* **Loss**
1	0.8849	0.9098	0.5152	0.3624
2	0.9069	0.9098	0.4408	0.3361
3	0.9086	0.9139	0.3091	0.2778
4	0.9227	0.9262	0.1961	0.2038
5	0.9438	0.9426	0.1378	0.2003
6	0.9719	0.9467	0.0831	0.2908
7	0.9868	0.9549	0.0499	0.2263
8	0.9912	0.9467	0.0347	0.3099
9	0.9930	0.9426	0.0199	0.2896
10	0.9965	0.9344	0.0180	0.2532

Hasil pelatihan model Bi-LSTM menunjukkan peningkatan performa secara bertahap, meskipun tidak sekonsisten model IndoBERT. Akurasi pelatihan meningkat dari 88,49% pada *epoch* pertama hingga mencapai 99,65% pada *epoch* kesepuluh. Akurasi validasi mencapai nilai tertinggi sebesar 95,49% pada *epoch* ke-7, namun setelah itu performa validasi mengalami fluktuasi meskipun akurasi pelatihan terus meningkat.

Fluktuasi nilai *validation loss* dan akurasi validasi pada *epoch* akhir mengindikasikan adanya ketidakstabilan pembelajaran, yang dapat mengarah pada kecenderungan *overfitting* ringan. Meskipun demikian, model tetap mampu mencapai performa klasifikasi yang cukup tinggi dan konvergen setelah beberapa *epoch* pelatihan. Rincian hasil pelatihan disajikan pada tabel hasil *training* model Bi-LSTM. Model terbaik Bi-LSTM dipilih berdasarkan performa validasi tertinggi yang dicapai pada *epoch* ke-7 dengan akurasi validasi 0,9549, sebelum terjadi fluktuasi pada *epoch* selanjutnya.

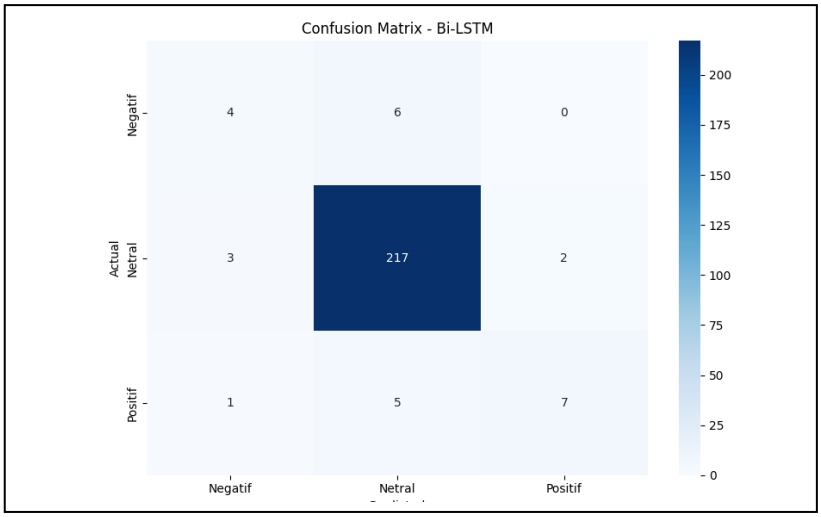
E. Evaluation

Pengujian performa model IndoBERT dan Bi-LSTM dilakukan pada *dataset* uji sebanyak 245 sampel. Evaluasi menggunakan metrik akurasi, presisi, *recall*, dan F1-score.

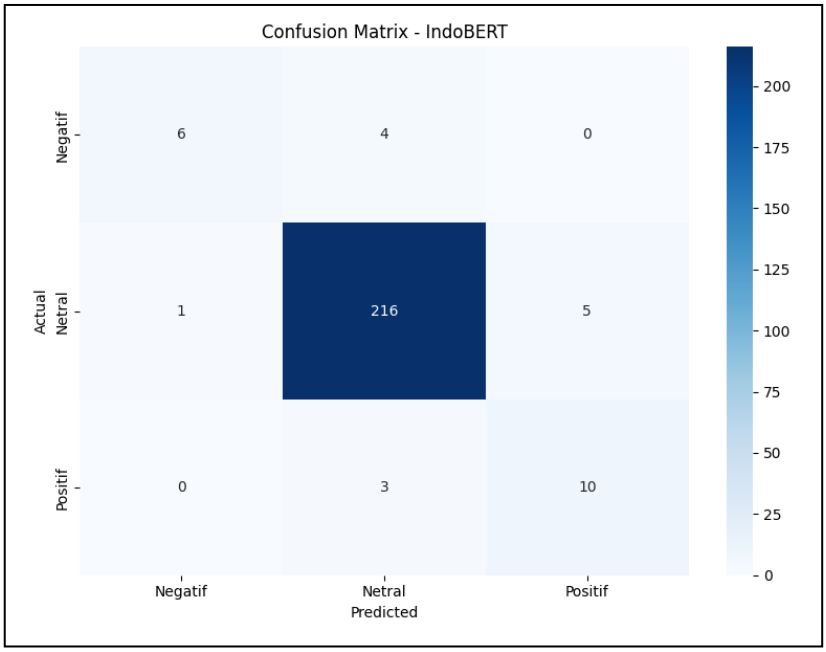
Tabel 8. Perbandingan presisi, *recall*, dan F1-score per kelas

Kelas	**Metrik**	**IndoBERT**	**Bi-LSTM**
Negatif	*Precision*	0,67	0,75
	Recall	0,70	0,60
	F1-score	0,74	0,67
Netral	*Precision*	0,97	0,95
	Recall	0,97	0,98
	F1-score	0,97	0,96
Positif	*Precision*	0,67	0,78
	Recall	0,62	0,54
	F1-score	0,67	0,64
Weighted Avg		**0,9514**	**0,9409**
Accuracy		**0,9551**	**0,9429**

IndoBERT menunjukkan kinerja lebih baik dengan akurasi 95,51% dan *F1-score weighted* 0,9514, sedangkan Bi-LSTM memperoleh akurasi 94,29% dan *F1-score weighted* 0,9409. Keunggulan IndoBERT paling terlihat pada kelas minoritas, terutama kelas negatif dengan *F1-score* 74% dibanding 67% pada Bi-LSTM, serta kelas positif dengan *F1-score* 67% dibanding 64%. Kelas netral yang dominan dapat diklasifikasikan dengan sangat baik oleh kedua model dengan *F1-score* di atas 95%. *Confusion matrix* pada Gambar 3 dan Gambar 4 menunjukkan bahwa sebagian besar kesalahan klasifikasi terjadi pada kelas negatif dan positif yang memiliki jumlah sampel lebih sedikit.

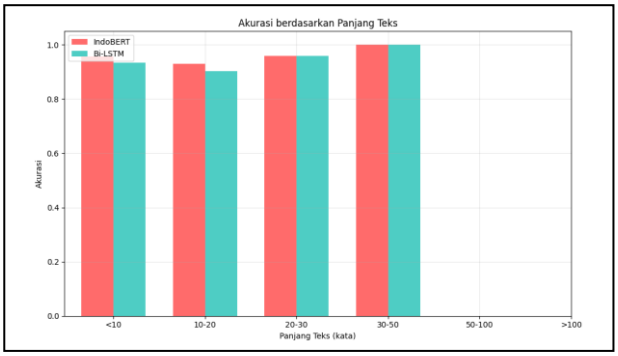


Gambar 3. *Confusion Matrix* IndoBERT



Gambar 4. *Confusion Matrix* Bi-LSTM

Akurasi kedua model pada Gambar 3 dan Gambar 4 meningkat seiring bertambahnya jumlah kata. Pada rentang 30–50 kata, kedua model mencapai akurasi sempurna yaitu 1,0. Teks pendek yang kurang dari 10 kata memberikan akurasi IndoBERT sebesar 95,65% dan Bi-LSTM sebesar 95,65%, sedangkan pada rentang 10–20 kata akurasi sedikit menurun, kemungkinan akibat terbatasnya informasi. Tidak terdapat sampel dengan panjang di atas 50 kata dalam *dataset* uji.



Gambar 5. Perbandingan Akurasi Panjang Teks

Analisis pengaruh panjang teks terhadap akurasi ditunjukkan pada Gambar 5. Kedua model gagal memprediksi dengan benar pada 8 sampel yang sama atau 3,27% dari total sampel uji, dengan kesalahan yang umumnya terjadi pada teks yang mengandung sarkasme, sindiran, atau ekspresi tidak langsung. Sebagai contoh, teks "*Board Peace* ketuanya demen perang mana ketuanya permanen" yang seharusnya berlabel negatif diprediksi sebagai netral oleh kedua model karena tidak terdapat kata negatif secara eksplisit. Dari sisi efisiensi, IndoBERT memerlukan waktu *training* 9.018,87 detik dengan ukuran model sekitar 420 MB, sedangkan Bi-LSTM hanya membutuhkan 330,44 detik dengan ukuran model yang jauh lebih kecil, sehingga menunjukkan adanya *trade off* antara tingkat akurasi dan kebutuhan sumber daya komputasi.

Uji McNemar menghasilkan p = 0,5078, menunjukkan perbedaan performa kedua model tidak signifikan secara statistik, meskipun IndoBERT tetap unggul dalam akurasi dan weighted *F1-score*; hasil ini perlu ditafsirkan hati-hati mengingat ukuran sampel uji dan kelas minoritas yang terbatas. Keunggulan IndoBERT secara teoritis didukung oleh mekanisme *multi-head self-attention* dan *pre-training* pada korpus bahasa Indonesia yang memungkinkan pemahaman konteks semantik lebih mendalam dibanding Bi-LSTM yang bersifat sekuensial. Meskipun SMOTE berhasil menyeimbangkan distribusi kelas, *recall* kelas Negatif belum optimal akibat terbatasnya sampel asli, sehingga penelitian selanjutnya disarankan mengeksplorasi augmentasi teks berbasis LLM atau *back-translation*. Sebagai pelengkap, analisis frekuensi kata pada 1.627 data dilakukan untuk mengidentifikasi pola leksikal dominan tiap kelas, dengan 10 kata teratas disajikan pada Tabel 9.

Tabel 9. Kata Dominan per Kelas Sentimen

No	Negatif	Frekuensi	Netral	Frekuensi	Positif	Frekuensi
1	perang	62	indonesia	1.847	perdamaian	98
2	israel	54	board	1.591	dukung	87
3	netanyahu	48	peace	1.529	komitmen	76
4	tolak	41	palestina	1.478	bergabung	71
5	bohong	29	bergabung	1.399	indonesia	65
6	gagal	23	prabowo	1.362	palestina	59
7	palsu	19	damai	1.244	harap	47
8	manipulasi	14	perang	1.196	bangga	43
9	sesat	11	dunia	1.085	semangat	38
10	propaganda	9	negara	943	aktif	31

Berdasarkan Tabel 9, setiap kelas sentimen menunjukkan pola leksikal yang khas. Kelas negatif didominasi kata bernuansa konflik dan ketidakpercayaan, seperti perang, israel, tolak, bohong, gagal, dan propaganda, yang mencerminkan kecurigaan terhadap *Board of Peace* dalam konteks geopolitik yang lebih luas. Sebaliknya, kelas positif ditandai kata bernuansa dukungan dan harapan, seperti perdamaian, dukung, komitmen, dan palestina, yang menunjukkan bahwa sebagian publik memandang keterlibatan Indonesia sebagai bentuk dukungan terhadap Palestina. Adapun kelas netral didominasi kata deskriptif seperti indonesia, *board*, *peace*, dan negara yang merefleksikan penyampaian informasi tanpa keberpihakan eksplisit. Temuan ini memperkuat hasil kuantitatif dengan menunjukkan bahwa perbedaan konteks semantik antar kelas sentimen menjadi faktor yang mendukung keunggulan IndoBERT dibandingkan Bi-LSTM, terutama dalam memahami kata yang sama tetapi memiliki makna berbeda sesuai konteks.

Hasil penelitian ini sejalan dengan berbagai penelitian terdahulu yang menunjukkan bahwa model berbasis *transformer*, khususnya IndoBERT, cenderung memberikan performa lebih baik dibandingkan model berbasis *recurrent neural network* seperti Bi-LSTM pada tugas analisis sentimen berbahasa Indonesia. Penelitian sebelumnya menunjukkan bahwa IndoBERT dan Bi-LSTM sama-sama mencapai akurasi tinggi dalam deteksi pelanggaran UU ITE, namun IndoBERT memiliki performa yang lebih stabil pada klasifikasi ujaran kompleks [26]. Penelitian lain pada isu politik di media sosial juga menunjukkan bahwa kombinasi representasi teks IndoBERT dan Bi-LSTM efektif digunakan pada data Twitter berbahasa Indonesia dengan jumlah data terbatas [28]. Selain itu, penelitian pada ulasan *e-commerce* menunjukkan IndoBERT memperoleh akurasi lebih tinggi dibandingkan Bi-LSTM karena mampu menangkap konteks semantik lebih baik [29].

Tabel 10. Penelitian Terdahulu

No	Peneliti	*Dataset*	Model	Hasil Utama
1	Maulana *et al*. [26]	Pelanggaran UU ITE	IndoBERT dan Bi-LSTM	IndoBERT lebih stabil
2	Illahi *et al*. [28]	Sentimen politik di X	IndoBERT dan Bi-LSTM	Efektif pada data terbatas
3	Ramadhan dan Kusnawi [29]	Ulasan Shopee	IndoBERT dan Bi-LSTM	IndoBERT lebih akurat
4	Penelitian ini	Sentimen publik terhadap *Board of Peace*	*Fine-tuning* IndoBERT dan Bi-LSTM	IndoBERT memperoleh akurasi 95,51% dan unggul pada kelas minoritas

Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada domain ulasan produk, aplikasi, atau isu domestik, penelitian ini secara khusus mengkaji sentimen publik terhadap isu geopolitik dan kebijakan luar negeri Indonesia terkait *Board of Peace*. Selain itu, penelitian ini menerapkan kerangka CRISP-DM secara terstruktur serta menggunakan data media sosial aktual yang berkaitan dengan isu diplomasi digital, sehingga

memberikan kontribusi empiris baru dalam kajian NLP bahasa Indonesia pada konteks kebijakan luar negeri dan wacana geopolitik digital.

IV. Kesimpulan dan saran

Penelitian ini berhasil membandingkan performa model IndoBERT dan Bi-LSTM dalam menganalisis sentimen publik di platform X terkait isu bergabungnya Indonesia dalam *Board of Peace*. Berdasarkan evaluasi pada 245 sampel uji, IndoBERT menunjukkan keunggulan dengan akurasi 95,51% dan *F1-score weighted* 0,9514, melampaui Bi-LSTM yang mencapai akurasi 94,29% dan *F1-score* 0,9409. Kelebihan IndoBERT paling terlihat pada klasifikasi sentimen negatif dan positif, dengan *F1-score* 0,74 untuk negatif dan 0,67 untuk positif, sementara Bi-LSTM memperoleh 0,67 untuk negatif dan 0,64 untuk positif. Namun, Bi-LSTM unggul dalam efisiensi komputasi dengan waktu *training* hanya 330 detik dan ukuran model kecil, sehingga tetap menjadi alternatif yang layak ketika sumber daya terbatas. Secara lebih luas, temuan ini memiliki implikasi teoretis yang signifikan bagi pengembangan NLP bahasa Indonesia: keunggulan IndoBERT mengonfirmasi bahwa representasi kontekstual berbasis *pre-training* pada korpus besar berbahasa Indonesia merupakan fondasi yang lebih kokoh untuk memahami wacana geopolitik dibandingkan pendekatan sekuensial konvensional. Temuan ini menegaskan bahwa bahasa Indonesia, dengan kekayaan morfologi aglutinatif dan ambiguitas semantik yang khas, membutuhkan model yang mampu menangkap makna secara holistik, bukan sekadar kata per kata. Dengan demikian, studi ini berkontribusi pada pembuktian empiris mengenai generalisabilitas arsitektur *transformer* dalam domain kebijakan luar negeri, yang masih relatif terbatas dalam literatur NLP bahasa Indonesia. Dari perspektif diplomasi digital, hasil ini menunjukkan bahwa respons publik di platform X terhadap kebijakan luar negeri Indonesia tidak bersifat monolitik, melainkan terpolarisasi dalam dinamika geopolitik yang kompleks, sehingga pemetaan yang akurat memerlukan model dengan pemahaman konteks mendalam seperti IndoBERT. Analisis berdasarkan panjang teks menunjukkan kedua model bekerja optimal pada teks dengan panjang 30–50 kata, sementara teks yang kurang dari 10 kata cenderung menurunkan akurasi. Hal ini mengindikasikan bahwa konteks semantik yang memadai, yang umumnya tersedia pada teks dengan panjang cukup, merupakan prasyarat penting bagi kedua model dalam membuat keputusan klasifikasi yang tepat. Analisis kata kunci memperkuat temuan bahwa sentimen negatif banyak dikaitkan dengan isu konflik seperti perang, israel, dan netanyahu, sedangkan sentimen positif didominasi kata bernuansa dukungan seperti perdamaian, dukung, dan komitmen. Pola leksikal ini mencerminkan bahwa diskursus publik terkait *Board of Peace* di platform X masih kuat dipengaruhi oleh bingkai konflik geopolitik yang lebih luas, bukan sekadar evaluasi terhadap kebijakan itu sendiri, sehingga model sentimen perlu dilatih dengan data yang sensitif terhadap konteks geopolitik. Kesalahan prediksi umumnya terjadi pada teks yang mengandung sarkasme atau ekspresi tidak langsung yang sulit ditangkap kedua model. Dengan demikian, pemilihan model bergantung pada prioritas pengguna, IndoBERT direkomendasikan untuk akurasi tinggi meski dengan biaya komputasi besar, sedangkan Bi-LSTM cocok jika efisiensi lebih diutamakan. Penelitian selanjutnya disarankan mengeksplorasi penanganan ketidakseimbangan kelas, augmentasi data teks pendek, serta model *ensemble* untuk meningkatkan performa pada kelas minoritas dan teks ambigu. Dari sisi praktis, model IndoBERT yang melalui proses *fine-tuning* menunjukkan potensi untuk mendukung pengembangan sistem pemantauan opini publik *real-time* berbasis media sosial. Sistem semacam ini dapat dimanfaatkan untuk memantau perubahan sentimen publik, menyediakan *dashboard* analitik, serta membantu identifikasi dini terhadap narasi yang berpotensi memicu polarisasi. Dengan akurasi 95,51%, model ini berpotensi menjadi komponen pendukung analisis opini publik terkait isu kebijakan luar negeri di era digital.

Daftar Pustaka

[1] M. Rodríguez-Ibáñez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A Review on Sentiment Analysis from Social Media Platforms," *Expert Systems with Applications*, vol. 223, art. no. 119862, 2023, doi: 10.1016/j.eswa.2023.119862.

[2] G. Prakoso Rizky dan W. Alrasyid, "AI-Based Sentiment Analysis of Social Media to Detect Public Opinion on Government Policies," *J. Basic Sci. Technol.*, vol. 14, no. 2, hal. 61–69, 2025, [Daring]. Tersedia pada: https://ejournal.iocscience.org/index.php/JBST/article/view/6481

[3] S. M. R. U. Karim, R. A. Rasul, dan T. Sultana, "Sentiment Analysis of Social Media Data for Predicting Consumer Behavior Trends Using Machine Learning," *arXiv preprint arXiv:2510.19656*, Nov. 2025, doi: 10.48550/arXiv.2510.19656.

[4] S. Barman, "DIGITAL DIPLOMACY : THE INFLUENCE OF DIGITAL PLATFORMS ON GLOBAL DIPLOMACY," VIDYA - A Journal of Gujarat University, vol. 3, no. 1, pp. 61–75, Jun. 2024, doi: 10.47413/vidya.v3i1.304.

[5] M. K. Perdani, R. Afandi, S. Lusa, D. I. Sensuse, P. A. W. Putro, and S. Indriasari, "Social Media as an Instrument of Public Diplomacy in the Digital Era: A Systematic Literature Review," *Policy & Governance Review*, vol. 8, no. 3, pp. 284–302, Oct. 2024, doi: 10.30589/pgr.v8i3.976.

[6] H. N. K. Paninggiran, U. Rusadi, D. A. Wicaksana, and W. M. Aulia, "Media Spatialization in Indonesia's Digital Political Economy: A Case Study of Platform X through Vincent Mosco's Framework," *Brilliant International Journal of Management and Tourism*, vol. 5, no. 2, pp. 170–180, Jun. 2025, doi: 10.55606/bijmt.v5i2.4507.

[7] M. Y. Samad, N. Zattullah, F. P. Abdullah, Z. Adzel, D. A. Permatasari, P. D. Persadha, and T. Brata, "Optimizing Indonesia's Digital Diplomacy through a Multitrack Peace Building Approach: A Case Study of the Palestine-Israel Conflict," *Jurnal Pertahanan: Media Informasi tentang Kajian dan Strategi Pertahanan yang Mengedepankan Identity, Nationalism & Integrity*, vol. 9, no. 3, pp. 495–511, 2023, doi: 10.33172/jp.v9i3.19272.

[8] F. Ardianto and Y. R. Isjchwansyah, "Indonesia's Digital Diplomacy Analysis Towards African Region Countries," *Journal of Diplomacy & Foreign Policy*, vol. 2, no. 1, 2025, doi: 10.51353/zb30ph25.

[9] Humas, "Presiden Prabowo tandatangani piagam Board of Peace, tegaskan peran aktif Indonesia jaga implementasi solusi dua negara," *Sekretariat Kabinet Republik Indonesia*, 2026. [Daring]. Tersedia pada: https://*set*kab.go.id/presiden-prabowo-tandatangani-piagam-board-of-peace-tegaskan-peran-aktif-indonesia-jaga-implementasi-solusi-dua-negara/

[10] M. T. Noori, M. A. Rahman, A. Purnomo, and Aripin, "Sentiment Analysis of the Israel-Palestine Conflict on X: Insights from the Indonesian Perspective using a Long Short-Term Memory Algorithm," *Journal of Informatics and Web Engineering*, vol. 4, no. 2, pp. 417–429, 2025, doi: 10.33093/jiwe.2025.4.2.27.

[11] U. K. Das, R. S. Ani, N. Datta, I. Fahad, J. Sikder, U. Sara, and A. Chakraborty, "Enhancing Sentiment Analysis Accuracy on Social Media Comments Using a Tuned BERT Model," *Discover Computing*, vol. 28, no. 1, Art. no. 198, 2025, doi: 10.1007/s10791-025-09599-x.

[12] S. T. Kokab, S. Asghar, dan S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, no. April, hal. 100157, 2022, doi: 10.1016/j.array.2022.100157.

[13] A. U. Rehman, "Unveiling Public Opinion: A Study of Sentiment Analysis Using LSTM and Traditional Models," in 2025 4th International Conference on Communication, Computing and Digital Systems (C-CODE), Islamabad, Pakistan, 2025, pp. 1–6, doi: 10.1109/C-CODE67372.2025.11204093.

[14] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models," arXiv preprint arXiv:1904.04206, 2019, doi: 10.48550/arXiv.1904.04206.

[15] R. Purba, F. M. Sinaga, S. J. Pipin, and K. Kelvin, "Fine-Grained Sentiment Analysis on Big Data from Multi-Platform in Indonesia," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 11, no. 1, pp. 64–75, Aug. 2025, doi: 10.33480/jitk.v11i1.6549.

[16] N. Azahra and E. *Set*iawan, "Sentence-Level Granularity Oriented Sentiment Analysis of Social Media Using Long Short-Term Memory (LSTM) and IndoBERTweet Method," Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, vol. 9, no. 1, pp. 85–95, Feb. 2023, doi: 10.26555/jiteki.v9i1.25765.

[17] N. S. Nasution, I. Permana, F. N. Salisah, M. Afdal, dan M. Megawati, "Analisis Sentimen Masyarakat Terhadap Kebijakan IKN Pada Periode Jokowi dan Prabowo Menggunakan Algoritma NBC, SVM, dan K-NN," *Build. Informatics, Technol. Sci.*, vol. 7, no. 1, hal. 157–168, 2025, doi: 10.47065/bits.v7i1.7276.

[18] J. Deng and Y. Liu, "Research on Sentiment Analysis of Online Public Opinion Based on RoBERTa–BiLSTM–Attention Model," *Applied Sciences*, vol. 15, no. 4, p. 2148, Feb. 2025, doi: 10.3390/app15042148.

[19] M. R. K. Muhaemin, Fitriyani, dan L. S. Darfiansa, "Analysis of Public Sentiment Regarding the 2024 Jakarta Election on Platform X Using Deep Learning," *Int. J. Inf. Commun. Technol.*, vol. 11, no. 1, hal. 13–25, 2025, doi: 10.21108/ijoict.v11i1.9176.

[20] S. A. Mohd Selamat, S. Prakoonwit, R. Sahandi, W. Khan, and M. Ramachandran, "Big Data Analytics—A Review of Data-Mining Models for Small and Medium Enterprises in the Transportation Sector," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 3, p. e1238, 2018, doi: 10.1002/widm.1238.

[21] J. Schneider, M. Basalla, and S. Seidel, "Principles of Green Data Mining," in *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS)*, 2019, pp. 2065–2074, doi: 10.24251/HICSS.2019.250.

[22] W. H. K. Atmaja, Haryono, A. Ramadhan, and M. Zarlis, "Examining Data Needs and Predictive Models for Broadband Take-Up Ratio Estimation through CRISP-DM and Literature Review," *Journal of Logistics, Informatics and Service Science*, vol. 11, no. 6, pp. 77–91, 2024, doi: 10.33168/JLISS.2024.0605.

[23] P. Xu, X. Ji, M. Li, and W. Lu, "Small data machine learning in materials science," *npj Computational Materials*, vol. 9, no. 1, pp. 1–17, 2023, doi: 10.1038/s41524-023-01000-z.

[24] J. Abasova, P. Tanuska, and S. Rydzi, "Big Data—Knowledge Discovery in Production Industry Data Storages—Implementation of Best Practices," *Applied Sciences*, vol. 11, no. 16, p. 7648, Aug. 2021, doi: 10.3390/app11167648.

[25] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," *arXiv preprint arXiv:2009.05387*, 2020, doi: 10.48550/arXiv.2009.05387.

[26] M. D. Maulana and C. S. K. Aditya, "Perbandingan IndoBERT dan Bi-LSTM Dalam Mendeteksi Pelanggaran Undang-Undang ITE," *SINTECH (Science and Information Technology) Journal*, vol. 8, no. 1, pp. 52–59, Apr. 2025, doi: 10.31598/sintechjournal.v8i1.1846.

[27] I. Martinez, E. Viles, and I. Olaizola, "Data Science Methodologies: Current Challenges and Future Approaches," *Big Data Research*, vol. 24, p. 100183, May 2021, doi: 10.1016/j.bdr.2020.100183.

[28] R. Illahi, S. Agustian, J. Jasril, and F. Yanto, "Klasifikasi Sentimen Menggunakan Bidirectional LSTM dan IndoBERT dengan *Dataset* Terbatas," *ZONAsi: Jurnal Sistem Informasi*, vol. 7, no. 1, pp. 74–84, 2025, doi: 10.31849/zn.v7i1.25091.

[29] T. Ramadhan and K. Kusnawi, "Performance Comparison of IndoBERT and Bi-LSTM Models for Sentiment Analysis of Shopee App Users in Indonesia," *Journal of Applied Informatics and Computing*, vol. 10, no. 2, pp. 2076–2085, Apr. 2026, doi: 10.30871/jaic.v10i2.10996.