KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP AKSI 17+8 DI MEDIA SOSIAL MENGGUNAKAN LSTM DAN BERT

Sonia Roselina Correia[1*], Sumarlin[2], Dewi Anggraini[3]
[1,2,3] Program Studi Teknik Informatika STIKOM Uyelindo Kupang.
[1*]roselinania12@gmail.com, [2]shumarlin@gmail.com, [3]thewifoeh@gmail.com
DOI: https://doi.org/10.46880/mtk.v12i2.6096

ABSTRACT

This study aims to compare the performance of Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) models in classifying public sentiment on platform X (Twitter) regarding the "17+8 Tuntutan Rakyat" movement. The dataset consists of 1,000 Indonesian tweets collected between August and September 2025, with a subset of 200 data evaluated using a 5-Fold Cross Validation scheme. The average evaluation results show that the LSTM model achieved an accuracy of 0.6900, whereas BERT achieved 0.5400. However, per-class metric analysis reveals that LSTM suffered from severe majority-class bias by predicting all instances as neutral (F1-score of 0.0000 for both positive and negative classes), whereas BERT demonstrated discrimination capability on minority classes (negative recall of 26.83% and positive recall of 14.29%). This research is limited by a small evaluation subset size and the absence of class imbalance handling techniques, which implies the crucial need for GPU acceleration and resampling methods in future social media text sentiment analysis studies.

Keywords: Sentiment Analysis, LSTM, BERT, Twitter/X, Imbalanced Dataset, Indonesian Language.

I. PENDAHULUAN

Perkembangan media sosial telah menghasilkan volume data teks yang sangat besar dan menjadi sumber informasi penting untuk memahami opini masyarakat terhadap berbagai isu. Analisis terhadap data teks tersebut memerlukan teknik Natural Language Processing (NLP) agar informasi yang terkandung dalam bahasa alami dapat diolah secara otomatis oleh komputer. NLP memungkinkan proses ekstraksi informasi, klasifikasi, serta analisis sentimen dari data teks sehingga banyak diterapkan dalam berbagai penelitian berbasis media sosial [1]. Sebelum dilakukan proses klasifikasi, data teks umumnya melalui tahapan text preprocessing seperti case folding, tokenizing, stopword removal, dan stemming untuk mengurangi noise serta meningkatkan kualitas data yang akan dianalisis [2], [3]. Khususnya pada dinamika sosial-politik Indonesia di platform X (Twitter) periode Agustus–September 2025, isu Aksi 17+8 Tuntutan Rakyat muncul sebagai diskursus publik berskala nasional yang dipicu oleh isu-isu krusial seperti wafatnya aktivis Affan Kurniawan [12], [13]. Teks media sosial pada fenomena ini sangat dinamis, sarat emosi, dan memiliki ketidakseimbangan kelas opini

Perkembangan Machine Learning dan Deep Learning semakin meningkatkan kemampuan sistem dalam memahami pola bahasa alami. Deep Learning mampu mempelajari representasi data secara otomatis sehingga menghasilkan performa yang lebih baik dibandingkan metode konvensional pada berbagai tugas klasifikasi teks [4],[5]. Salah satu perkembangan penting dalam Deep Learning adalah arsitektur Transformer yang menjadi dasar pengembangan model Bidirectional Encoder Representations from Transformers (BERT). Model ini mampu memahami hubungan antar kata berdasarkan konteks kalimat sehingga menghasilkan representasi bahasa yang lebih baik. Pada bahasa Indonesia, proses fine-tuning BERT juga terbukti mampu meningkatkan pemahaman terhadap ambiguitas dan konteks kalimat [6].

Selain BERT, model Long Short-Term Memory (LSTM) juga banyak digunakan dalam analisis sentimen karena memiliki kemampuan menyimpan informasi jangka panjang pada data sekuensial. Model ini efektif dalam mengolah teks yang memiliki ketergantungan antar kata sehingga sering digunakan untuk klasifikasi sentimen pada media sosial [5]. Perlu dicatat bahwa perlakuan preprocessing pada kedua arsitektur memiliki karakteristik berbeda; LSTM membutuhkan stemming dan stopword removal kaku, sedangkan BERT memanfaatkan WordPiece Tokenizer yang mempertahankan konteks alami kata [6].

Kajian literatur terkini (2020–2025) menunjukkan bahwa model berbasis Deep Learning mampu memberikan hasil yang baik pada tugas analisis sentimen dalam beragam konteks. Bouchra et al. menerapkan metode Deep Learning untuk menganalisis sentimen masyarakat terhadap tayangan televisi nasional Penelitian lain oleh Wicaksono dan Nastiti terhadap komentar YouTube menghasilkan akurasi 99,61% menggunakan LSTM [9]. Pada konteks sentimen politik Indonesia, Illahi et al. membuktikan kombinasi Bi-LSTM dan IndoBERT tetap mampu menghasilkan performa baik pada data terbatas [7], sementara Irianti et al. memperlihatkan bahwa model berbasis BERT efektif diterapkan untuk menganalisis opini masyarakat pada platform X/Twitter terkait isu politik, khususnya dalam konteks Putusan Mahkamah Konstitusi No. 60/PUU-XXII/2024 [8]. Hasil penelitian Malasari dan Ramli juga menunjukkan bahwa integrasi BERT dan LSTM dapat meningkatkan kemampuan model dalam mengidentifikasi sentimen pada media sosial [10].

Meskipun berbagai penelitian telah menerapkan Deep Learning, sebagian besar masih berfokus pada topik umum seperti komentar YouTube atau isu politik lain [2], [9]. Penelitian yang secara khusus mengkaji sentimen masyarakat terhadap gerakan 17+8 Tuntutan Rakyat masih sangat terbatas [12], [13], [15]. Selain gap topik, terdapat gap metodologis yaitu dominasi evaluasi berbasis akurasi agregat. Pada dataset yang sangat tidak seimbang (imbalanced dataset), nilai akurasi rata-rata

dapat menimbulkan ilusi performa (accuracy paradox) yang menyembunyikan kegagalan model dalam mengenali kelas minoritas.

Pertanyaan penelitian dan hipotesis. Berdasarkan gap tersebut, penelitian ini menjawab pertanyaan: model manakah (LSTM atau BERT) yang memberikan performa klasifikasi lebih adil dan akurat untuk sentimen gerakan 17+8 Tuntutan Rakyat pada platform X, khususnya pada kondisi dataset terbatas dan tidak seimbang. Penelitian ini menghipotesiskan bahwa pada kondisi *imbalanced* tanpa *resampling*, LSTM berpotensi mencatatkan akurasi agregat lebih tinggi akibat bias kelas mayoritas, sedangkan BERT memiliki daya diskriminasi konteks yang lebih baik pada kelas minoritas [7]. Penelitian ini memberikan tiga kontribusi utama:

1. Penyediaan Dataset Kasus Khusus Aksi 17+8: Menghadirkan dan mendokumentasikan dataset publik baru berbahasa Indonesia mengenai isu Aksi 17+8 Tuntutan Rakyat dari platform X (Twitter) yang telah divalidasi dan dikalibrasi oleh pakar bahasa.
2. Evaluasi Komparatif LSTM vs. BERT pada Data Imbalanced: Menyajikan analisis komparatif mendalam antara model recurrent (LSTM from scratch) dan pre-trained Transformer (IndoBERT) pada teks berbahasa Indonesia yang sangat tidak seimbang (imbalanced dataset), dengan mengungkap fenomena majority-class bias dan accuracy paradox melalui pembedahan metrik per-kelas serta uji signifikansi statistik paired t-test.
3. Rekomendasi Metodologis untuk Studi Lanjutan: Memberikan panduan praktis dan rekomendasi metodologis mengenai bahaya evaluasi berbasis metrik akurasi agregat/tunggal pada analisis sentimen media sosial, pentingnya penerapan teknik penyeimbangan kelas (class weighting/SMOTE), serta pertimbangan kebutuhan infrastruktur komputasi (akselerasi GPU).

Penelitian ini memiliki keterbatasan berupa dataset yang imbalanced (73,8% netral, 17,1% negatif, 9,1% positif), penggunaan subset evaluasi 200 data (5-Fold Cross Validation) akibat keterbatasan perangkat keras CPU-only, serta belum diterapkannya metode balancing data (SMOTE/class weighting).

II. KAJIAN TEORI

Penelitian Terdahulu

Berbagai penelitian telah menerapkan metode Deep Learning untuk melakukan analisis sentimen pada data media sosial. Penelitian yang dilakukan oleh Bouchra et al. menggunakan model Deep Learning untuk menganalisis sentimen masyarakat terhadap tayangan televisi nasional berdasarkan 515.492 data dari platform X. Penelitian tersebut menunjukkan bahwa model yang digunakan mampu menghasilkan performa klasifikasi yang baik dengan akurasi mencapai 96,4%, sehingga efektif dalam mengidentifikasi kecenderungan opini masyarakat terhadap tayangan televisi nasional [14].

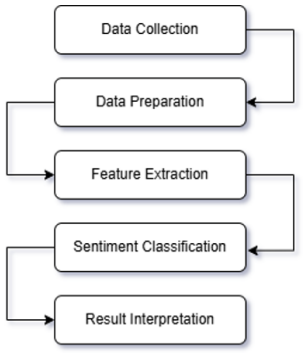
Penelitian lainnya dilakukan oleh Khadapi dan Pakpahan yang membandingkan model Long Short-Term Memory (LSTM) dan Bidirectional Encoder Representations from Transformers (BERT) pada klasifikasi sentimen diskusi Twitter mengenai Pemilu 2024 menggunakan 4.376 tweet. Hasil penelitian menunjukkan bahwa BERT memiliki kemampuan yang lebih baik dalam memahami konteks bahasa, sedangkan LSTM memberikan performa yang lebih stabil pada proses klasifikasi sentimen [8].

Selanjutnya, Wicaksono dan Nastiti menerapkan metode LSTM dan Bi-LSTM untuk menganalisis komentar pada kanal YouTube Indonesia Lawyers Club. Penelitian tersebut menghasilkan akurasi sebesar 99,61% menggunakan LSTM dan 98,11% menggunakan Bi-LSTM, sehingga menunjukkan bahwa model berbasis jaringan saraf berulang mampu memberikan performa yang sangat baik dalam klasifikasi sentimen [15].

Berdasarkan penelitian-penelitian tersebut dapat disimpulkan bahwa LSTM memiliki keunggulan dalam memproses data sekuensial, sedangkan BERT lebih unggul dalam memahami konteks semantik suatu kalimat. Meskipun demikian, penelitian yang menerapkan kedua model tersebut untuk menganalisis sentimen masyarakat terhadap Gerakan 17+8 masih sangat terbatas sehingga penelitian ini dilakukan untuk mengisi kekosongan tersebut.

Analisis Sentimen

Analisis sentimen merupakan salah satu penerapan Natural Language Processing (NLP) yang bertujuan mengidentifikasi, mengekstraksi, serta mengelompokkan opini atau emosi yang terkandung dalam suatu teks ke dalam kategori positif, negatif, maupun netral. Teknik ini banyak dimanfaatkan untuk menganalisis opini masyarakat pada media sosial karena mampu mengolah data dalam jumlah besar secara otomatis sehingga dapat memberikan gambaran mengenai kecenderungan persepsi publik terhadap suatu isu [2].



Gambar 1. Sistem Analisis Sentimen

Gambar tersebut menunjukkan tahapan analisis sentimen yang terdiri atas proses pengumpulan data (*data collection*), prapemrosesan (*data preparation*), ekstraksi fitur (*feature extraction*), klasifikasi sentimen (*sentiment classification*), dan interpretasi hasil (*result interpretation*). Tahapan tersebut dilakukan secara berurutan agar informasi yang dihasilkan lebih akurat dan mudah dianalisis [2].

Aksi 17+8

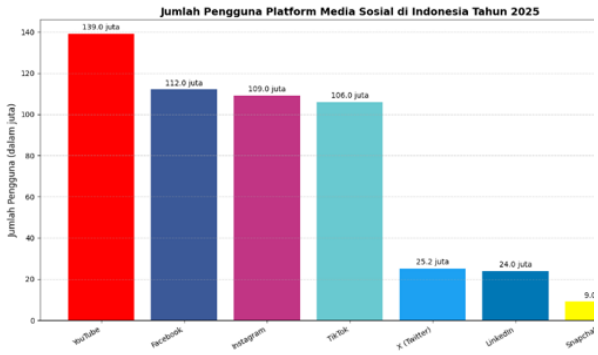
Gerakan 17+8 Tuntutan Rakyat merupakan aksi demonstrasi yang muncul pada Agustus 2025 sebagai bentuk penyampaian aspirasi masyarakat terhadap berbagai kebijakan pemerintah. Gerakan ini berkembang melalui demonstrasi di berbagai daerah dan mendapat perhatian

luas di media sosial sehingga memicu diskusi publik dalam skala nasional [10].

Perhatian masyarakat semakin meningkat setelah terjadinya peristiwa meninggalnya Affan Kurniawan saat pengamanan aksi demonstrasi. Kasus tersebut menjadi salah satu topik yang banyak diperbincangkan di berbagai media nasional maupun media sosial sehingga memunculkan beragam opini dari masyarakat.

Media Sosial

Media sosial X (sebelumnya Twitter) merupakan salah satu platform yang banyak digunakan masyarakat untuk menyampaikan informasi, berdiskusi, serta memberikan tanggapan terhadap berbagai isu publik. Fitur seperti reply, repost, dan quote memungkinkan penyebaran informasi berlangsung dengan cepat sehingga menjadikan platform ini sebagai salah satu sumber data yang banyak dimanfaatkan dalam penelitian analisis sentimen [8].



Gambar 2. Grafik Jumlah Pengguna Twitter di Indonesia Tahun 2025

Berdasarkan laporan DataReportal, jumlah pengguna X di Indonesia mencapai sekitar 25,2 juta pengguna pada tahun 2025. Jumlah tersebut menunjukkan bahwa platform ini masih relevan sebagai sumber data dalam penelitian analisis sentimen. Penelitian Khadapi dan Pakpahan juga membuktikan bahwa data Twitter/X mampu memberikan gambaran mengenai opini masyarakat terhadap suatu isu secara efektif [8].

Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang kecerdasan buatan yang mempelajari bagaimana komputer dapat memahami, mengolah, serta menghasilkan bahasa manusia secara otomatis. NLP memungkinkan komputer melakukan berbagai tugas seperti klasifikasi teks, analisis sentimen, penerjemahan bahasa, hingga ekstraksi informasi dari dokumen teks. Dalam penelitian ini, NLP digunakan sebagai dasar untuk mengolah data tweet sebelum dilakukan proses klasifikasi sentimen [1].

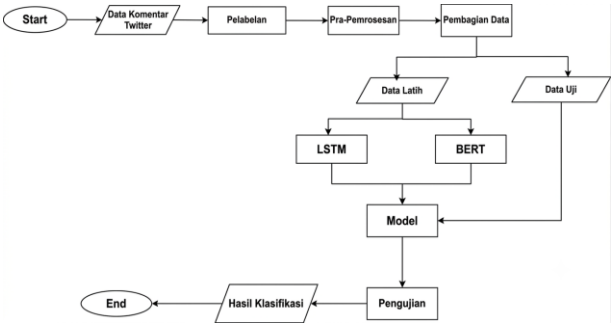
Deep Learning

Deep Learning merupakan salah satu cabang Machine Learning yang memanfaatkan jaringan saraf tiruan berlapis (deep neural network) untuk mempelajari pola dari data secara otomatis. Metode ini banyak digunakan dalam pengolahan teks karena mampu menangkap hubungan yang kompleks antar kata dibandingkan metode konvensional [4].

Penelitian ini menggunakan dua model Deep Learning, yaitu Long Short-Term Memory (LSTM) dan Bidirectional Encoder Representations from Transformers (BERT). LSTM merupakan pengembangan dari Recurrent Neural Network (RNN) yang dirancang untuk mengatasi masalah long-term dependency melalui mekanisme gate, sehingga efektif dalam memproses data sekuensial [5]. Sementara itu, BERT merupakan model berbasis arsitektur Transformer yang memanfaatkan mekanisme self-attention untuk memahami hubungan antar kata secara dua arah (bidirectional) sehingga mampu menghasilkan representasi bahasa yang lebih kontekstual. Kemampuan tersebut menjadikan BERT sebagai salah satu model yang banyak digunakan dalam berbagai tugas NLP, termasuk analisis sentimen [7].

III. METODE

Penelitian ini bertujuan mengklasifikasikan sentimen masyarakat terhadap isu "17+8 Tuntutan Rakyat" di media sosial X (Twitter) menggunakan metode Deep Learning dengan membandingkan model Long Short-Term Memory (LSTM) dan Bidirectional Encoder Representations from Transformers (BERT).



Gambar 3. Tahapan Penelitian

Secara umum, alur dan tahapan penelitian ini mengikuti lima proses utama yang saling terintegrasi seperti yang ditunjukkan pada Gambar 1, meliputi pengumpulan data (*Data Collection*), prapemrosesan data (*Data Preparation*), ekstraksi fitur (*Feature Extraction*), klasifikasi sentimen (*Sentiment Classification*), dan interpretasi hasil (*Result Interpretation*).

Pengumpulan Data (Data Collection)

Data yang digunakan dalam penelitian ini berupa teks unggahan (*tweet*) berbahasa Indonesia dari platform media sosial X (Twitter). Pengambilan data dilakukan selama periode Agustus hingga September 2025 menggunakan kombinasi *X API* dan pustaka *Tweet Harvest*. Kata kunci (*keywords*) dan tagar (*hashtags*) yang digunakan mencakup #17Plus8, #WargaJagaWarga, dan #TuntutanRakyat. Dari proses ekstraksi tersebut, diperoleh dataset mentah sebanyak 1.000 data tweet.

Pelabelan Data dan Prapemrosesan Teks (Data Preparation)

Proses Pelabelan data sentimen dilakukan menggunakan pendekatan *hybrid* (kombinasi leksikon dan validasi pakar:

1. Pelabelan Awal (Lexicon InSet): Dataset dialokasikan ke dalam tiga kelas sentimen (Positif, Netral, Negatif) menggunakan leksikon *InSet* berdasarkan bobot polaritas kata.

2. Validasi Pakar (Human Annotator): Seluruh 1.000 data hasil pelabelan leksikon divalidasi dan dikoreksi secara manual oleh 1 (satu) orang pakar bahasa untuk memastikan akurasi konteks sosial-politik dan bahasa informal/sarkasme.
3. Uji Kesesuaian (Cohen's Kappa): Untuk mengukur tingkat kesesuaian antara hasil pelabelan otomatis berbasis leksikon dan pelabelan pakar, dilakukan pengujian statistik *Cohen's Kappa* (κ). Hasil perhitungan menunjukkan nilai $\kappa = 0{,}04$. Berdasarkan skala Landis & Koch, nilai ini masuk dalam kategori *slight agreement* (kesesuaian sangat rendah). Nilai Kappa yang sangat rendah ini membuktikan secara empiris bahwa pelabelan otomatis berbasis leksikon tidak mampu menangkap konteks bahasa informal, singkatan, dan emosi publik pada media sosial, sehingga pembetulan oleh pakar bahasa menjadi tahap yang mutlak diperlukan.

Tabel 1. Distribusi Kelas Dataset Sentiment

Kelas Sentimen	Jumlah Tweet	Presentase (%)
Netral	738	73,8%
Negatif	171	17,1
Positif	91	9,1%
Total	1000	100,0%

Mengingat arsitektur LSTM dan BERT memiliki karakteristik pemrosesan teks yang berbeda, tahapan *preprocessing* disesuaikan untuk masing-masing model:

1. Preprocessing untuk LSTM: Meliputi *cleansing* (menghapus URL, @mention, #tagar, angka, dan karakter spesial), *case folding* (mengubah ke huruf kecil), normalisasi kata *slang*/singkatan (sesuai kamus KBBI), *stopword removal* (menggunakan Sastrawi), *stemming* (algoritma Nazief-Adriani), serta *tokenization & padding* ($max\_length = 64$).
2. Preprocessing untuk BERT: Meliputi *cleansing* sederhana, tanpa stemming dan tanpa stopword removal agar struktur dan konteks kalimat tetap utuh. Teks kemudian diproses menggunakan *AutoTokenizer* berbasis *WordPiece* dari Hugging Face (indobenchmark/indobert-lite-base-p1) dengan menyisipkan token [CLS] dan [SEP] serta *padding* ($max\_length = 64$).

Arsitektur dan Konfigurasi Model (*Sentiment Classification*)

Model pertama adalah LSTM yang dibangun dari awal (from scratch), mewakili pendekatan Recurrent Neural Network (RNN) untuk mengolah data sekuensial. Arsitekturnya terdiri dari Embedding Layer, Bidirectional LSTM Layer (dua arah), Dropout Layer, dan Fully Connected Layer sebagai lapisan keluaran 3 kelas sentimen. Representasi teks pada LSTM menggunakan vocabulary dan integer encoding hasil preprocessing dan stemming. Secara matematis, unit LSTM bekerja melalui mekanisme gerbang (gate):

Forget gate:

$$f\_t = \sigma(W\_f \cdot [h\_(t-1), x\_t] + b\_f) \dots (1)$$

Input gate:

$$i\_t = \sigma(W\_i \cdot [h\_(t-1), x\_t] + b\_i) \dots (2)$$

Kandidat cell state:

$$\tilde{C}\_t = tanh(W\_C \cdot [h\_(t-1), x\_t] + b\_C) \dots (3)$$

Pembaruan cell state:

$$C\_t = f\_t * C\_(t-1) + i\_t * \tilde{C}\_t \dots (4)$$

Output gate:

$$o\_t = \sigma(W\_o \cdot [h\_(t-1), x\_t] + b\_o) \dots (5)$$

Hidden state:

$$h\_t = o\_t * \tanh(C\_t) \dots (6)$$

di mana σ adalah fungsi sigmoid, W dan b adalah bobot dan bias tiap gerbang, sedangkan f_t, i_t, dan o_t masing-masing adalah forget, input, dan output gate yang mengatur informasi yang disimpan atau dilupakan pada cell state C_t.

Model kedua memanfaatkan IndoBERT Lite Base P1 (indobenchmark/indobert-lite-base-p1), model pre-trained berbasis Transformer yang dikembangkan khusus untuk Bahasa Indonesia, diaplikasikan melalui proses fine-tuning. Kemampuan BERT memahami konteks dua arah diperoleh melalui mekanisme self-attention:

$$\text{Attention}(Q,K,V) = \text{softmax}(QK^T / \sqrt{d_k})\, V \dots (7)$$

Di mana Q, K, dan V adalah representasi Query, Key, dan Value dari input, serta d_k adalah dimensi Key sebagai faktor penskalaan.

Tabel 2 Ringkasan Hyperparameter Pelatihan Model.

Hyperparameter	Modul LSTM	Model BERT (Indo BERT)
Optimizer	Adam	Adam
Learning Rate	0,001	5×10^{-5}
Loss Function	CrossEntropyLoss	CrossEntropyLoss
Batch Size	32	32
Max Epochs/Early Stopping	5 (Patience = 7)	5 (Patience = 7)
Max Sequence Length	64	64

Kedua model dilatih menggunakan perangkat laptop Acer Aspire Go 14 dengan prosesor Intel Core i3-N305 (1,80 GHz), RAM 8 GB, dan penyimpanan SSD 512 GB, tanpa akselerasi GPU khusus. Keterbatasan perangkat keras ini menjadi faktor utama dibatasinya jumlah epoch dan penggunaan subset data pengujian agar proses pelatihan berjalan stabil pada komputasi CPU.

Validasi dan Evaluasi Model (*Result Interpretation*)

Untuk pembagian data dan validasi performa, penelitian ini menerapkan skema 5-Fold Cross Validation menggunakan Stratified K-Fold guna menjaga proporsi distribusi kelas tetap seimbang pada tiap lipatan. Untuk mengakomodasi keterbatasan komputasi CPU-only, pengujian komparatif 5-Fold Cross Validation dijalankan pada subset 200 data yang dipilih secara acak terstratifikasi dari total 1.000 data. Pada setiap iterasi, subset dibagi menjadi 80% (160 data) sebagai data latih dan 20% (40 data) sebagai data uji.

Mengingat distribusi kelas pada dataset penelitian ini bersifat imbalanced dengan sentimen netral mendominasi sebanyak 73,8% dari total data, diikuti sentimen negatif 17,1% dan positif hanya 9,1% penelitian ini tidak menerapkan teknik penanganan ketidakseimbangan data

tambahan seperti class weighting, oversampling, maupun SMOTE. Satu-satunya mekanisme yang diterapkan untuk memitigasi dampak imbalance adalah penggunaan Stratified K-Fold, yang menjaga agar proporsi kelas pada setiap fold tetap representatif terhadap proporsi kelas keseluruhan, meskipun tidak mengubah komposisi kelas itu sendiri. Hal ini diakui sebagai salah satu keterbatasan penelitian yang turut memengaruhi performa model, khususnya rendahnya nilai precision pada kelas minoritas (positif dan negatif), dan menjadi salah satu arah pengembangan pada penelitian selanjutnya.

Performa akhir dari model LSTM dan BERT kemudian dievaluasi serta dibandingkan secara komparatif berdasarkan metrik evaluasi standar confusion matrix, yang meliputi Accuracy, Precision, Recall, dan F1-Score, dengan rumus sebagai berikut:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \dots (8)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \dots (9)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \dots (10)$$

$$\text{F1} - \text{Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \dots (11)$$

di mana TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative) dihitung per-kelas sentimen. Selain metrik evaluasi per-kelas, dilakukan pula Uji Statistik Paired t-Test untuk mengetahui apakah perbedaan nilai akurasi rata-rata antara LSTM dan BERT bersifat signifikan secara statistik ($\alpha = 0{,}05$).

IV. HASIL DAN PEMBAHASAN

Distribusi Kelas Dataset

Penelitian ini menggunakan dataset yang diperoleh melalui proses *crawling* pada platform X (Twitter) terkait isu Aksi 17+8 Tuntutan Rakyat dengan total sebanyak 1.000 komentar. Sebelum memasuki tahap pemodelan, dataset terlebih dahulu melalui proses pelabelan sentimen ke dalam tiga kategori, yaitu positif, netral, dan negatif. Hasil distribusi data untuk masing-masing kelas sentimen ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Data per Kelas Sentimen

Kelas Label	Jumlah Data	Presentase (%)
Positif	91	9,1%
Netral	738	73,8%
Negatif	171	17,1%
Total	1000	100%

Berdasarkan Tabel 3, terlihat bahwa distribusi kelas pada dataset bersifat sangat tidak seimbang (*highly imbalanced dataset*), di mana sentimen netral mendominasi sebesar 73,8% (738 data) dari total 1.000 komentar. Dominasi sentimen netral menunjukkan bahwa sebagian besar pengguna platform X memberikan tanggapan yang bersifat informatif, deskriptif, atau tidak secara eksplisit menunjukkan kecenderungan mendukung maupun menolak. Sementara itu, sentimen negatif berjumlah 17,1% (171 data) dan sentimen positif hanya 9,1% (91 data).

Pengujian dan Evaluasi Peforma Model

Evaluasi performa model Long Short-Term Memory (LSTM) dan Bidirectional Encoder Representations from Transformers (BERT) dilakukan menggunakan skema 5-Fold Cross Validation. Karena keterbatasan perangkat keras CPU-only, proses pelatihan dan evaluasi komparatif dijalankan pada subset 200 data yang dipilih secara acak terstratifikasi dari total 1.000 data. Pada setiap iterasi, subset dibagi menjadi 80% (160 data) sebagai data latih dan 20% (40 data) sebagai data uji.

Ringkasan proses pelatihan per-fold untuk LSTM dan BERT disajikan pada Tabel 4 dan Tabel 5, sedangkan rekapitulasi evaluasi rata-rata kedua model ditunjukkan pada Tabel 6 dan Tabel 7.

Tabel 4. Ringkasan Training per Fold LSTM

Fold	Train Acc	Val Acc	Train Loss	Val Loss	Best Epoch
1	56.25%	67.50%	0.7141	0.8695	4
2	47.50%	70.00%	0.7854	0.7314	4
3	60.62%	70.00%	0.6760	0.7562	5
4	57.50%	70.00%	0.6904	0.8048	5
5	51.25%	67.50%	0.6919	0.7738	5

Tabel 5. Ringkasan Training per fold BERT

Fold	Train Acc	Val. Acc	Train Loss	Val. Loss	Best Epoch
1	21.25%	35.00%	1.0984	1.0205	5
2	56.88%	70.00%	0.8653	0.8084	5
3	25.00%	70.00%	0.9502	0.8994	5
4	20.63%	42.50%	1.0599	0.9901	5
5	22.50%	52.50%	1.0005	0.9666	5

Tabel 6. Hasil Evaluasi Model LSTM

Fold	Accuracy	Precision	Recall	F1-Score
1	0.6750	0.4556	0.6750	0.5440
2	0.7000	0.4900	0.7000	0.5765
3	0.7000	0.4900	0.7000	0.5765
4	0.7000	0.4900	0.7000	0.5765
5	0.6750	0.4556	0.6750	0.5440
Rata-rata	**0.6900**	**0.4763**	**0.6900**	**0.5635**

Tabel 7. Hasil Evaluasi Model BERT

Fold	Accuracy	Precision	Recall	F1-Score
1	0.3500	0.4594	0.3500	0.3685
2	0.7000	0.4900	0.7000	0.5765
3	0.7000	0.4900	0.7000	0.5765
4	0.4250	0.6250	0.4250	0.4669
5	0.5250	0.5886	0.5250	0.5133
Rata-rata	**0.5400**	**0.5306**	**0.5400**	**0.5003**

Berdasarkan metrik rata-rata agregat (Tabel 6 dan 7), LSTM mencatatkan Accuracy (0,6900) dan F1-Score (0,5635) yang lebih tinggi dibandingkan BERT (Accuracy 0,5400 dan F1-Score 0,5003). Namun, penilai kinerja tidak boleh hanya mengandalkan metrik agregat semata pada dataset yang tidak seimbang (accuracy paradox).

Evaluasi Metrik per Kelas dan Confusion Matrix

Untuk memahami performa nyata dari kedua arsitektur dalam memprediksi tiap kategori sentimen, Tabel 8 menyajikan rincian metrik per kelas beserta jumlah data aktual (*support*) pada subset pengujian 200 data.

Tabel 8. Metrik Evaluasi per Kelas Sentimen pada Subset 200 Data Uji

Model	Kelas	Precision	Recall	F1-Score	Jumlah Data
LSTM	Positif	0,0000	0,0000	0,0000	21
LSTM	Netral	0,6900	1,0000	0,8166	138
LSTM	Negatif	0,0000	0,0000	0,0000	41
BERT	Positif	0,1250	0,1429	0,1333	21
BERT	Netral	0,7231	0,6812	0,7015	138
BERT	Negatif	0,2391	0,2683	0,2529	41

Hasil pada Tabel 6 mengungkap temuan penting yang tidak tampak pada metrik rata-rata: model LSTM memperoleh nilai precision, recall, dan F1-score sebesar 0,0000 pada kelas positif maupun negatif, yang berarti LSTM tidak berhasil mengenali satu pun data pada kedua kelas minoritas dari seluruh 200 data uji, dan secara konsisten memprediksi seluruh data sebagai kelas netral. Sebaliknya, model BERT berhasil mengenali sebagian data pada kelas negatif (recall 26,83%) dan kelas positif (recall 14,29%), menunjukkan kemampuan diskriminasi antar kelas yang lebih baik meskipun belum optimal. Peta pergeseran dan kesalahan prediksi dari seluruh 200 data uji (hasil akumulasi dari 5 fold) disajikan dalam bentuk *Confusion Matrix* pada Tabel 9 dan Tabel 10.

Tabel 9. Confusion Matrix Akumulatif LSTM

Aktual\Prediksi	Negatif	Netral	Positif
Negatif	0	41	0
Netral	0	138	0
Positif	0	21	0

Tabel 10. Confusion Matrix Akumulatif BERT

Aktual\Prediksi	Negatif	Netral	Positif
Negatif	11	23	7
Netral	30	94	14
Positif	5	13	3

Analisis confusion matrix akumulatif (Tabel 7 dan Tabel 8) menunjukkan pola kesalahan yang kontras antara kedua model. Model LSTM secara konsisten memprediksi seluruh data sebagai kelas netral pada kelima fold pengujian, tanpa terkecuali (Tabel 7), mengindikasikan model mempelajari distribusi kelas mayoritas (*majority-class bias*) alih-alih karakteristik linguistik yang membedakan sentimen secara substantif. Model BERT menunjukkan pola prediksi yang lebih bervariasi (Tabel 8), meskipun tetap didominasi kesalahan ke arah kelas netral.

Analisis Penyebab Bias Mayoritas pada LSTM dan Perbandingan Kinerja

Ketidakseimbangan kelas pada dataset penelitian ini (netral 73,8%, negatif 17,1%, positif 9,1%), tanpa disertai teknik penyeimbangan seperti class weighting atau SMOTE, menjadi faktor utama yang membentuk perilaku kedua model dalam penelitian ini. Model LSTM merespons ketidakseimbangan kelas dengan strategi paling sederhana, yaitu memprediksi kelas mayoritas untuk seluruh data, sebagaimana ditunjukkan pada Tabel 6 dan analisis confusion matrix. Strategi ini menghasilkan accuracy dan recall agregat yang tampak tinggi (69,00%), namun tidak mencerminkan kemampuan klasifikasi sentimen yang sesungguhnya, mengingat precision, recall, dan F1-score pada kelas positif dan negatif sama-sama bernilai 0,0000. Sebaliknya, model BERT, meskipun mencatat accuracy agregat lebih rendah (54,00%), menunjukkan upaya nyata dalam membedakan ketiga kelas sentimen.

Performa BERT yang belum optimal turut dipengaruhi oleh beberapa faktor teknis, yaitu keterbatasan jumlah data pada kelas minoritas dalam subset 200 data yang digunakan untuk evaluasi (rata-rata hanya 8 data negatif dan 4 data positif per fold pengujian), pembatasan jumlah epoch pelatihan (maksimum 5 epoch), serta proses pelatihan yang dilakukan tanpa akselerasi GPU (lihat Bagian III.C). Ketiga faktor ini secara umum diketahui dapat membatasi kemampuan model Transformer seperti BERT dalam mencapai konvergensi optimal selama proses fine-tuning. Tabel 8 dan Tabel 9 mengungkap fenomena *majority-class bias* yang ekstrem pada model LSTM. LSTM memperoleh nilai *Precision*, *Recall*, dan *F1-Score* sebesar 0,0000 pada kelas minoritas (Positif dan Negatif). Model ini secara konsisten memprediksi seluruh 200 data uji sebagai kelas Netral tanpa terkecuali. Ada tiga faktor utama yang menyebabkan LSTM memprediksi seluruh data sebagai kelas mayoritas (Netral):

1. Dampak Ketidakseimbangan Data Tanpa Resampling: Dataset didominasi oleh kelas Netral (69% pada subset uji / 73,8% pada dataset utuh). Tanpa teknik penanganan ketidakseimbangan data seperti SMOTE, *oversampling*, atau penerapan *class weighting* pada fungsi *CrossEntropyLoss*, fungsi kerugian (*loss function*) akan teroptimasi paling cepat jika model menebak kelas mayoritas.
2. Karakteristik Preprocessing dan *Word Embeddings* yang Dibangun dari Nol: Pada model LSTM *from scratch*, kata-kata dilatih menggunakan *embedding layer* acak yang dikombinasikan dengan *stemming* kaku (Nazief-Adriani). Proses *stemming* dan *stopword removal* sering kali menghilangkan kata tugas atau imbuhan penting yang membawa nuansa emosi/sarkasme pada teks bahasa informal.
3. Mekanisme *Loss Minimization* Akibat Data Minoritas Terlalu Sedikit: Dengan jumlah *support* kelas Positif yang hanya 21 data dan Negatif 41 data, gradien pembaruan bobot (*weight updates*) pada LSTM terdominasi oleh sampel Netral. Tebakan buta (*blind guess*) pada kelas mayoritas memberikan *loss* rata-rata terkecil bagi optimizer Adam.

Sebaliknya, meskipun BERT menghasilkan *Accuracy* rata-rata lebih rendah (0,5400), arsitektur Transformer pada IndoBERT terbukti memiliki daya diskriminasi (*contextual sensitivity*) terhadap kelas minoritas. BERT berhasil mengenali 11 dari 41 sampel negatif (*Recall* 26,83%) dan 3 dari 21 sampel positif (*Recall* 14,29%). BERT memanfaatkan *WordPiece Tokenizer* tanpa menghapus *stopwords* atau melakukan *stemming*, sehingga mampu mempertahankan struktur sintaksis dan hubungan semantik antar kata dua arah (*bidirectional*) secara lebih utuh.

Analisis Contoh Kesalahan Klasifikasi (*Error Analysis*)

Untuk memberikan gambaran kualitatif mengenai kega-

galan prediksi kedua model, Tabel 11 menampilkan beberapa contoh data uji nyata beserta hasil prediksi dari LSTM dan BERT.

Tabel 11. Contoh Kesalahan Klasifikasi Teks Tweet oleh LSTM dan BERT

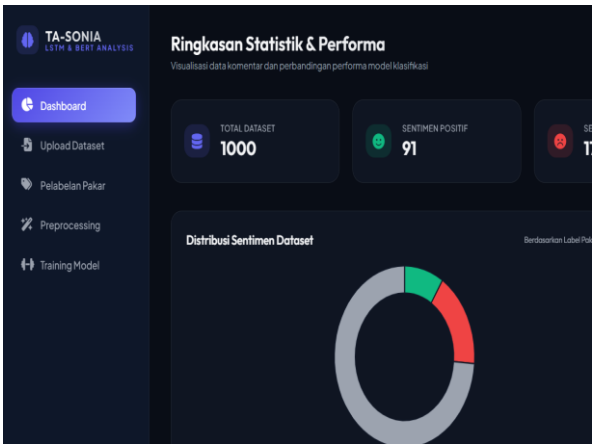
Teks Tweet Original/ Preprocessed	Label Pakar (Aktual)	Prediksi LSTM	Prediksi BERT	Analisis Kesalahan Prediksi
"17 tuntut jangka pendek 17+8 tuntut rakyat harus laku maksimal 5 september 2025"	Netral	Netral	Negatif	Over-sensitivity BERT: Kata *"tuntut"* dan pembatasan tanggal diinterpretasikan oleh BERT sebagai bentuk tekanan/kritik (negatif), padahal pakar menilai kalimat ini berupa kutipan pernyataan normatif (netral).
"gerakan 17+8 tuntut rakyat kait apa kabar wakil mid class kalau terus terus bakal kuat"	Negatif	Netral	Netral	Sarkasme dan Bahasa Informal: Kalimat mengandung kritikan tersirat/sarkasme sosial. Kedua model gagal mengidentifikasi ketiadaan kata sentimen eksplisit, sehingga terprediksi sebagai netral.
"aksi 17+8 tuntut rakyat wujud nyata kepedulian bersama warga jaga warga"	Positif	Netral	Positif	Kemampuan Kontekstual BERT: Kata *"wujud nyata kepedulian"* dan tagar *"warga jaga warga"* berhasil ditangkap oleh mekanisme *self-attention* BERT sebagai dukungan (positif), sedangkan LSTM terbias memprediksi netral.

Interpretasi Uji Signifikan Statistik (Paired t-Test)

Untuk menguji apakah perbedaan *Accuracy* rata-rata antara LSTM (0,6900) dan BERT (0,5400) terbukti signifikan secara statistik, dilakukan pengujian *Paired t-Test* terhadap nilai akurasi pada kelima fold pengujian. Pengujian ini menghasilkan nilai statistik hitung $t = 2{,}221$ dan p-value = 0,0905. Dengan mempertimbangkan hasil evaluasi per kelas (Tabel 6), analisis confusion matrix, dan uji signifikansi statistik (p-value = 0,0905) secara bersamaan, penelitian ini tidak dapat menyimpulkan bahwa LSTM lebih unggul dibandingkan BERT untuk tugas klasifikasi sentimen pada dataset ini. Metrik akurasi dan recall rata-rata semata terbukti dapat menyesatkan interpretasi performa model pada kondisi dataset yang sangat tidak seimbang. Sebaliknya, hasil penelitian ini mengindikasikan bahwa BERT berpotensi menjadi pilihan yang lebih baik untuk tugas klasifikasi sentimen semacam ini, apabila ketidakseimbangan data ditangani secara memadai pada penelitian selanjutnya. Interpretasi metodologis terhadap hasil uji statistik ini adalah sebagai berikut: a) Tidak Ada Perbedaan Signifikan secara Statistik ($\alpha = 0{,}05$): Karena nilai p-value (0,0905) lebih besar dari taraf signifikansi standar ($\alpha = 0{,}05$), hipotesis nol (H_0) gagal ditolak. Hal ini membuktikan bahwa secara statistik, tidak ada perbedaan kinerja akurasi yang signifikan antara LSTM dan BERT pada dataset ini. b) Keunggulan Semu Akurasi LSTM: Keunggulan nilai akurasi rata-rata LSTM sebesar 0,6900 dibanding BERT 0,5400 tidak boleh disimpulkan sebagai keunggulan performa arsitektur yang nyata. Angka ini murni merupakan implikasi dari keterbatasan ukuran sampel evaluasi ($n = 5$ fold) dan bias prediksi kelas mayoritas.c) Pentingnya Evaluasi Komprehensif: Hasil uji statistik ini mempertegas saran metodologis bahwa mengevaluasi model *Deep Learning* pada data media sosial yang tidak seimbang tidak boleh hanya bersandar pada metrik tunggal atau rata-rata agregat. Pembagian metrik per kelas dan analisis uji beda statistik mutlak diperlukan untuk menghindari penarikan kesimpulan yang menyesatkan (*false superiority*).

Implementasi Sistem

1. Tampilan Dashboard



Gambar 4. Tampilan Dashboard Sistem

Tampilan Dashboard merupakan halaman utama pada sistem yang berfungsi untuk menyajikan ringkasan informasi mengenai dataset, hasil analisis sentimen, proses pelabelan , serta performa model klasifikasi. Halaman ini dirancang dengan antarmuka yang sederhana, informatif, dan mudah dipahami sehingga pengguna dapat memantau seluruh proses analisis sentimen secara cepat tanpa harus membuka setiap menu secara terpisah

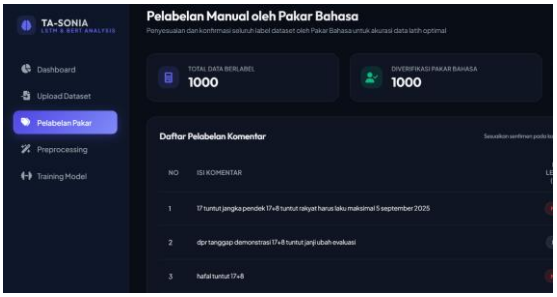
2. Tampilan Upload Dataset

Gambar 5. Tampilan Halaman Upload Dataset

Tampilan Upload Dataset merupakan fitur yang digunakan untuk mengelola dan mengunggah dataset yang telah diberi label sentimen ke dalam sistem. Halaman ini dirancang agar pengguna dapat memasukkan dataset penelitian dalam format Excel (.xlsx, .xls) atau CSV (.csv) dengan mudah melalui mekanisme drag and drop maupun pemilihan berkas dari komputer. Selain mendukung proses unggah dataset, halaman ini juga menampilkan ringkasan statistik dataset yang sedang aktif, meliputi jumlah total data, distribusi kelas sentimen positif, negatif, dan

netral, serta visualisasi distribusi data dalam bentuk diagram.

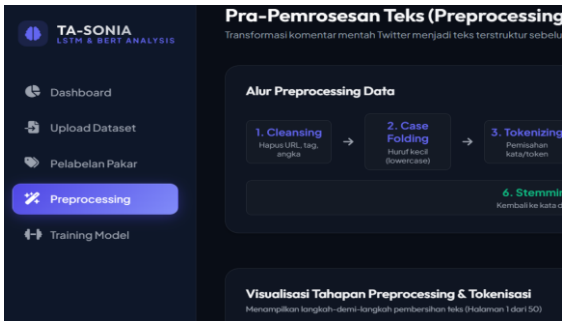
3. Pelabelan



Gambar 6. Tampilan Halaman Pelabelan

Tampilan Pelabelan merupakan fitur untuk memberikan label sentimen pada data komentar menggunakan metode Hybrid, yaitu kombinasi pelabelan otomatis dengan Lexicon InSet dan verifikasi manual oleh ahli bahasa. Halaman ini bertujuan meningkatkan kualitas label dataset agar lebih akurat, serta menyediakan fasilitas bagi pengguna untuk mengubah label jika hasil pelabelan otomatis tidak sesuai.

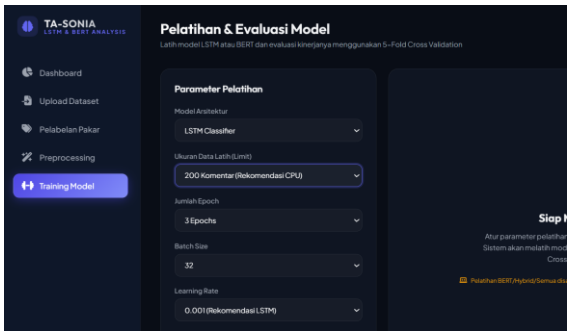
4. Tampilan Preprocesing



Gambar 7. Tampilan Halaman Preprocesing

Tampilan Preprocessing merupakan fitur yang digunakan untuk melakukan prapemrosesan data komentar hasil crawling sebelum digunakan pada proses pelabelan dan pelatihan model klasifikasi. Tahapan preprocessing bertujuan untuk membersihkan data dari karakter yang tidak diperlukan, menormalkan kata, serta mengubah komentar menjadi teks yang lebih terstruktur sehingga dapat meningkatkan kualitas data yang akan diproses oleh model LSTM-BERT.

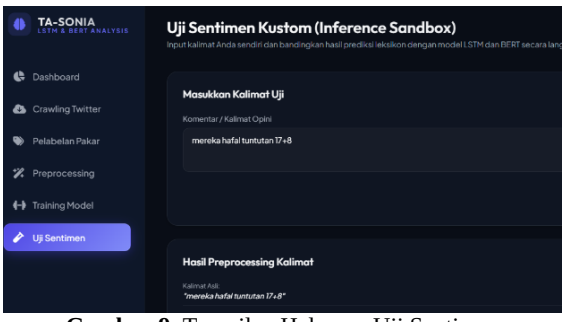
5. Tampilan Pelatihan dan Hasil Evaluasi Model



Gambar 8. Tampilan Halaman Pelatihan

Tampilan Pelatihan dan Evaluasi Model merupakan fitur yang digunakan untuk melatih model klasifikasi sentimen berdasarkan data yang telah melalui proses preprocessing dan pelabelan. Pada halaman ini, pengguna dapat menentukan parameter pelatihan, memilih arsitektur model yang akan digunakan, serta melihat hasil evaluasi kinerja model menggunakan metode 5-Fold Cross Validation. Selain itu, sistem juga menampilkan perbandingan performa antar model sehingga pengguna dapat mengetahui model yang memiliki hasil terbaik.

6. Tampilan Uji Sentimen



Gambar 9. Tampilan Halaman Uji Sentimen

Tampilan Uji Sentimen (Inference Sandbox) merupakan fitur yang digunakan untuk menguji hasil klasifikasi sentimen terhadap kalimat atau komentar yang dimasukkan secara langsung oleh pengguna. Halaman ini memungkinkan pengguna membandingkan hasil analisis menggunakan metode Lexicon InSet, LSTM, BERT, dan Hybrid BERT-LSTM sehingga pengguna dapat melihat perbedaan hasil prediksi dari setiap metode. Selain menampilkan hasil prediksi, sistem juga memperlihatkan tahapan preprocessing yang dilakukan terhadap kalimat masukan sebelum diproses oleh model pembelajaran mesin.

V. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai analisis sentimen publik di platform X (Twitter) terkait isu Aksi 17+8 Tuntutan Rakyat menggunakan pendekatan deep learning, diperoleh beberapa temuan utama. Dataset opini masyarakat sebanyak 1.000 entri menunjukkan ketidakseimbangan kelas yang signifikan, dengan dominasi sentimen netral sebesar 73,8% (738 entri), disusul sentimen negatif 17,1% (171 entri) dan sentimen positif 9,1% (91 entri).

Pengujian menggunakan 5-Fold Cross Validation pada subset 200 entri menunjukkan bahwa model Long Short-Term Memory (LSTM) mencapai rata-rata accuracy 0,6900, precision 0,4763, recall 0,6900, dan F1-score 0,5635, sedangkan model BERT memperoleh accuracy 0,5400, precision 0,5306, recall 0,5400, dan F1-score 0,5003. Secara agregat metrik tersebut menempatkan LSTM di atas BERT pada pengukuran akurasi dan F1-score rata-rata.

Analisis metrik per kelas dan confusion matrix mengungkapkan fenomena accuracy paradox pada LSTM akibat bias terhadap kelas mayoritas: LSTM memprediksi hampir seluruh data sebagai kelas netral sehingga precision, recall, dan F1-score untuk kelas positif dan negatif bernilai 0,0000. Sebaliknya, BERT menunjukkan kemampuan diskriminasi kontekstual yang lebih baik terhadap kelas minoritas dengan recall

26,83% untuk kelas negatif dan 14,29% untuk kelas positif. Uji signifikansi statistik (paired t-test) terhadap nilai accuracy menghasilkan p-value = 0,0905 (p > 0,05), yang menunjukkan tidak adanya perbedaan performa akurasi yang signifikan secara statistik antara LSTM dan BERT; hasil ini menegaskan keterbatasan metrik akurasi tunggal pada kondisi data yang tidak seimbang.

Penelitian ini memiliki keterbatasan utama berupa dominasi class imbalance tanpa penerapan teknik penyeimbangan data, keterbatasan infrastruktur komputasi (CPU-only tanpa akselerasi GPU), serta penggunaan subset evaluasi 200 entri yang membatasi generalisasi hasil.

Rekomendasi untuk penelitian selanjutnya adalah sebagai berikut. Pertama, terapkan teknik penyeimbangan data seperti class weighting pada fungsi loss, SMOTE (oversampling), atau undersampling untuk mengurangi bias terhadap kelas mayoritas dan meningkatkan sensitivitas terhadap kelas minoritas. Kedua, gunakan infrastruktur komputasi berbasis GPU sehingga memungkinkan fine-tuning BERT pada keseluruhan dataset (≥1.000 entri) dengan konfigurasi epoch yang lebih ekstensif untuk mencapai konvergensi yang lebih baik. Ketiga, lakukan eksperimen dengan model pembanding konvensional (misalnya SVM atau Naive Bayes) sebagai baseline komputasi ringan, serta terapkan teknik augmentasi teks atau varian pre-trained model bahasa Indonesia untuk menangani karakteristik bahasa informal dan sarkasme pada data media sosial.

REFERENSI

[1] M. Amien amien, "ELANG: Journal of Interdisciplinary Research Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia," hlm. 99–105, Agu 2023.

[2] F. Bouchra, I. M. A. D. Suarjaya, dan N. K. D. Rusjayanthi, "Analisis Sentimen Masyarakat terhadap Tayangan Televisi Nasional menggunakan Metode Deep Learning 89," Okt 2024.

[3] S. Khairunnisa, A. Adiwijaya, dan S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, hlm. 406, Apr 2021, doi: 10.30865/mib.v5i2.2835.

[4] B. Firmanto, A. S. Aziz, dan J. Sesoca, "Language models are few-shot learners," dalam *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, 2020. doi: 10.65525/svup.9788199778009.2026.224-230.

[5] A. Hasiholan, I. Cholissodin, dan N. Yudistira, "Analisis Sentimen Tweet Covid-19 Varian Omicron pada Platform Media Sosial Twitter menggunakan Metode LSTM berbasis Multi Fungsi Aktivasi dan GLOVE," 2022. [Daring]. Tersedia pada: http://j-ptiik.ub.ac.id

[6] A. S. Yazid dan E. Winarko, "Fine-Tuning BERT untuk Menangani Ambiguitas Pada POS Tagging Bahasa Indonesia," 2023.

[7] R. Illahi, S. Agustian, F. Yanto, P. H. Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau Jl Soebrantas, S. Baru, dan K. Pekanbaru, "Klasifikasi Sentimen Menggunakan Bidirectional Lstm Dan Indobert Dengan Dataset Terbatas," 2025.

[8] A. Irianti, H. Halimah, S. Sutedi, dan M. Agariana, "Integration of BERT and SVM in Sentiment Analysis of Twitter/X Regarding Constitutional Court Decision No. 60/PUU-XXII/2024," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, hlm. 469–482, Apr 2025, doi: 10.52436/1.jutif.2025.6.2.4068.

[9] B. Wicaksono dan V. R. S. Nastiti, "Analisis Sentimen dalam Opini Publik di Chanel Youtube Indonesia Lawyers Club Tentang Isu Populer dengan Menggunakan Metode LSTM dan Bi-LSTM," *Jurnal Algoritma*, vol. 21, no. 2, hlm. 241–251, Des 2024, doi: 10.33364/algoritma/v.21-2.1696.

[10] N. Malasari dan M. Ramli, "Analisis Sentimen Media Sosial Menggunakan Algoritma BERT dan LSTM," *Journal of Computer Science and Information Technology*, vol. 1, no. 3, hlm. 85–92, Des 2025, doi: 10.70716/jocsit.v1i3.318.

[11] Tempo, "Asal Muasal Gerakan 17+8 Tuntutan Rakyat," Tempo.

[12] Emir Yanwardhana, "Affan Kurniawan Tewas Dilindas Rantis Brimob, Ini Tanggapan Menko BG," CNBC Indonesia.

[13] Tempo dan Myesha Fatina Rachman, "Kronologi Pemeriksaan 7 Anggota Brimob yang Lindas Affan Kurniawan dengan Kendaraan Taktis," Tempo.

[14] BBC News Indonesia, "Demo tolak kenaikan tunjangan DPR memicu gerakan 17+8," BBC News Indonesia.

[15] Kompas dan Yefta Christopherus Asia Sanjaya, "Kronologi Tragis Affan Kurniawan: Driver Ojol Tewas Terlindas Rantis Brimob Saat Antar Orderan," Kompas.com.