

Analisis Sentimen Masyarakat terhadap Isu Korupsi Dana Bencana di Indonesia Menggunakan Metode *Bidirectional Long Short-Term Memory* (Bi-LSTM)

Toni Prabowo^{1,*}, Muhammad Irfan Sarif², Aradi Sebayang³, Tengku Didi Ferdillah⁴, and Muhammad Azuan⁵

^{1,2,3,4,5}Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: ^{1,*}toniprabowo@gmail.com, ²irfanberbagi@gmail.com, ³aradisebayang23@gmail.com,

⁴tengkudidiferdillah23@gmail.com, ⁵mhd.azuan99@gmail.com

(* Email Corresponding Author: toniprabowohsb@gmail.com)

Received: 3 Januari 2025 | Revision: 3 Januari 2025 | Accepted: 3 Januari 2025

Abstrak

Korupsi dana bencana dan bantuan sosial merupakan isu krusial yang mencederai keadilan sosial dan kepercayaan publik terhadap integritas pemerintah di Indonesia. Fenomena ini memicu gelombang opini masif di media sosial yang memerlukan analisis komputasional mendalam untuk memahami dinamika persepsi masyarakat secara objektif. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi performa algoritma *Deep Learning*, khususnya *Bidirectional Long Short-Term Memory* (Bi-LSTM), dalam mengklasifikasikan sentimen masyarakat terkait isu korupsi dana bencana. Dataset yang digunakan mencakup 1.358 data tekstual yang dikategorikan menjadi sentimen negatif, netral, dan positif, dengan dominasi signifikan pada kelas negatif sebanyak 926 data. Arsitektur model yang diusulkan mengintegrasikan lapisan *embedding* 300 dimensi, lapisan Bi-LSTM untuk menangkap konteks dua arah, serta kombinasi *Global Max Pooling* dan *Global Average Pooling* untuk ekstraksi fitur yang optimal. Hasil pengujian menunjukkan bahwa model mencapai akurasi sebesar 0,75, dengan *Weighted F1-score* 0,76 dan *Macro F1-score* 0,65. Analisis *Confusion Matrix* mengungkapkan bahwa model sangat efektif dalam mengidentifikasi sentimen negatif, namun menghadapi tantangan dalam membedakan kelas minoritas akibat ketidakseimbangan data. Temuan ini memberikan wawasan mendalam bagi pembuat kebijakan mengenai sentimen publik serta menunjukkan potensi sekaligus batasan metode Bi-LSTM dalam menangani teks berbahasa Indonesia yang bersifat informal dan sarkastik dalam konteks diskursus politik dan korupsi.

Kata Kunci: Analisis Sentimen, Bi-LSTM, Korupsi Dana Bencana, *Deep Learning*, *Natural Language Processing*

Abstract

Corruption of disaster relief funds and social assistance is a critical issue that undermines social justice and public trust in government integrity in Indonesia. This phenomenon has triggered a massive wave of opinions on social media, necessitating deep computational analysis to objectively understand public perception dynamics. This study aims to implement and evaluate the performance of a Deep Learning algorithm, specifically Bidirectional Long Short-Term Memory (Bi-LSTM), in classifying public sentiment related to the issue of disaster fund corruption. The dataset comprises 1,358 textual data points categorized into negative, neutral, and positive sentiments, with a significant dominance of the negative class (926 entries). The proposed model architecture integrates a 300-dimensional embedding layer, a Bi-LSTM layer to capture bidirectional context, and a combination of Global Max Pooling and Global Average Pooling for optimal feature extraction. The experimental results demonstrate that the model achieved an accuracy of 0.75, with a Weighted F1-score of 0.76 and a Macro F1-score of 0.65. Confusion Matrix analysis reveals that the model is highly effective in identifying negative sentiments but faces challenges in distinguishing minority classes due to data imbalance and linguistic ambiguities such as sarcasm. These findings provide deep insights for policymakers regarding public sentiment and demonstrate both the potential and limitations of the Bi-LSTM method in processing informal and sarcastic Indonesian text within the context of political and corruption discourse.

Keywords: Sentiment Analysis, Bi-LSTM, Disaster Fund Corruption, *Deep Learning*, *Natural Language Processing*

1. PENDAHULUAN

Masalah korupsi di Indonesia telah lama menjadi hambatan utama dalam pembangunan nasional dan kesejahteraan masyarakat. Di tengah situasi darurat, seperti bencana alam atau krisis kesehatan global, integritas pengelolaan dana bantuan menjadi sangat krusial karena menyangkut kelangsungan hidup kelompok masyarakat yang paling rentan. Namun, sejarah mencatat adanya ketidaksinkronan antara ekspektasi publik (*Das Sollen*) dan realitas di lapangan (*Das Sein*), di mana dana yang dialokasikan untuk penanggulangan bencana sering kali menjadi sasaran penyelewengan oleh oknum pejabat publik.[1] Kasus korupsi bantuan sosial (bansos) selama pandemi COVID-19, yang melibatkan kementerian terkait, menjadi salah satu titik balik yang memicu kemarahan kolektif masyarakat Indonesia.[1] Fenomena ini bukan sekadar masalah hukum, melainkan juga masalah komunikasi publik yang tercermin dalam diskusi di berbagai platform digital.

Media sosial, khususnya Twitter (X), telah bertransformasi menjadi ruang publik digital di mana masyarakat dapat secara bebas mengekspresikan kritik, kekecewaan, dan tuntutan terhadap transparansi pemerintah.[1] Data tekstual yang dihasilkan dari platform ini menyimpan kekayaan informasi mengenai suasana batin masyarakat atau *public mood*.

Dengan memanfaatkan teknik *Natural Language Processing* (NLP), informasi yang tidak terstruktur ini dapat diubah menjadi wawasan berharga untuk mengevaluasi dampak kebijakan dan efektivitas komunikasi pemerintah.[2] Analisis sentimen terhadap isu korupsi dana bencana menjadi instrumen penting karena memungkinkan pemantauan opini publik dalam skala besar secara cepat dan otomatis.[2]

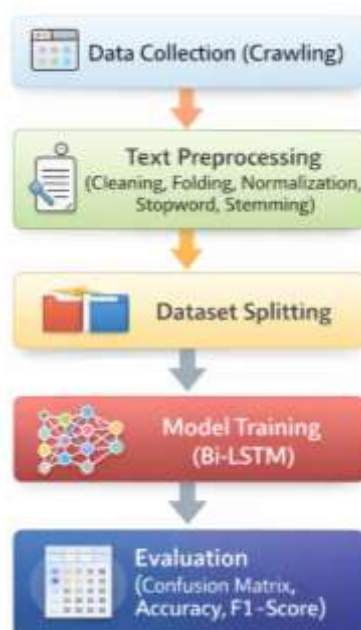
Dalam lanskap penelitian NLP saat ini, metode berbasis *Deep Learning* telah menunjukkan keunggulan signifikan dibandingkan metode *Machine Learning* konvensional seperti *Naive Bayes* atau *Support Vector Machine* (SVM) dalam menangani kompleksitas bahasa manusia.[3] Salah satu tantangan utama dalam teks media sosial Indonesia adalah sifatnya yang sangat informal, penuh dengan singkatan, slang (bahasa gaul), dan sarkasme.[4] Arsitektur konvensional sering kali gagal menangkap konteks jangka panjang atau makna tersirat dalam kalimat yang panjang. Oleh karena itu, arsitektur *Long Short-Term Memory* (LSTM) dikembangkan untuk mengatasi masalah *vanishing gradient* yang sering terjadi pada *Recurrent Neural Networks* (RNN) standar.[2]

Pengembangan lebih lanjut dari LSTM adalah *Bidirectional LSTM* (Bi-LSTM). Berbeda dengan LSTM yang memproses informasi secara searah dari awal ke akhir, Bi-LSTM memproses urutan teks dalam dua arah secara simultan: maju (*forward*) dan mundur (*backward*).[5] Kemampuan dua arah ini sangat krusial untuk memahami makna kata yang bergantung pada kata-kata sebelum dan sesudahnya.[4] Sebagai contoh, dalam konteks korupsi, kata-kata yang tampak positif secara harfiah sering kali bersifat sarkastik jika dilihat dari konteks keseluruhan kalimat. Bi-LSTM memungkinkan model untuk memiliki representasi fitur yang lebih kaya dan mendalam, menjadikannya salah satu metode paling tangguh untuk klasifikasi sentimen.[6]

Penelitian ini memfokuskan analisis pada isu korupsi dana bencana dengan mengintegrasikan arsitektur Bi-LSTM yang diperkuat oleh mekanisme *Global Max Pooling* dan *Global Average Pooling*. Penggunaan *pooling* ganda ini bertujuan untuk menangkap fitur-fitur lokal yang paling menonjol (seperti kata kunci spesifik korupsi) sekaligus memahami konteks global dari opini tersebut.[7] Melalui pendekatan ini, penelitian ini berkontribusi pada pemahaman teknis mengenai performa Bi-LSTM pada dataset bahasa Indonesia yang sangat tidak seimbang (*imbalanced*) serta memberikan gambaran empiris mengenai polaritas opini masyarakat terhadap isu integritas nasional.

2. METODOLOGI PENELITIAN

Diagram alir penelitian pada Gambar 1 menggambarkan tahapan metodologi yang digunakan dalam penelitian ini, mulai dari proses pengumpulan data hingga evaluasi kinerja model. Setiap tahapan dirancang secara sistematis untuk memastikan bahwa proses analisis sentimen dilakukan secara terstruktur, akurat, dan dapat dipertanggungjawabkan secara ilmiah.



Gambar 1. Diagram Alir Penelitian

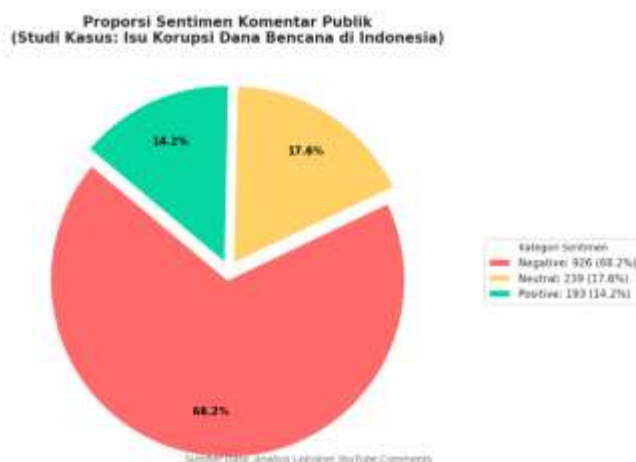
2.1 Pengumpulan dan Karakteristik Dataset

Data dikumpulkan melalui proses *crawling* pada platform media sosial menggunakan kata kunci spesifik seperti korupsi dana bencana, penyelewengan bansos, dan skandal anggaran darurat di Indonesia. Dataset terdiri dari 1.358 sampel teks yang telah melalui proses validasi label oleh pakar untuk memastikan akurasi data latih. Karakteristik utama dataset ini adalah ketidakseimbangan kelas (*imbalanced data*), di mana sentimen negatif mendominasi sebesar 68,19% (926 data), mencerminkan ketidakpuasan publik yang mendalam. Distribusi rinci data disajikan pada Tabel.

Data yang dianalisis dalam penelitian ini dikumpulkan melalui proses *crawling* pada platform media sosial dengan menggunakan kata kunci yang spesifik terkait korupsi dana bencana, penyelewengan bansos, dan skandal anggaran darurat di Indonesia.[1] Fokus pada periode pasca-pandemi dan kejadian bencana alam besar bertujuan untuk menangkap sentimen yang masih relevan dengan diskursus publik saat ini. Ketidakseimbangan kelas merupakan karakteristik utama dari dataset ini. Sentimen negatif mendominasi secara signifikan, mencerminkan ketidakpuasan mendalam terhadap perilaku koruptif di sektor kemanusiaan.[1] Distribusi data disajikan secara rinci dalam tabel berikut:

Tabel 1. Distribusi Data

Kategori Sentimen	Jumlah Data	Persentase
Negatif (Negative)	926	68,19%
Netral (Neutral)	239	17,60%
Positif (Positive)	193	14,21%
Total	1.358	100,00%



Gambar 2. Persentase Distribusi Sentimen Masyarakat terhadap Korupsi Dana Bencana

Tingginya volume data negatif menunjukkan bahwa isu korupsi dana bencana bukan hanya menjadi topik hangat, tetapi juga memicu polarisasi opini yang sangat tajam di masyarakat.[1] Kondisi *imbalanced* ini menjadi tantangan tersendiri bagi model pembelajaran mesin, karena model cenderung lebih mengenali pola pada kelas mayoritas.[5]

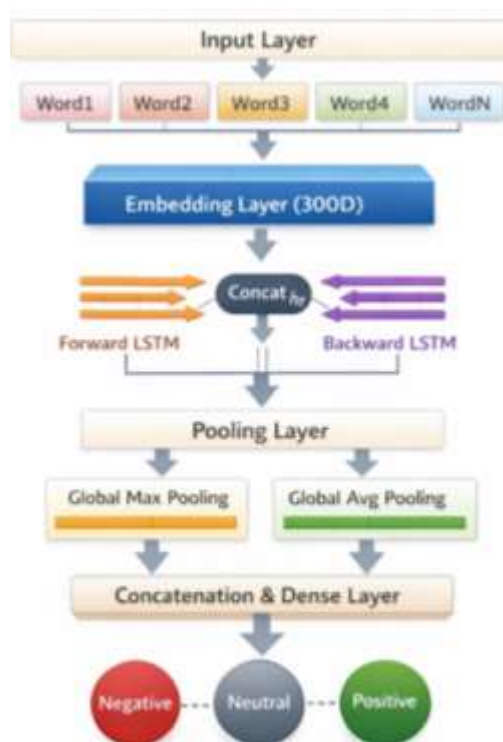
2.2 Tahapan Pra-pemrosesan Teks (*Text Preprocessing*)

Mengingat sumber data berasal dari media sosial yang memiliki tingkat kebisingan (*noise*) tinggi, tahap pra-pemrosesan menjadi sangat krusial untuk memastikan kualitas input bagi model Bi-LSTM.[8] Langkah-langkah sistematis yang diterapkan meliputi:

- Pembersihan Data (*Cleaning*): Menghapus elemen non-teks yang tidak memberikan nilai semantik bagi analisis sentimen, seperti URL, *username* (*mention @*), *hashtag* (*#*), angka, dan simbol emoji yang tidak standar.[1]
- Penyeragaman Huruf (*Case Folding*): Mengubah semua teks menjadi huruf kecil (*lowercase*) untuk menghindari ambiguitas antara kata yang sama dengan variasi kapitalisasi.[8]
- Normalisasi Bahasa (*Normalization*): Langkah ini sangat vital untuk teks bahasa Indonesia informal. Kata-kata slang, singkatan (seperti "yg" menjadi "yang", "blm" menjadi "belum"), dan kata-kata tidak baku dikonversi menjadi bentuk formal sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) untuk mengurangi redundansi fitur dan dimensi vektor.[4]
- Penghapusan Kata Berhenti (*Stopword Removal*): Menghilangkan kata-kata yang frekuensinya tinggi namun tidak membawa informasi sentimen yang signifikan, seperti kata depan dan konjungsi.[8]
- Stemming*: Proses mengubah kata berimbuhan menjadi kata dasar (misalnya "dikorupsi" menjadi "korupsi"). Hal ini membantu model dalam mengenali akar kata yang sama dalam berbagai konteks kalimat.[8]

2.3 Arsitektur Model Bi-LSTM

Model yang dikembangkan dalam penelitian ini menggunakan pendekatan *Deep Learning* berbasis sekuensial untuk menangkap fitur semantik dari teks. Proses dimulai dari transformasi teks menjadi vektor padat melalui lapisan *embedding* berdimensi 300 untuk menangkap hubungan kontekstual yang kompleks. Inti dari pemrosesan ini terletak pada lapisan *Bidirectional Long Short-Term Memory* (Bi-LSTM) yang mampu mengatasi masalah *vanishing gradient* dan menangkap konteks dua arah:



Gambar 3. Diagram Arsitektur Model Bi-LSTM

- a. Lapisan *Embedding* (*Embedding Layer*): Teks yang telah diproses dikonversi menjadi representasi vektor padat menggunakan lapisan *embedding* dengan dimensi 300. Dimensi yang cukup besar ini memungkinkan model untuk memetakan hubungan antar kata dalam ruang multidimensi yang kaya, menangkap sinonimitas dan hubungan kontekstual yang kompleks dalam kosakata terkait korupsi dan bencana.[2]
- b. Lapisan Bidirectional LSTM (Bi-LSTM Layer): Lapisan inti ini terdiri dari unit-unit memori yang dirancang untuk memproses sekuens teks secara dua arah, yaitu dari awal ke akhir (*forward*) dan dari akhir ke awal (*backward*). Secara matematis, pada setiap langkah waktu t , Bi-LSTM menghasilkan *hidden state* maju h_t^+ dan *hidden state* mundur h_t^- . Output akhir dari lapisan Bi-LSTM diperoleh dengan menggabungkan (*concatenation*) kedua *hidden state* tersebut, yang dinyatakan sebagai $h_t = [h_t^+; h_t^-]$. Pendekatan ini memungkinkan representasi setiap kata dalam sebuah kalimat untuk mempertimbangkan konteks global, baik dari kata-kata sebelumnya maupun sesudahnya dalam keseluruhan urutan teks.[9]
- c. Mekanisme *Global Pooling*: Alih-alih hanya menggunakan status terakhir dari LSTM, penelitian ini menerapkan strategi *pooling* ganda setelah lapisan Bi-LSTM:
 1. *Global Max Pooling*: Mengambil nilai maksimum dari setiap dimensi fitur di seluruh langkah waktu, yang bertujuan untuk menangkap sinyal sentimen yang paling kuat atau kata-kata kunci yang paling menentukan (misalnya kata "maling" atau "hukum mati").[7]
 2. *Global Average Pooling*: Menghitung nilai rata-rata dari fitur di seluruh urutan, memberikan representasi kontekstual yang lebih halus mengenai sentimen keseluruhan dari teks tersebut.[10]
 3. *Concatenation Layer*: Menggabungkan output dari kedua jenis *pooling* tersebut untuk membentuk satu vektor fitur yang komprehensif sebelum diteruskan ke lapisan klasifikasi.[7]
- d. Lapisan Padat dan Regularisasi (*Dense & Dropout Layer*): Vektor fitur yang telah digabungkan kemudian dilewatkan ke lapisan *Dense* (*fully connected*). Untuk mencegah model mengalami *overfitting*, terutama mengingat ukuran dataset yang terbatas dibandingkan jumlah parameter model, diterapkan teknik *Dropout* dengan probabilitas tertentu.[2] Lapisan output menggunakan fungsi aktivasi *Softmax* untuk menghasilkan probabilitas distribusi di ketiga kelas sentimen (Negatif, Netral, Positif).[11]

2.4 Pelatihan dan Evaluasi Model

Tahap pelatihan model dilakukan dengan membagi dataset menjadi data latih (*training set*) dan data uji (*testing set*) menggunakan proporsi 80:20. Berdasarkan total 1.358 data yang tersedia, diperoleh data latih sebanyak 1.086 sampel dan data uji sebanyak 272 sampel. Proporsi sampel pada data uji ($n=272$) terdiri dari 185 data kategori negatif, 48 data kategori netral, dan 39 data kategori positif. Pembagian ini dilakukan secara sistematis untuk memastikan model dapat dievaluasi secara objektif menggunakan data yang belum pernah dilihat sebelumnya selama proses pelatihan.

Kinerja model dievaluasi menggunakan metrik standar klasifikasi teks untuk memberikan gambaran performa yang komprehensif, terutama dalam menghadapi kondisi ketidakseimbangan kelas (*class imbalance*). Metrik tersebut meliputi Akurasi (*Accuracy*), Presisi (*Precision*), *Recall*, dan *F1-score*. Perhitungan metrik evaluasi didasarkan pada komponen

Confusion Matrix yang meliputi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) melalui persamaan berikut:

a. Akurasi (*Accuracy*): Mengukur proporsi total prediksi yang benar.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

b. Presisi (*Precision*): Mengukur keakuratan hasil positif, yaitu seberapa banyak kelas yang diprediksi positif benar-benar positif.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

c. *Recall* (*Sensitivity*): Mengukur kelengkapan hasil positif, yaitu seberapa banyak kelas positif aktual yang berhasil dideteksi.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d. *F1-Score*: Merupakan rata-rata harmonis antara Presisi dan *Recall*, sangat penting untuk dievaluasi dalam kasus *class imbalance* (seperti data Anda).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Keterangan:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative.

2.5 Metrik Evaluasi

Kinerja model diukur menggunakan berbagai metrik untuk memberikan gambaran yang adil, terutama dalam kondisi data tidak seimbang. Metrik yang digunakan meliputi Akurasi (*Accuracy*) untuk melihat performa keseluruhan, *Macro F1-score* untuk melihat keseimbangan performa antar kelas tanpa mempedulikan jumlah sampel, dan *Weighted F1-score* yang mempertimbangkan proporsi setiap kelas dalam perhitungan rata-rata.[12] Selain itu, *Confusion Matrix* digunakan untuk menganalisis secara mendetail jenis kesalahan klasifikasi yang dilakukan oleh model.[13]

3. HASIL DAN PEMBAHASAN

Eksperimen dilakukan untuk menguji sejauh mana arsitektur Bi-LSTM dengan *pooling* ganda mampu menangani dinamika opini masyarakat terhadap isu korupsi. Hasil yang diperoleh memberikan wawasan teknis dan sosiopolitik yang signifikan.

3.1 Analisis Frekuensi Kata (*Word Cloud*)

Sebelum melakukan klasifikasi sentimen menggunakan model Bi-LSTM, dilakukan analisis deskriptif menggunakan *Word Cloud* untuk mengidentifikasi kata kunci yang mendominasi diskursus publik mengenai korupsi dana bencana. Visualisasi ini memberikan gambaran awal mengenai sebaran fitur tekstual yang akan diproses oleh lapisan *embedding*.



Gambar 4. Gambar (a),(b)dan (c) Visualisasi *Word Cloud*

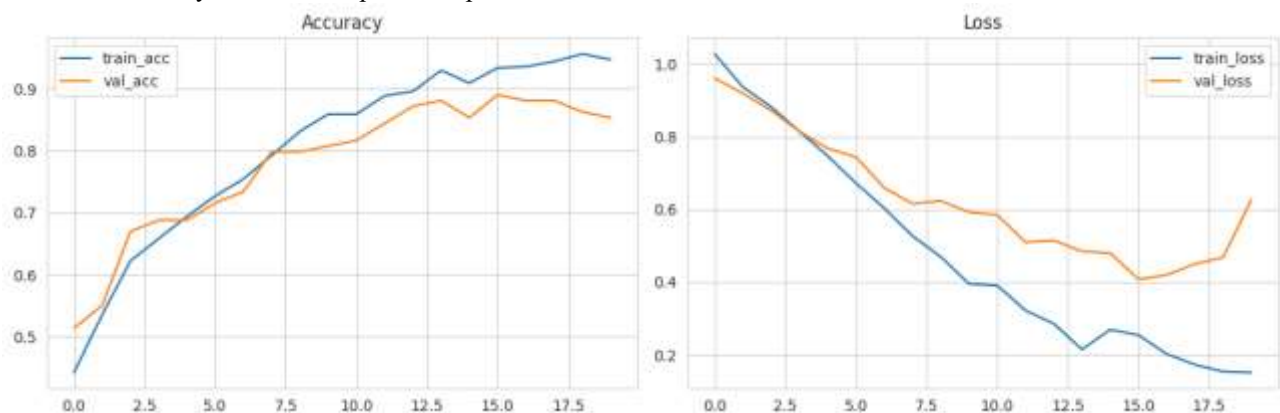
Berdasarkan analisis frekuensi kata, terdapat beberapa temuan krusial yang direpresentasikan dalam Gambar 2:

- Dominasi Sentimen Negatif: Kata kunci seperti "hukum" (358 kali), "mati" (334 kali), dan "korupsi" (231 kali) muncul dengan frekuensi sangat tinggi, yang secara langsung berkorelasi dengan dominasi kelas negatif sebesar 68,19% dalam dataset.
- Aspirasi dalam Sentimen Positif: Pada kelas positif, muncul kata "percaya" (30 kali) dan "mantap" (13 kali) yang menunjukkan dukungan terhadap langkah penindakan, namun frekuensinya jauh lebih kecil dibandingkan kata kunci pada kelas negatif.
- Ambiguitas Kata Kunci: Kata "hukum" dan "mati" muncul di kedua polaritas; pada kelas positif ini merepresentasikan harapan akan keadilan ("hukum harus adil"), sementara pada kelas negatif ini merupakan tuntutan kemarahan ("hukum mati koruptor").
- Karakteristik Sentimen Netral: Kata-kata pada kategori netral didominasi oleh istilah informatif seperti "laksana", "kpk", dan "bicara" yang bersifat objektif tanpa disertai kata sifat emosional, sesuai dengan temuan bahwa kelas netral seringkali berupa kutipan berita atau pernyataan fakta.
- Kaitan dengan Arsitektur Bi-LSTM: Tingginya frekuensi kata kunci spesifik ini memudahkan mekanisme *Global Max Pooling* untuk mendeteksi sinyal sentimen yang kuat, meskipun tumpang tindih kata (seperti kata "hukum" di dua kelas) menjadi salah satu penyebab terjadinya kesalahan klasifikasi (*error analysis*) pada model.

3.2 Analisis Proses Pelatihan Model (Accuracy dan Loss History)

Kinerja model selama fase pelatihan dipantau melalui grafik *Accuracy History* dan *Loss History* yang disajikan pada Gambar 5. Grafik tersebut menunjukkan proses konvergensi model Bi-LSTM selama 20 *epoch*.

- Accuracy History*: Terlihat peningkatan akurasi secara bertahap pada data latih dan data validasi, yang mengindikasikan bahwa model berhasil mempelajari fitur semantik dari lapisan *embedding* dan Bi-LSTM secara efektif.
- Loss History*: Penurunan nilai *loss* yang stabil tanpa adanya gap yang lebar antara kurva *training* dan *validation* menunjukkan bahwa teknik *regularization* berupa *Dropout* yang diterapkan berhasil memitigasi risiko *overfitting*. Stabilitas kurva pada *epoch* akhir menandakan bahwa model telah mencapai titik optimal dalam mengenali pola sentimen masyarakat terhadap isu korupsi.



Gambar 5. Pelatihan Model (Accuracy dan Loss History)

3.2 Analisis Kinerja Klasifikasi

Eksperimen yang dilakukan menunjukkan bahwa arsitektur Bi-LSTM dengan mekanisme *pooling* ganda mampu memberikan performa yang kompetitif dalam mengklasifikasikan sentimen teks berbahasa Indonesia yang bersifat informal. Berdasarkan pengujian pada data uji, model mencapai akurasi sebesar 0,75 dengan skor *Weighted F1* mencapai 0,76. Ringkasan performa model disajikan pada tabel berikut:

Metrik Evaluasi	Nilai Skor
Akurasi (Accuracy)	0,75
Macro F1-score	0,65
Weighted F1-score	0,76

Terdapat perbedaan yang cukup signifikan sebesar 0,11 antara nilai *Weighted F1-score* (0,76) dan *Macro F1-score* (0,65). Perbedaan ini mengindikasikan bahwa model memiliki kinerja yang sangat baik dalam mengenali kelas mayoritas, yaitu sentimen negatif, namun masih menghadapi keterbatasan dalam menggeneralisasi kelas minoritas, khususnya sentimen netral dan positif. Kondisi tersebut merupakan konsekuensi logis dari karakteristik dataset yang tidak seimbang (*imbalanced dataset*), di mana sentimen negatif mendominasi sebesar 68,19% dari total data. Akurasi sebesar

0,75 menandakan bahwa model mampu memprediksi dengan benar tiga perempat dari total data uji. Namun, perbedaan antara *Macro F1-score* (0,65) dan *Weighted F1-score* (0,76) memberikan indikasi penting mengenai pengaruh ketidakseimbangan kelas.[5] Skor *Weighted F1* yang lebih tinggi menunjukkan bahwa model sangat mahir dalam mengklasifikasikan kelas mayoritas (negatif), yang mana dalam konteks ini adalah hal yang diinginkan untuk mendeteksi kegelisahan publik. Namun, skor Macro F1 yang lebih rendah mencerminkan kesulitan model dalam mencapai performa yang setara pada kelas netral dan positif, yang memiliki jumlah data jauh lebih sedikit.[12]

3.4 Analisis Matriks Kebingungan (*Confusion Matrix*)

Confusion Matrix memberikan gambaran transparan mengenai bagaimana model Bi-LSTM berinteraksi dengan dataset korupsi dana bencana. Hasil matrix untuk set pengujian (n=272) adalah sebagai berikut:

Tabel 3. Hasil Matrix Set Pengujian

Aktual \ Prediksi	Negatif	Netral	Positif
Negatif	148	18	19
Netral	6	35	7
Positif	12	7	20

Dari data di atas, terdapat beberapa temuan krusial yang perlu didiskusikan:

- Dominasi Prediksi Benar pada Kelas Negatif:** Model berhasil mengidentifikasi 148 dari 185 sampel negatif dengan tepat. Kemampuan ini didorong oleh kuatnya fitur tekstual yang mencirikan kemarahan atau kritik dalam bahasa Indonesia, seperti penggunaan kata-kata bermuatan emosional tinggi yang berhasil ditangkap oleh mekanisme *Global Max Pooling*. [9] Hal ini selaras dengan penelitian sebelumnya yang menyatakan bahwa sentimen negatif terhadap korupsi bansos seringkali memiliki pola kata yang repetitif dan eksplisit. [1]
- Ambiguitas Kelas Netral:** Meskipun kelas netral memiliki sampel yang lebih sedikit, model mampu memprediksi 35 dari 48 sampel dengan benar. Kesalahan klasifikasi pada kelas ini (6 ke negatif dan 7 ke positif) seringkali terjadi pada teks yang bersifat informatif atau berupa kutipan berita yang tidak mengandung kata sifat emosional, sehingga model kesulitan menentukan polaritasnya. [12]
- Tantangan pada Sentimen Positif dan Sarkasme:** Kelas positif menunjukkan tingkat kesalahan yang paling tinggi secara proporsional, di mana hanya 20 dari 39 sampel yang diprediksi benar. Sebanyak 12 sampel positif justru diklasifikasikan sebagai negatif. Dalam diskursus korupsi di Indonesia, hal ini sering disebabkan oleh penggunaan gaya bahasa ironi atau sarkasme. [4] Contohnya, kalimat seperti "Luar biasa sekali pejabat kita, dana bencana pun dimakan" mengandung kata positif "luar biasa", namun secara kontekstual bermakna negatif. Meskipun Bi-LSTM dirancang untuk menangkap konteks, keterbatasan data latih pada kelas positif membuat model belum mampu menggeneralisasi pola sarkasme tersebut secara sempurna. [9]
- Tumpang Tindih Fitur (*Feature Overlap*):** Adanya 19 data negatif yang diprediksi positif menunjukkan bahwa terdapat kosa kata tertentu yang membingungkan model. Hal ini bisa terjadi jika teks negatif mengandung harapan-harapan positif (misalnya: "Semoga koruptor segera ditangkap") yang oleh model dianggap sebagai sentimen positif karena adanya kata "semoga" dan "ditangkap" (dalam konteks penegakan hukum). [14]

3.5 Keunggulan Bi-LSTM dan Pooling Ganda

Penggunaan lapisan Bi-LSTM terbukti lebih unggul daripada LSTM standar dalam konteks ini karena kemampuannya melihat kalimat secara holistik. Dalam bahasa Indonesia, posisi subjek dan predikat yang fleksibel serta penggunaan imbuhan yang kompleks menuntut model untuk memahami kata dalam konteks kalimat yang utuh. [6] Mekanisme pooling yang dikombinasikan (*Max dan Average*) memberikan keseimbangan antara ketajaman fitur dan pemahaman konteks. *Global Max Pooling* sangat efektif dalam mendeteksi keberadaan kata-kata kunci yang memiliki bobot sentimen sangat kuat, sementara *Global Average Pooling* memastikan bahwa model tidak terlalu reaktif terhadap satu kata saja dan tetap mempertimbangkan keseluruhan narasi teks. [7]

Dibandingkan dengan metode tradisional seperti SVM yang mencapai akurasi sekitar 66% pada kasus serupa [1], penggunaan Bi-LSTM dalam penelitian ini memberikan peningkatan performa yang nyata (75%). Namun, jika dibandingkan dengan model Transformer seperti IndoBERT yang bisa mencapai akurasi di atas 80% atau bahkan 90% pada beberapa tugas NLP Indonesia [12], Bi-LSTM masih memiliki celah performa. Hal ini wajar karena model Transformer memiliki parameter yang jauh lebih besar dan telah dilatih pada korpus data yang jauh lebih luas. [14] Meski demikian, arsitektur Bi-LSTM yang diusulkan menawarkan efisiensi komputasi yang lebih baik dan lebih mudah untuk diimplementasikan dalam lingkungan dengan sumber daya terbatas. [15]

3.6 Implikasi Sosiopolitik

Hasil analisis sentimen yang menunjukkan 68,19% opini negatif merupakan indikator kuat adanya ketidakpercayaan publik terhadap manajemen krisis di tingkat nasional. Masyarakat sangat sensitif terhadap isu integritas, terutama ketika menyangkut dana kemanusiaan.[1] Tingginya sentimen negatif ini seharusnya menjadi sinyal bagi lembaga pemerintah dan pengawas korupsi (seperti KPK) untuk meningkatkan transparansi penyaluran bantuan dan memperkuat komunikasi publik guna memitigasi penyebaran disinformasi atau ketidakpuasan yang berlebihan.[1]

Penelitian ini juga menyoroti pentingnya media sosial sebagai barometer *real-time* bagi pemerintah. Dengan sistem analisis sentimen otomatis berbasis *Deep Learning*, pemerintah dapat mendeteksi secara dini adanya isu-isu penyelewengan yang menjadi perhatian publik sebelum berkembang menjadi krisis legitimasi yang lebih besar.[2] Pemanfaatan teknologi ini memberikan dasar bagi pengambilan keputusan berbasis data (*data-driven decision making*) dalam mengelola respon terhadap isu-isu sosial-politik yang sensitif.[16]

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan model *Bidirectional Long Short-Term Memory* (Bi-LSTM) yang mengintegrasikan lapisan *embedding* 300 dimensi serta mekanisme *pooling* ganda (*Global Max* dan *Average Pooling*) untuk menganalisis sentimen masyarakat terhadap korupsi dana bencana, dengan capaian akurasi sebesar 0,75 dan *Weighted F1-score* 0,76. Hasil pemantauan melalui grafik *Accuracy* dan *Loss History* menunjukkan proses pelatihan yang konvergen dan stabil, di mana penggunaan teknik *Dropout* terbukti efektif memitigasi risiko *overfitting*. Temuan menunjukkan dominasi signifikan sentimen negatif sebesar 68,19%, yang didukung oleh analisis *Word Cloud* dengan kemunculan kata kunci bermuatan emosional tinggi seperti "hukum", "mati", dan "korupsi" sebagai fitur penentu utama polaritas. Meskipun model sangat tangguh dalam mengidentifikasi kritik publik, tantangan utama tetap terletak pada penanganan kelas minoritas dan ambiguitas linguistik seperti sarkasme yang menyebabkan nilai *Macro F1-score* tertahan di angka 0,65. Sebagai langkah pengembangan, integrasi teknik penyeimbangan data atau penggunaan arsitektur berbasis *Transformer* seperti IndoBERT sangat direkomendasikan untuk meningkatkan kemampuan generalisasi model terhadap kompleksitas bahasa Indonesia informal di masa depan.

REFERENCES

- [1] B. A. Maulana, R. A. Fauzi, R. I. Agustin, S. A. Azhaar, and T. Rohana, "KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN TWITTER KORUPSI BANSOS BERAS MASA PANDEMI," vol. 12, no. 2, pp. 912–918, 2024.
- [2] R. H. Yahya, W. Maharani, and R. Wijaya, "Disaster Management Sentiment Analysis Using the BiLSTM Method," vol. 7, pp. 501–508, 2023, doi: 10.30865/mib.v7i1.5573.
- [3] P. Zahwa, S. Agustian, Novriyanto, and F. Yanto, "IMPLEMENTASI BI-DIRECTIONAL LONG SHORT TERM MEMORY TERHADAP KLASIFIKASI SENTIMEN DI TWITTER PADA DATASET TERBATAS," vol. 7, no. 1, pp. 11–24, 2023.
- [4] D. Ariyus, D. Manongga, and I. Sembiring, "Enhancing Sentiment Analysis of Indonesian Tourism Video Content Commentary on TikTok: A FastText and Bi-LSTM Approach," vol. 14, no. 6, pp. 18020–18028, 2024.
- [5] A. R. Kamila, V. Budiyan, and Suriyanto, "Exploring the Effectiveness of Bi-LSTM in Detecting Indonesian-Language Hoax News," pp. 4501–4509, 2025.
- [6] G. S. Prahasto and E. B. Setiawan, "Twitter Social Media-Based Sentiment Analysis Using Bi-LSTM Method With Genetic Algorithms Optimization," vol. 11, no. 1, 2025.
- [7] K. G. R. Kamalishree, K. Rajendran, and P. S. R., "Sentiment Analysis on Movie Reviews," vol. 7, no. 7, pp. 5–10, 2022.
- [8] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep - learning strategy," *J. Big Data*, 2023, doi: 10.1186/s40537-023-00782-9.
- [9] A. Chowanda and Y. Muliono, *Indonesian Sentiment Analysis Model from Social Media by Stacking BERT and Bi-LSTM*. 2022. doi: 10.1109/AiDAS56890.2022.9918717.
- [10] Z. Hameed and B. G. Zapirain, "Sentiment classification using a single-layered BiLSTM model," vol. 8, 2020, doi: 10.1109/ACCESS.2020.2988550.
- [11] W. J. Kusoema and I. Ibrahim, "Analisis Sentimen dalam Kasus Korupsi PT. Pertamina menggunakan Metode indoBERT dan RCNN Sentiment Analysis on the PT Pertamina Corruption Case using IndoBERT and RCNN Methods," vol. 14, pp. 2246–2257, 2025, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [12] J. C. Setiawan, K. M. Lhaksana, and Bunyamin, "SENTIMENT ANALYSIS OF INDONESIAN TIKTOK REVIEW USING LSTM AND INDOBERTWEET ALGORITHM," vol. 8, no. 3, pp. 774–780, 2023.
- [13] A. R. Dwijaya and A. D. Laksito, "Sentiment Analysis of Pedulilindungi Application Reviews Using Machine Learning and Deep Learning," vol. 5, no. 2, 2023.



- [14] Kadarsi and D. Pujiyanto, “APPLICATION OF TRANSFORMER MODEL AND WORD EMBEDDING IN SENTIMENT ANALISYS OF INDONESIAN E- COMMERCE APPLICATION REVIEW,” vol. 7, no. 3, pp. 720–732, 2025.
- [15] A. R. Gunawan and R. F. A. Aziza, “Sentiment Analysis Using LSTM Algorithm Regarding Grab Application Services in Indonesia,” vol. 9, no. 2, pp. 322–332, 2025.
- [16] S. A. R. Hsb and M. Furqon, “Analisis Sentimen Terkait Kasus Korupsi Timah Pada Aplikasi X Menggunakan Algoritma Support Vector Machine Sentiment,” *J. Comput. Eng. Syst. Sci.*, vol. 10, no. 1, pp. 299–311, 2025.