



Analisis Sentimen Penggunaan Artificial Intelligence Terhadap Penyelenggaraan Pemilu 2024 Menggunakan Metode LSTM-RNN

Yaommi Juanpasha Juliansyah, Atiqah Meutia Hilda*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta
Jl. Tanah Merdeka No.20, RT.11/RW.2, Rambutan, Kec. Ciracas, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta,
Indonesia

Email: ¹jyaommi@gmail.com, ^{2,*}atiqahmeutihilda@uhamka.ac.id

Email Penulis Korespondensi: atiqahmeutihilda@uhamka.ac.id

Submitted: 14/10/2024; Accepted: 26/10/2024; Published: 27/10/2024

Abstrak—Penggunaan kecerdasan buatan (AI) dalam Pemilu 2024 di Indonesia menghadirkan peluang dan tantangan baru. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap penerapan AI dalam pemilu, dengan fokus pada dampaknya terhadap transparansi, etika, dan privasi dalam proses politik. Data dikumpulkan dari berbagai platform media sosial, dengan total 768 data yang terdiri dari opini dan komentar publik mengenai peran AI dalam pemilu. kemudian diproses untuk mendapatkan 272 data bersih, dari data bersih tersebut, ditemukan 70 data positif, 81 data netral, dan 121 data negatif. Split data dilakukan dengan 217 data latih (80%) dan 55 data uji (20%). Metode yang digunakan adalah Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) karena kemampuannya menganalisis urutan teks untuk memprediksi dan mengidentifikasi sentimen masyarakat, baik positif, negatif, maupun netral. Hasil penelitian menunjukkan bahwa AI memiliki potensi untuk meningkatkan transparansi dan efisiensi dalam pemilu, namun juga menimbulkan kekhawatiran tentang etika dan privasi. Model LSTM-RNN yang digunakan dalam penelitian ini mampu mencapai tingkat akurasi 0,71 setelah diterapkan metode up-sampling pada kelas minoritas, menunjukkan pentingnya penanganan ketidak seimbangan kelas dalam analisis sentimen. Hasil ini menekankan bahwa penerapan AI dalam pemilu perlu diimbangi dengan kebijakan dan regulasi yang memastikan transparansi, etika, dan privasi tetap terjaga. Penelitian ini juga menemukan bahwa masih banyak masyarakat Indonesia yang tidak suka atau ragu dalam penggunaan AI di pemilu. Penelitian ini memberikan wawasan penting bagi para pembuat kebijakan dan praktisi untuk memahami dampak penerapan AI dalam pemilu dan bagaimana memitigasi risiko yang mungkin timbul.

Kata Kunci: Kecerdasan Buatan; Pemilu 2024; Analisis Sentimen; LSTM-RNN; Twitter

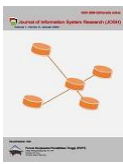
Abstract—The use of artificial intelligence (AI) in the 2024 Indonesian elections presents both new opportunities and challenges. This research aims to analyze public sentiment towards the implementation of AI in elections, focusing on its impact on transparency, ethics, and privacy in the political process. Data was collected from various social media platforms, totaling 768 data points consisting of public opinions and comments regarding the role of AI in elections. After processing, 272 clean data points were obtained, of which 70 were positive, 81 were neutral, and 121 were negative. Data splitting was performed with 217 training data (80%) and 55 testing data (20%). The method used was the Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) due to its ability to analyze text sequences to predict and identify public sentiment, whether positive, negative, or neutral. The results of the study show that AI has the potential to increase transparency and efficiency in elections, but also raises concerns about ethics and privacy. The LSTM-RNN model used in this study was able to achieve an accuracy of 0.71 after applying the up-sampling method to the minority class, demonstrating the importance of handling class imbalance in sentiment analysis. These results emphasize that the implementation of AI in elections needs to be balanced with policies and regulations that ensure transparency, ethics, and privacy are maintained. This study also found that many Indonesian people are still hesitant or opposed to the use of AI in elections. This research provides important insights for policymakers and practitioners to understand the impact of AI implementation in elections and how to mitigate potential risks.

Keywords: Artificial Intelligence; 2024 Election; Sentiment Analysis; LSTM-RNN; Twitter

1. PENDAHULUAN

Pemilihan umum (pemilu) di Indonesia merupakan salah satu pilar utama dalam sistem demokrasi, memberikan kesempatan kepada masyarakat untuk menentukan pemimpin dan wakil mereka. Proses ini tidak hanya menjadi sarana bagi individu untuk memilih, tetapi juga merupakan cerminan dari harapan dan aspirasi masyarakat terhadap masa depan politik dan sosial negara. Namun, seiring dengan perkembangan teknologi informasi dan komunikasi, terutama dalam penggunaan media sosial dan kecerdasan buatan (Artificial Intelligence/AI), muncul tantangan baru yang memengaruhi dinamika pemilu [1].

Dalam beberapa tahun terakhir, media sosial telah menjadi platform utama bagi kampanye politik [2]. Partai politik dan calon legislatif semakin memanfaatkan platform ini untuk menyampaikan pesan kampanye, membangun citra, dan berinteraksi dengan pemilih. Masyarakat kini dapat mengakses informasi dengan lebih cepat dan mudah, serta terlibat dalam diskusi politik secara real-time [3]. Namun, penggunaan media sosial juga membawa konsekuensi yang signifikan, termasuk risiko penyebaran disinformasi dan manipulasi opini publik [4]. Banyak penelitian menunjukkan bahwa informasi yang salah atau menyesatkan di media sosial dapat memengaruhi persepsi pemilih dan berpotensi merubah hasil pemilu. Dengan meningkatnya penggunaan teknologi digital dalam konteks pemilu, sangat penting untuk melakukan analisis mendalam terhadap bagaimana masyarakat merespons dan berinteraksi dengan informasi yang disampaikan di platform tersebut.



Salah satu tantangan yang muncul adalah penggunaan teknologi AI dalam proses politik. Penerapan AI dalam kampanye politik dan analisis pemilih berpotensi meningkatkan efisiensi dan efektivitas strategi kampanye. Namun, hal ini juga menimbulkan kekhawatiran terkait etika, privasi, dan penggunaan data [5]. Penggunaan AI dalam pemilu memiliki potensi untuk meningkatkan transparansi dan akurasi. Namun, hal ini juga dapat menimbulkan risiko baru dalam hal manipulasi dan pengaruh negatif terhadap pemilih. Menjawab pertanyaan-pertanyaan ini sangat penting, mengingat dampaknya terhadap integritas proses demokrasi

Selain itu, dalam penyelenggaraan pemilu, Komisi Pemilihan Umum (KPU) Indonesia telah mengimplementasikan Sistem Perhitungan Suara (SITUNG). SITUNG bertujuan untuk meningkatkan transparansi dan efisiensi dalam penghitungan suara, dengan memungkinkan publik untuk mengakses hasil perhitungan secara real-time. Sistem ini memungkinkan KPU untuk menghimpun dan mengolah data suara dari seluruh tempat pemungutan suara (TPS) di Indonesia. Meskipun SITUNG diharapkan dapat mengurangi kecurangan dan meningkatkan akuntabilitas, penggunaannya juga dapat menimbulkan tantangan baru. Misalnya, adanya potensi manipulasi data dan disinformasi terkait hasil penghitungan yang dapat beredar di media sosial, yang selanjutnya dapat memengaruhi persepsi masyarakat terhadap hasil pemilu [6].

Penelitian ini diharapkan dapat memberikan solusi untuk masalah yang dihadapi dengan menganalisis sentimen masyarakat terhadap bagaimana AI berperan dalam berbagai aspek penyelenggaraan Pemilu 2024. AI memiliki potensi untuk meningkatkan efisiensi dan keakuratan melalui analisis data pemilih, membantu memprediksi tren suara, serta menyusun strategi kampanye yang lebih terarah berdasarkan preferensi dan sentimen masyarakat. Dengan mengembangkan model analisis sentimen berbasis Long Short-Term Memory (LSTM), diharapkan para pemangku kepentingan seperti tim kampanye, partai politik, dan lembaga penyelenggara pemilu dapat memahami opini publik dengan lebih baik. Melalui analisis ini, diharapkan dapat diperoleh wawasan yang akurat dan relevan mengenai persepsi masyarakat, sehingga strategi kampanye dapat disesuaikan dan responsif terhadap kebutuhan pemilih.

Penelitian terkait analisis sentimen menggunakan metode Long Short-Term Memory (LSTM) dan penerapan kecerdasan buatan (AI) telah menunjukkan kemajuan signifikan dalam berbagai sektor, termasuk analisis teks, media sosial, dan bidang kesehatan. Misalnya, penelitian oleh N Murthy et al. [7] menunjukkan bahwa metode LSTM dapat meningkatkan akurasi dalam klasifikasi sentimen berbasis teks dengan hasil yang lebih baik saat jumlah data pelatihan diperbanyak. Penelitian Kurniasari & Setyanto [8] juga mengaplikasikan LSTM untuk analisis sentimen di Bahasa Indonesia, menunjukkan akurasi yang sangat tinggi sebesar 95%. Demikian pula, Wahyudi & Sibaroni [9] mengkaji analisis sentimen multi-aspek dengan metode LSTM pada aplikasi TikTok, dengan hasil akurasi yang bervariasi berdasarkan aspek yang dianalisis, seperti bisnis dan konten. Penelitian di bidang kesehatan oleh Eka Putra [10] memanfaatkan AI untuk sistem pengawasan pasien, dan meskipun konteksnya berbeda, AI terbukti mampu meningkatkan efisiensi dalam sektor lain. Penelitian oleh Indrayuni & Nurhadi [11] menggunakan metode Naïve Bayes dengan optimasi Particle Swarm Optimization (PSO) untuk menganalisis sentimen terkait perkembangan AI di Twitter, menunjukkan peningkatan akurasi dalam analisis sentimen umum mengenai AI.

Namun, meskipun penelitian-penelitian sebelumnya telah mengkaji berbagai metode analisis sentimen, masih terdapat gap yang belum terisi, khususnya dalam konteks penggunaan AI dalam pemilu. Kebanyakan penelitian terkait berfokus pada analisis sentimen berbasis teks atau media sosial secara umum, tanpa mengeksplorasi lebih dalam bagaimana AI memengaruhi proses politik, khususnya dalam pemilu. Penelitian oleh N Murthy et al. [7] dan Kurniasari & Setyanto [8], meskipun sukses dalam membuktikan keandalan LSTM dalam analisis sentimen, tidak menyentuh isu spesifik seperti penggunaan AI dalam pemilu atau dampaknya terhadap preferensi politik publik. Wahyudi & Sibaroni [9] memang menggunakan LSTM untuk media sosial, namun penelitian mereka terbatas pada satu platform (TikTok) dan tidak membahas AI dalam konteks pemilu. Penelitian Eka Putra [10] berfokus pada AI dalam sistem pengawasan pasien, yang meskipun relevan dalam hal penerapan AI, tidak membahas tantangan AI dalam politik, seperti transparansi dan kepercayaan publik. Sementara itu, penelitian oleh Indrayuni & Nurhadi [11] membahas sentimen terkait perkembangan AI, tetapi tidak menargetkan penerapan AI dalam pemilu dan lebih menggunakan metode tradisional seperti Naïve Bayes, yang kurang efektif dibandingkan LSTM untuk menangani data sekuensial yang lebih kompleks. Penelitian muslih [12] menggunakan dataset berisi 10.351 ulasan aplikasi Bing: Chat with AI & GPT-4 dari Google Play Store, dengan 6.040 ulasan negatif dan 4.311 ulasan positif. Klasifikasi sentimen dilakukan menggunakan metode Support Vector Machine (SVM) dan menunjukkan bahwa sentimen pengguna cenderung negatif, dengan 58,37% ulasan negatif dan 41,66% ulasan positif. Akurasi pengujian mencapai 89,44%.

Dari penelitian-penelitian terdahulu ini, dapat disimpulkan bahwa belum ada kajian yang secara khusus membahas bagaimana AI memengaruhi sentimen publik dalam penyelenggaraan pemilu, terutama dalam konteks Pemilu 2024 di Indonesia. Selain itu, metode yang digunakan dalam penelitian sebelumnya, seperti Naïve Bayes dan optimasi PSO, meskipun memberikan hasil yang cukup baik dalam analisis sentimen umum, belum sepenuhnya menangkap kompleksitas data sekuensial yang diperlukan untuk memahami dinamika sentimen politik dalam pemilu. Hal ini menimbulkan gap yang signifikan dalam literatur, di mana penerapan LSTM-RNN dalam analisis sentimen terhadap penggunaan AI dalam pemilu belum banyak dieksplorasi.

Oleh karena itu, penelitian ini bertujuan untuk mengisi gap tersebut dengan fokus pada analisis sentimen masyarakat terhadap penggunaan AI dalam penyelenggaraan Pemilu 2024 di Indonesia. Dengan menggunakan

metode LSTM-RNN yang terbukti efektif dalam menangani data sekuensial dan kompleks, penelitian ini diharapkan dapat memberikan wawasan baru tentang bagaimana AI memengaruhi preferensi politik publik. Analisis sentimen ini akan membantu penyelenggara pemilu, tim kampanye, dan partai politik untuk memahami dinamika opini masyarakat terkait penggunaan AI, serta memungkinkan mereka merespons isu-isu yang muncul dengan lebih tepat, sekaligus menjaga transparansi dan integritas proses pemilu di era digital.

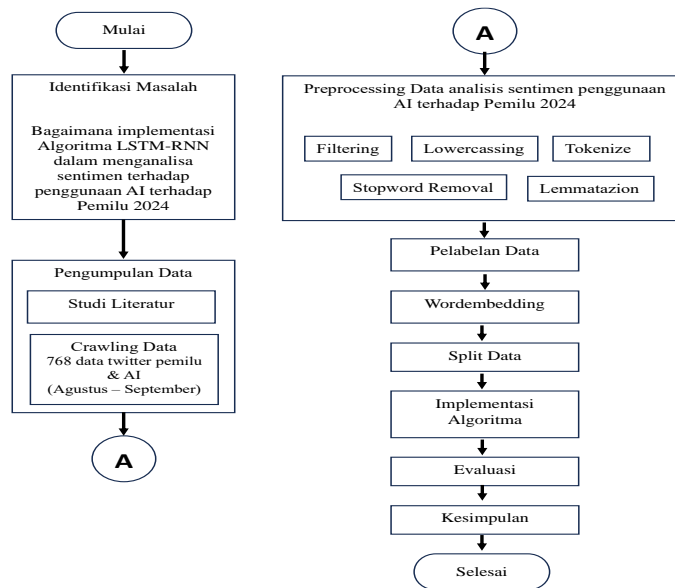
Dengan meningkatnya penggunaan teknologi dalam kampanye politik dan pemungutan suara, analisis yang mendalam diperlukan untuk mengetahui bagaimana masyarakat merespons penerapan AI di bidang politik. Melalui penerapan metode Long Short-Term Memory Recurrent Neural Network (LSTM-RNN), penelitian ini bertujuan untuk mengidentifikasi pola sentimen publik, baik positif maupun negatif, terhadap teknologi AI dalam konteks pemilu. Penelitian ini juga diharapkan dapat mengisi gap yang ada dalam literatur, di mana belum banyak kajian yang secara spesifik membahas peran AI dalam pemilu, terutama di Indonesia.

Selain itu, tujuan lain dari penelitian ini adalah untuk memberikan kontribusi empiris yang dapat digunakan oleh penyelenggara pemilu, partai politik, dan tim kampanye untuk merumuskan strategi yang lebih tepat dalam menghadapi tantangan politik di era digital. Dengan memahami sentimen publik secara lebih mendalam, penelitian ini diharapkan mampu mendukung terciptanya pemilu yang lebih transparan, adil, dan responsif terhadap kebutuhan masyarakat, sekaligus meminimalisasi potensi kesalahpahaman atau ketidakpercayaan yang mungkin timbul akibat penggunaan teknologi AI dalam proses demokrasi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Alur penelitian adalah serangkaian langkah sistematis yang diikuti oleh peneliti untuk menjawab pertanyaan penelitian atau menguji hipotesis. Proses ini penting untuk memastikan bahwa penelitian dilakukan dengan cara yang terstruktur, valid, dan dapat diandalkan. Peneliti mulai dari merumuskan masalah hingga menyajikan hasil penelitian.



Gambar 1. Alur Penelitian

Pada Gambar 1 merupakan proses alur penelitian, proses ini mencakup dari identifikasi masalah, pengumpulan data, pengolahan data, pemilihan metode, evaluasi, sampai mendapatkan kesimpulan dari penelitian.

2.2 Metode Pengumpulan Data

a. Studi Literatur

Studi literatur dilakukan dalam upaya mendapatkan sumber data yang diperlukan dalam penelitian. Data yang diperlukan dalam penelitian ini dibagi menjadi dua, yaitu data primer dan data sekunder. Data primer digunakan sebagai data utama dari penelitian ini. Data ini diperoleh secara langsung dari platform media sosial Twitter dan digunakan sebagai acuan penelitian, sementara Data Sekunder ini digunakan untuk memperkuat bahan penelitian yang dilakukan peneliti. Data ini diperoleh dari sumber-sumber sekunder seperti jurnal, buku, internet, dan penelitian sejenis.

b. Crawling Data

Crawling data merupakan sebuah proses otomatisasi dalam mengumpulkan informasi dengan cara mengambil data informasi dari sumber internet seperti website ataupun sosial media [13]. Pada penelitian ini data

diperoleh dari platform Twitter menggunakan pemrograman python dengan pustaka tweet-harvest, Data ini kemudian disimpan untuk analisis lebih lanjut.

2.3 Preprocessing Data

Preprocessing data berguna untuk mengekstrak informasi dari ‘tweets’ pengguna Twitter, mengubah kata-kata yang tidak terstruktur menjadi bentuk standar. Tujuan dari preprocessing adalah untuk membersihkan, mentransformasi, atau menyesuaikan data sehingga menjadi lebih sesuai untuk analisis atau pembelajaran model. Berikut merupakan langkah-langkah dalam melakukan preprocessing data:

- Filtering:** menghapus beberapa karakter spesial yang tidak diperlukan seperti titik (.), koma (,), tanda tanya (?), angka (0-9), dan segala sesuatu yang dianggap tidak perlu dalam proses pelatihan mendatang. Karakter-karakter tersebut dihilangkan karena dianggap tidak memiliki efek yang signifikan dalam menentukan label [14].
- Lowercasing:** Proses pengubahan semua huruf menjadi huruf kecil, hal ini diperlukan untuk menyeragamkan teks [15].
- Stopword Removal:** Stopword adalah proses menghilangkan kata-kata umum dan kata yang tidak memiliki banyak informasi khusus tentang konten atau makna yang signifikan dalam kalimat atau teks [16].
- Tokenisasi:** Tokenisasi merupakan proses memecah teks, atau string menjadi serangkaian potongan yang bermakna, yang disebut token. Seseorang dapat menganggap token sebagai bagian, mirip dengan bagaimana kata dapat menjadi token dalam suatu kalimat, dan sebuah kalimat bisa menjadi token dalam suatu paragraf [7]. Proses ini merupakan langkah umum dalam pemrosesan bahasa alami dan analisis teks.
- Lemmatization:** proses normalisasi teks yang bertujuan untuk mengubah kata-kata yang bervariasi bentuknya menjadi kata dasar yang seragam. Proses ini dilakukan untuk meningkatkan kinerja model [17]. Contoh kata yang ada pada dataset penelitian ini adalah kata “buatan” menjadi “buat”, “dimanfaatkan” menjadi “manfaat”.

2.4 Pelabelan Data

Proses pelabelan data dilakukan untuk mengidentifikasi sentimen yang terkandung di dalamnya. Data yang di dapat kemudian diberi label satu per satu pada setiap tweets akan diberi label sesuai dengan sentimen atau opini yang terkandung didalamnya, misalnya alimat seperti " Artificial intelligence bisa meningkatkan efektivitas kampanye di pemilu 2024" akan diberi label sebagai "positif", sementara kalimat "Ancaman Deepfakes dan Artificial Intelligence pada Pemilu" akan diberi label "negatif". Proses pelabelan dilakukan secara manual dan di bantu oleh 3 orang teman, dengan melakukan diskusi bersama dalam menentukan yang mana kata yang mengandung sentimen positif, sentimen negatif ataupun sentimen netral.

2.5 Word Embedding

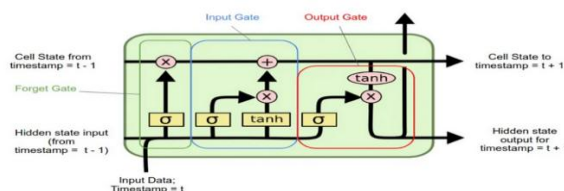
Word embedding adalah representasi vektor dari kata-kata dalam bentuk ruang vektor berdimensi rendah. Pada penelitian ini menggunakan model Word2Vec, model ini dikembangkan oleh Tomas Mikolov, Kai Chen, Greg Corrado, dan Jeffrey Dean dari Google pada tahun 2013. Word2Vec mengambil kata dari korpus sebagai input dan menghasilkan vektor kata sebagai output. Ini berfungsi secara optimal dalam menebak kata-kata secara akurat jika terdapat data yang memadai, kebergunaan, dan konteks dari setiap kata [8].

2.6 Split Data

Pada tahap ini memisahkan dataset menjadi data latih dan data uji. Data latih digunakan untuk melatih model. Model mempelajari pola dan karakteristik data selama fase ini. Data uji digunakan untuk menguji model yang telah dilatih. Dengan membagi dataset menjadi data training dan data uji, tujuannya adalah untuk mengevaluasi kinerja model. Data dibagi menjadi 80% data latih dan 20% data uji. Pembagian data ini penting untuk memastikan bahwa model tidak mengalami overfitting dan dapat diandalkan dalam menganalisis sentimen pada teks yang belum pernah ditemukan model sebelumnya [18].

2.7 Implementasi Algoritma

Setelah melakukan pembagian data, selanjutnya adalah membangun model algoritma. Model algoritma yang digunakan pada penelitian ini, yaitu algoritma Long Short-Term Memory (LSTM). LSTM adalah jenis arsitektur Recurrent Neural Network (RNN) yang dirancang khusus untuk menangani masalah pembelajaran mesin yang melibatkan urutan dan waktu. Pada gambar 2 menunjukkan sebuah arsitektur algoritma LSTM.



Gambar 2. Arsitektur LSTM



Algoritma LSTM adalah tipe khusus dari Recurrent Neural Network (RNN) yang dirancang untuk mengatasi masalah vanishing gradient, yang sering terjadi pada RNN biasa. LSTM memiliki kemampuan untuk menyimpan informasi dalam jangka panjang melalui mekanisme unik yang melibatkan tiga gerbang utama: Forget Gate, Input Gate, dan Output Gate [19]. Berikut adalah langkah-langkah kerja dari LSTM:

a. Forget Gate (Gerbang Lupa)

LSTM pertama-tama memutuskan informasi apa yang harus dilupakan dari memori sebelumnya (cell state). Gerbang ini menggunakan fungsi aktivasi sigmoid yang menghasilkan nilai antara 0 (lupakan) dan 1 (ingat) [20]. Dengan begitu, LSTM bisa menghapus informasi yang dianggap tidak penting.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

b. Input Gate (Gerbang Input)

Langkah berikutnya adalah menambahkan informasi baru ke dalam memori. Gerbang input menentukan seberapa banyak informasi baru yang harus disimpan, dan ini dihasilkan melalui kombinasi dari input saat ini dan memori sebelumnya. Nilai kandidat memori baru dihitung menggunakan fungsi aktivasi tanh, yang menghasilkan informasi baru yang akan ditambahkan ke memori.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

c. Memperbarui Memori

Setelah Forget Gate dan Input Gate bekerja, LSTM memperbarui cell state atau memori internal dengan menggabungkan informasi yang telah dilupakan dengan informasi baru yang telah dipilih. Dengan demikian, memori lama diperbarui dengan cara yang dinamis dan sesuai dengan kebutuhan input saat itu.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

d. Output Gate (Gerbang Output)

Pada tahap ini, Output Gate memutuskan bagian mana dari memori yang akan dijadikan output. Gerbang ini menggunakan fungsi sigmoid untuk menentukan seberapa banyak informasi yang akan dikeluarkan, dan output akhir kemudian dihitung berdasarkan memori yang telah diperbarui dan fungsi aktivasi tanh.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

e. Backpropagation Through Time (BPTT)

Model LSTM dilatih menggunakan metode propagasi balik yang disebut Backpropagation Through Time. Dalam metode ini, kesalahan yang dihasilkan oleh model dihitung dan digunakan untuk memperbarui bobot melalui optimisasi, seperti algoritma Adam atau SGD. Tujuannya adalah untuk meminimalkan fungsi kerugian (loss), seperti cross-entropy loss untuk tugas klasifikasi [21].

f. Prediksi

Setelah dilatih, LSTM dapat digunakan untuk melakukan prediksi dengan memproses data sekuensial baru. Model ini memanfaatkan memori yang telah "diingat" dari data sebelumnya untuk membuat keputusan yang lebih akurat di langkah-langkah selanjutnya [22].

2.8 Evaluasi

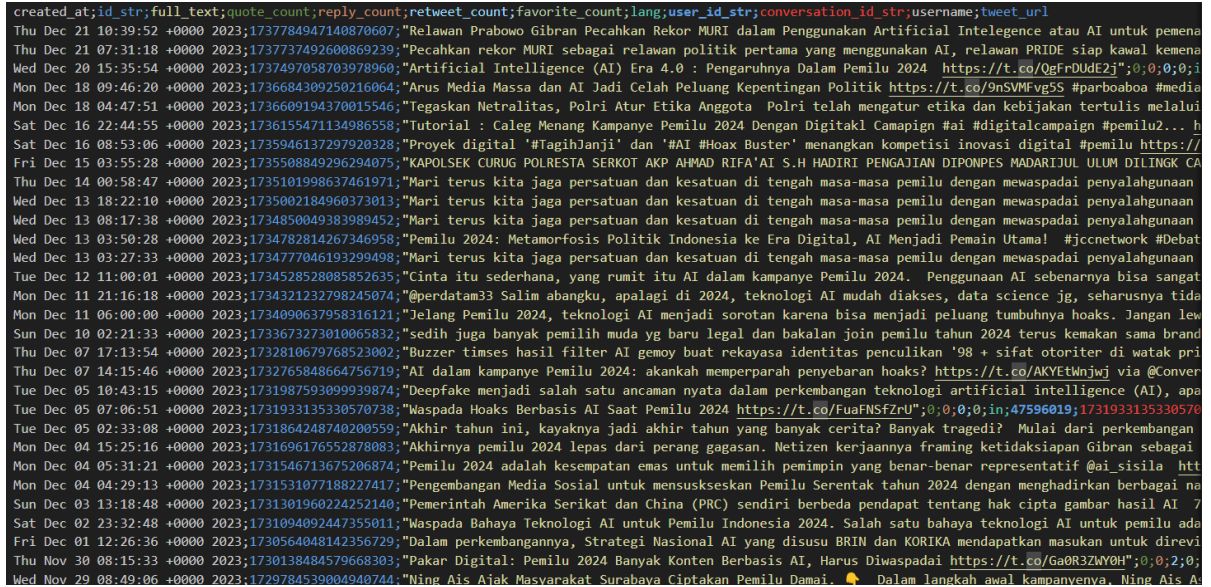
Evaluasi dalam analisis sentimen menggunakan LSTM (Long Short-Term Memory) adalah proses menilai kinerja model LSTM dalam mengklasifikasikan teks berdasarkan sentimen, seperti positif, negatif, atau netral [23]. Evaluasi ini biasanya melibatkan beberapa metrik berikut:

- Akurasi (Accuracy): Persentase dari data yang benar diklasifikasikan oleh model.
- Precision: Persentase dari data yang diklasifikasikan sebagai positif yang sebenarnya positif.
- Recall: Persentase dari semua data positif yang berhasil diklasifikasikan oleh model.
- F1-Score: Rata-rata harmonis dari precision dan recall, memberikan gambaran yang lebih komprehensif mengenai performa model.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Crawling Data

Pada penelitian ini, peneliti melakukan pengumpulan data secara otomatis dari berbagai sumber internet, seperti media sosial, situs web, atau platform lainnya (crawling). Hasil dari crawling data ini menyediakan basis informasi yang sangat penting untuk dianalisis lebih lanjut, terutama dalam menggali pola, tren, dan wawasan yang tersembunyi.



Gambar 3. Hasil Crawling Data Twitter

Pada gambar 3 merupakan hasil crawling data di media sosial twitter. Proses ini dilakukan dengan menggunakan tweet-harvest dan data di simpan dengan format csv. Data di ambil dengan menggunakan kata kunci “pemilu 2024” dan “Artificial Intelligence” yang berbahasa indonesia saja. Data yang telah dikumpulkan sebanyak 768 data mentah melalui proses crawling. setelah proses dilakukan, kita dapat mulai melakukan eksplorasi mendalam untuk memahami sentimen, perilaku, atau opini publik terkait penggunaan AI dalam pemilu 2024.

3.2 Preprocessing Data

Setelah proses crawling data berhasil mengumpulkan sejumlah data mentah, langkah berikutnya adalah proses preprocessing data. Preprocessing data adalah tahap pembersihan dan penyiapan data agar siap untuk dianalisis. Data mentah yang diambil dari berbagai sumber biasanya mengandung banyak elemen yang tidak relevan, seperti noise, duplikasi, atau format yang tidak konsisten.

Tabel 1. Tahap Pre-processing Data Twitter

Tahap Pre-processing	Hasil
Data Awal	Artificial intelligence (AI) bisa membuat konten palsu yang meniru atau mengatasnamakan tokoh politik tertentu dan ini akan banyak muncul saat Pemilu 2024 https://t.co/tc2kR8ZyVC
Filtering	Artificial intelligence AI bisa membuat konten palsu yang meniru atau mengatasnamakan tokoh politik tertentu dan ini akan banyak muncul saat Pemilu 2024
Lowercassing	artificial intelligence ai bisa membuat konten palsu yang meniru atau mengatasnamakan tokoh politik tertentu dan ini akan banyak muncul saat pemilu 2024
Stopword Removal	artificial intelligence ai membuat konten palsu meniru mengatasnamakan tokoh politik tertentu banyak muncul pemilu 2024
Tokenize	['artificial', 'intelligence', 'ai', 'membuat', 'konten', 'palsu', 'meniru', 'mengatasnamakan', 'tokoh', 'politik', 'tertentu', 'banyak', 'muncul', 'pemilu', '2024']
Lemmatization	['artificial', 'intelligence', 'ai', 'buat', 'konten', 'palsu', 'tiru', 'atas', 'tokoh', 'politik', 'tentu', 'banyak', 'muncul', 'pemilu', '2024']

Tabel 1 menunjukkan tahapan preprocessing data sentimen twitter. Melalui tahapan ini, Filtering, Lower Cassing, Stopword Removal, Tokenize, dan Lemmetazion data sentimen twitter dilakukan. Dari proses tersebut data yang telah di proses dilakukan pelabelan sentimen secara manual, supaya data lebih akurat dan dapat diolah secara efisien untuk melatih model analisis LSTM.

3.3 Pelabelan Data

Pelabelan data merupakan langkah penting untuk memberikan identitas atau kategori pada setiap data, yang akan digunakan dalam model prediksi. Data yang telah dilabeli membantu dalam proses supervised learning, di mana model dapat dilatih untuk mengenali pola berdasarkan label yang diberikan. Pada tahap ini, pelabelan data dilakukan secara manual untuk memastikan bahwa model analisis sentimen memiliki data yang berkualitas tinggi. Pelabelan data ini mencakup pengelompokan teks ke dalam kategori sentimen positif, negatif, atau netral.

Tabel 2. Label Sentimen

Teks	Label
Artificial Intelligence Bisa Meningkatkan Efektivitas Kampanye di Pemilu 2024	Positif
penggunaan AI untuk pemenangan pemilu 2024 kayaknya jadi judul penelitian yang menarik	Netral
Artificial intelligence AI bisa membuat konten palsu yang meniru atau mengatasnamakan tokoh politik tertentu dan ini akan banyak muncul saat Pemilu 2024	Negatif

Pada tabel 2 memperlihatkan hasil pelabelan data, di mana setiap teks telah dikategorikan sesuai dengan sentimen yang diidentifikasi. Pelabelan data dilakukan secara manual, data bersih diperoleh sebanyak 273 data sentimen twitter yang berkaitan dengan topik penelitian.

3.4 Word Embedding

Proses word embedding atau mengubah kata-kata menjadi representasi numerik di ruang vektor. Pada penelitian ini, metode yang digunakan untuk menghasilkan word embedding adalah Word2Vec. Pada gambar 4 merupakan hasil dari vektorisasi data sentimen twitter menggunakan metode Word2Vec.

```
array([[ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
         0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
         0,  0,  0,  0,  0,  0,  0,  0,  3,  2, 11, 100,  7,
         6, 383, 270, 271],
       [ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
         0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
         0,  0,  0,  0,  0,  0,  0,  0,  0,  0, 25,  4, 384,
         34,  7,  6, 270]])
```

Gambar 4. Hasil Word Embedding Menggunakan Word2Vec

3.5 Split Data

Selanjutnya melakukan pembagian data menjadi dua set, yaitu data latih (training set) dan data uji (test set). Pembagian ini dilakukan untuk melatih model LSTM dalam mengidentifikasi pola sentimen dari teks yang diberikan dan mengevaluasi kinerjanya pada data yang tidak pernah dilihat sebelumnya. Proporsi pembagian yang digunakan adalah 217 data latih (80%) dan 55 data uji (20%). Pembagian data ini penting untuk memastikan bahwa model tidak mengalami overfitting dan dapat diandalkan dalam menganalisis sentimen pada teks yang belum pernah ditemukan model sebelumnya. Pada gambar 5 merupakan proses pembagian data menjadi dua set, dengan mengatur data uji sebesar 0.2 atau 20%.

```
X_train, X_test, y_train, y_test = train_test_split(X_pad, y_encoded,
                                                    test_size=0.2, random_state=42)
```

Gambar 5. Proses Split Data

3.6 Implementasi Algoritma

Setelah melakukan pembagian data, selanjutnya adalah membangun model algoritma. Model algoritma yang digunakan pada penelitian ini, yaitu algoritma Long Short-Term Memory (LSTM). Pada tahap ini data akan digunakan untuk melatih model. Proses tersebut bisa diamati pada gambar 6. Model dibuat dengan menggunakan fungsi aktivasi 'softmax' dan optimasi 'adam'.

```
embed_dim = 128
lstm_out = 196

model = Sequential()
model.add(Embedding(max_features, embed_dim))
model.add(SpatialDropout1D(0.4))
model.add(LSTM(lstm_out, dropout=0.2, recurrent_dropout=0.2))
model.add(Dense(y_encoded.shape[1], activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

Gambar 6. Model Algoritma LSTM

Selama proses pelatihan, LSTM secara otomatis menyesuaikan bobot dan bias agar mampu memprediksi sentimen dengan tepat berdasarkan pola-pola dalam teks. Hasil dari model ini dievaluasi pada data uji untuk memastikan generalisasi dan performa dalam situasi nyata. Pada gambar 7 merupakan proses pelatihan model lstm dengan epoch sebanyak 15, dan batch size sebanyak 64.

```

model.fit(X_train, y_train, epochs=15, batch_size=64, verbose=1, class_weight=class_weights)
Epoch 1/15
5/5 ----- 3s 80ms/step - accuracy: 0.3439 - loss: 1.5419
Epoch 2/15
5/5 ----- 1s 107ms/step - accuracy: 0.3977 - loss: 1.4981
Epoch 3/15
5/5 ----- 0s 78ms/step - accuracy: 0.4709 - loss: 1.4696
Epoch 4/15
5/5 ----- 0s 80ms/step - accuracy: 0.5367 - loss: 1.4221
Epoch 5/15
5/5 ----- 0s 82ms/step - accuracy: 0.5426 - loss: 1.3306
Epoch 6/15
5/5 ----- 0s 77ms/step - accuracy: 0.5706 - loss: 1.1875
Epoch 7/15
5/5 ----- 0s 80ms/step - accuracy: 0.6782 - loss: 1.0024
Epoch 8/15
5/5 ----- 0s 91ms/step - accuracy: 0.7490 - loss: 0.7616
Epoch 9/15
5/5 ----- 0s 77ms/step - accuracy: 0.8559 - loss: 0.5406
Epoch 10/15
5/5 ----- 0s 79ms/step - accuracy: 0.9412 - loss: 0.3533
Epoch 11/15
5/5 ----- 0s 79ms/step - accuracy: 0.9368 - loss: 0.3386
Epoch 12/15
5/5 ----- 0s 83ms/step - accuracy: 0.9501 - loss: 0.2374
Epoch 13/15
5/5 ----- 0s 81ms/step - accuracy: 0.9581 - loss: 0.2005
Epoch 14/15
5/5 ----- 0s 89ms/step - accuracy: 0.9716 - loss: 0.1092
Epoch 15/15
5/5 ----- 0s 80ms/step - accuracy: 0.9681 - loss: 0.1546

```

Gambar 7. Proses Pelatihan Model LSTM

Setelah proses pelatihan model selesai, LSTM berhasil mencapai akurasi sebesar 0.96 pada data latih, yang menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan tingkat ketepatan yang tinggi. Hasil ini mengindikasikan bahwa model LSTM dapat menangkap pola-pola sentimen secara efektif, meskipun masih perlu dilakukan evaluasi lebih lanjut pada data uji untuk memastikan performanya tetap konsisten pada data yang belum pernah dilihat sebelumnya.

3.7 Evaluasi

Setelah model LSTM dilatih, langkah selanjutnya adalah melakukan evaluasi untuk menilai kinerjanya. Proses evaluasi ini melibatkan penggunaan data pengujian yang sebelumnya dipisahkan untuk mengukur akurasi dan metrik lainnya, seperti presisi, recall, dan F1-score.

Tabel 3. Hasil Evaluasi Data Uji LSTM

	Precision	Recall	f1-score	support
0	0.43	0.71	0.53	17
1	0.60	0.57	0.59	21
2	0.71	0.29	0.42	17
Accuracy			0.53	55
Macro avg	0.58	0.52	0.51	55
Weighted avg	0.58	0.53	0.52	55

Hasil evaluasi model LSTM yang terdapat pada **Error! Reference source not found.7** menunjukkan tingkat akurasi sebesar 0,53, karena disebabkan oleh jumlah data uji yang sedikit, sehingga tidak cukup representatif untuk mengukur kemampuan model secara akurat. Oleh karena itu, langkah selanjutnya adalah melakukan up-sample pada kelas minoritas untuk menyeimbangkan distribusi data, dengan cara menambahkan lebih banyak contoh dari kelas yang kurang terwakili. Proses ini dapat meningkatkan performa model dengan memberikan informasi yang lebih seimbang selama pelatihan, sehingga model dapat belajar dengan lebih baik dan memberikan hasil yang lebih akurat pada evaluasi. Pada tabel menunjukkan hasil dari proses up-sample pada kelas minoritas.

Tabel 4. Hasil Evaluasi Data Uji Setelah Proses Up-sample Kelas Minor

	precision	recall	f1-score	support
0	0.85	0.57	0.68	30
1	0.71	0.81	0.76	21
2	0.62	0.82	0.71	22
accuracy			0.71	73
macro avg	0.73	0.73	0.71	73
weighted avg	0.74	0.71	0.71	73

Melakukan proses up-sample pada kelas minoritas menunjukkan peningkatan akurasi menjadi 0,71. Peningkatan ini menunjukkan bahwa dengan menyediakan lebih banyak contoh data dari kelas yang kurang



terwakili. Model dapat belajar pola yang lebih baik, sehingga mampu memberikan prediksi yang lebih akurat terhadap data baru. Proses up-sampling terbukti efektif dalam memperbaiki performa model.

3.8 Pembahasan

Dalam konteks pemilu 2024, penggunaan kecerdasan buatan (AI) semakin penting untuk menganalisis sentimen publik dan memahami pandangan masyarakat terhadap calon dan isu-isu yang diangkat. Dalam penelitian ini, data dikumpulkan melalui crawling, menghasilkan 768 data kotor yang kemudian diproses untuk mendapatkan 272 data bersih. Proses ini mencakup penghapusan data yang tidak relevan dan normalisasi teks. Setelah mendapatkan data bersih, dilakukan embedding kata menggunakan Word2Vec untuk merepresentasikan kata-kata dalam bentuk vektor. Selanjutnya, data dibagi menjadi 217 data latih (80%) dan 55 data uji (20%). Model LSTM dengan menggunakan optimizer Adam serta fungsi aktivasi softmax. Model dilatih dengan epoch sebanyak 15 dan batch size 64, model LSTM menunjukkan akurasi tinggi pada data latih sebesar 0,96. Selanjutnya dilakukan pengujian model LSTM. Evaluasi pada data uji menunjukkan bahwa akurasi model menurun menjadi 0,53, yang mengindikasikan bahwa model belum mampu menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Untuk meningkatkan performa model, dilakukan evaluasi lebih lanjut dengan metode up-sampling pada kelas minoritas. Proses ini bertujuan untuk menyeimbangkan distribusi data dan memberikan lebih banyak contoh bagi model dalam mempelajari pola kelas minoritas. Hasil dari evaluasi setelah up-sampling menunjukkan peningkatan akurasi menjadi 0,71. Hal ini mencerminkan bahwa dengan pengolahan data yang lebih baik, model LSTM dapat memberikan hasil yang lebih akurat dan andal dalam analisis sentimen menjelang pemilu.

4. KESIMPULAN

Penggunaan kecerdasan buatan (AI) dalam analisis sentimen untuk pemilu 2024 menunjukkan efektivitas yang signifikan dalam menggali pandangan publik terhadap isu-isu yang berkembang. Dalam penelitian ini, data dikumpulkan melalui crawling, menghasilkan 768 data kotor yang kemudian diproses menjadi 272 data bersih. Dari data bersih tersebut, ditemukan 70 data positif, 81 data netral, dan 121 data negatif, yang menunjukkan kecenderungan sentimen publik yang lebih negatif terhadap penggunaan AI dalam pemilu 2024. Data ini kemudian dibagi menjadi 217 data latih dan 55 data uji dengan rasio 8:2, untuk memastikan model dapat dilatih dengan data yang cukup serta diuji keakuratannya pada data yang belum pernah dilihat sebelumnya. Melalui metode embedding kata menggunakan Word2Vec dan pelatihan model LSTM, dengan optimizer Adam dan fungsi aktivasi softmax, model mencapai akurasi tinggi pada data latih sebesar 0,96. Namun, akurasi yang lebih rendah pada melakukan uji model, yaitu 0,53. Evaluasi lebih lanjut dilakukan menggunakan metode up-sampling kelas minoritas berhasil meningkatkan akurasi menjadi 0,71, menandakan bahwa penanganan ketidakseimbangan kelas dapat meningkatkan performa model. Untuk meningkatkan efektivitas penggunaan AI dalam analisis sentimen di masa mendatang, disarankan agar penelitian berikutnya fokus pada beberapa langkah penting. Pertama, perluasan pengumpulan data dengan volume yang lebih besar dan keragaman yang lebih tinggi dapat membantu model dalam menangkap konteks dan nuansa yang lebih kompleks. Kedua, eksplorasi metode ensemble atau kombinasi beberapa model machine learning dapat menjadi strategi yang bermanfaat untuk meningkatkan akurasi dan keandalan prediksi. Terakhir, penting untuk melakukan pemantauan dan evaluasi secara berkala terhadap model yang telah dilatih, guna memastikan bahwa model tetap relevan dan adaptif.

REFERENCES

- [1] M. Chinen, *The International Governance of Artificial Intelligence*. Edward Elgar Publishing Limited, 2023.
- [2] B. Prakoso, atul Himmah, and F. Kurnia Illahi, "Dinamika Politik Menuju Pemilihan Presiden 2024 Di Indonesia: Studi Social Network Analysis," *Jurnal Lanskap Politik*, vol. 1, pp. 107–133, Sep. 2023, doi: 10.31942/jlp.2023.1.2.
- [3] A. Komarudin, A. Meutia Hilda, and C. Author, "Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Pada Play Store Menggunakan Metode Naïve Bayes," 2024. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/co-science>
- [4] A. Daisy, Z. Fadhilah, and S. Retnoningsih, "PERANCANGAN KAMPANYE DIGITAL MELAWAN DISINFORMASI MELALUI ARTIFICIAL INTELLIGENCE DAN DEEPAKE DI KALANGAN PRA LANSIA USIA 45-55 TAHUN," 2024.
- [5] I. Muhammad and M. Mirza, "Implementasi Artificial Intelligence Dalam Iklan Politik Menuju Masyarakat Indonesia 5.0," *JURNAL VISUAL IDEAS*, vol. 3, no. 2, 2023.
- [6] T. Akbar, R. Imanda, I. A. Abstrak, and K. Kunci, "Perbandingan Analisis Sentimen pada Aplikasi SIREKAP dengan aplikasi SITUNG di Media Sosial X Menggunakan Algoritma Support Vector Machine," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, Aug. 2024, doi: 10.33022/ijcs.v13i4.4084.
- [7] G. S. N Murthy, S. Rao Allu, B. Andhavarapu, M. Bagadi, and M. Belusonti, "Text based Sentiment Analysis using LSTM; Text based Sentiment Analysis using LSTM," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 05, May 2020. [Online]. Available: www.ijert.org
- [8] L. Kurniasari and A. Setyanto, "SENTIMENT ANALYSIS USING RECURRENT NEURAL NETWORK-LSTM IN BAHASA INDONESIA," 2020.
- [9] D. Wahyudi and Y. Sibaroni, "Deep Learning for Multi-Aspect Sentiment Analysis of TikTok App using the RNN-LSTM Method," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1665.



- [10] O. Eka Putra, “IMPLEMENTASI ARTIFICIAL INTELLIGENCE PADA SISTEM PENGAWASAN PASIEN RUMAH SAKIT,” *J Teknol*, vol. 10, no. 02, 2020.
- [11] E. Indrayuni and A. Nurhadi, “OPTIMASI NAIVE BAYES BERBASIS PSO UNTUK ANALISA SENTIMEN PERKEMBANGAN ARTIFICIAL INTELLIGENCE DI TWITTER,” *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 65–70, Aug. 2023, doi: 10.33480/inti.v18i1.4282.
- [12] Muslih, A. Meutia Hilda, and M. Jafar Elly, “Metode Klasifikasi Support Vector Machine (SVM) Untuk Analisis Sentimen Aplikasi Bing_ Chat with AI & GPT-4 Di Google Play Store,” *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, vol. 17, Mar. 2024, doi: 10.33322/petir.v17i1.2283.
- [13] M. A. Khder, “Web scraping or web crawling: State of art, techniques, approaches and application,” *International Journal of Advances in Soft Computing and its Applications*, vol. 13, no. 3, pp. 144–168, 2021, doi: 10.15849/ijasca.211128.11.
- [14] M. Z. Rahman, Y. A. Sari, and N. Yudistira, “Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM),” 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [15] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltip.2022.04.020.
- [16] S. Sarica and J. Luo, “Stopwords in technical language processing,” *PLoS One*, vol. 16, no. 8 August, Aug. 2021, doi: 10.1371/journal.pone.0254937.
- [17] J. Fehle, T. Schmidt, and C. Wolff, “Lexicon-based Sentiment Analysis in German: Systematic Evaluation of Resources and Preprocessing Techniques,” 2021. [Online]. Available: <https://www.springer.com/de>
- [18] I. O. Muraina, “IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND DATA ANALYSTS,” 2022. [Online]. Available: <https://www.researchgate.net/publication/358284895>
- [19] N. Yudistira, *Deep Learning: Teori, Contoh Perhitungan, dan Implementasi*. Deepublish, 2024.
- [20] T. Szandała, “Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks,” 2020.
- [21] J. Cahyani, S. Mujahidin, and T. P. Fiqar, “Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional,” *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 2, p. 346, Jul. 2023, doi: 10.26418/justin.v11i2.57395.
- [22] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, “A survey on long short-term memory networks for time series prediction,” in *Procedia CIRP*, Elsevier B.V., 2021, pp. 650–655. doi: 10.1016/j.procir.2021.03.088.
- [23] M. H. Al-Areef and K. Saputra, “Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM,” *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 22, pp. 270–279, Aug. 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jis/index>