

Implementasi NLP Klasifikasi Berita Pemilu Menggunakan Algoritma LSTM

¹Harry Vadilan Sianturi, ²Cahyono Budy Santoso

^{1,2}Universitas Pembangunan Jaya, Indonesia

¹Harry.Vadilansianturi@student.upj.ac.id; ²cahyono.budy@upj.ac.id;

Article Info

Article history:

Received, 2025-05-14

Revised, 2025-05-19

Accepted, 2025-06-03

Kata Kunci:

klasifikasi teks,
berita politik
pasangan calon
LSTM
pemilu Indonesia

Keywords:

text classification
political news
candidate pairs
LSTM
Indonesian election

ABSTRAK

Penelitian ini mengkaji penerapan metode klasifikasi teks berbasis Long Short-Term Memory (LSTM) untuk mengelompokkan berita pemilu berdasarkan entitas pasangan calon presiden dan wakil presiden. Permasalahan yang diangkat adalah tidak adanya sistem klasifikasi otomatis yang mampu mengidentifikasi afiliasi politik secara langsung dalam teks berita yang jumlahnya sangat besar di media digital. Dalam studi ini, data berita diperoleh dari sumber terbuka dan diberi label secara otomatis berdasarkan kemunculan kata kunci nama kandidat. Model dilatih dengan pendekatan pembelajaran terawasi menggunakan arsitektur LSTM untuk mengenali pola sekuensial dalam teks. Hasil evaluasi menunjukkan bahwa model berhasil mencapai akurasi validasi sebesar 95,44% dan menghasilkan nilai f1-score makro sebesar 0,95, menunjukkan kinerja klasifikasi yang sangat baik di seluruh kategori paslon. Selain itu, hasil prediksi pada data uji menunjukkan konsistensi dan stabilitas model dalam mengidentifikasi entitas politik. Penelitian ini menunjukkan bahwa pendekatan LSTM efektif dalam mengklasifikasikan berita pemilu berbasis entitas, dan berpotensi untuk diintegrasikan dalam sistem analitik media atau pemantauan wacana politik secara otomatis.

ABSTRACT

This study examined the application of a Long Short-Term Memory (LSTM)-based text classification method to categorize election news according to presidential and vice-presidential candidate entities. The core problem addressed was the lack of an automated classification system capable of identifying political affiliations directly within the vast volume of digital news content. In this research, news data were collected from open-access sources and automatically labeled based on the occurrence of candidate-related keywords. A supervised learning approach was implemented using the LSTM architecture to capture sequential patterns within the news text. The evaluation results demonstrated that the model achieved a validation accuracy of 95.44% and a macro-averaged F1-score of 0.95, indicating strong classification performance across all candidate categories. Furthermore, predictions on test data revealed the model's consistency and stability in recognizing political entities. This study confirmed the effectiveness of the LSTM-based approach for entity-based election news classification and highlighted its potential for integration into automated media analytics and political discourse monitoring systems.

This is an open access article under the [CC BY-SA](#) license.



Penulis Korespondensi:

Cahyono Budy Santoso,
Program Studi Sistem Informasi,
Universitas Pembangunan Jaya,
Email: cahyono.budy@upj.ac.id

1. PENDAHULUAN

Latar belakang masalah yang dihadapi dalam konteks pemberitaan pemilu semakin kompleks seiring dengan meningkatnya volume informasi yang tersedia di media sosial dan platform berita lainnya. Dalam konteks ini,

berita yang berhubungan dengan politik, khususnya tentang calon pemimpin (paslon), sering kali diwarnai oleh misinformasi dan disinformasi. Hal ini menjadikan penerapan teknik pemrosesan bahasa alami (*Natural Language Processing (NLP)*) untuk klasifikasi berita pemilu menjadi suatu kebutuhan mendesak. Penelitian menunjukkan bahwa teknik berbasis pembelajaran mendalam, seperti *Long Short-Term Memory (LSTM)*, dapat menghasilkan klasifikasi yang akurat untuk tugas dengan menggunakan model BiLSTM dalam klasifikasi berita palsu [1].

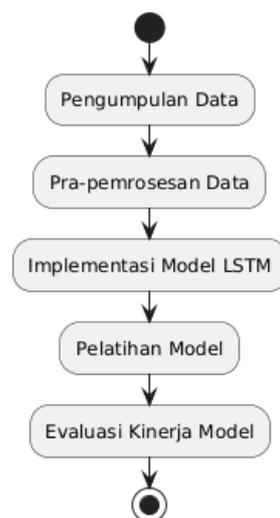
Tinjauan literatur menunjukkan bahwa banyak penelitian telah berfokus pada pengembangan algoritma untuk deteksi berita palsu dalam konteks media sosial, di mana perubahan cepat dan penyebaran informasi yang luas membuat tugas ini semakin sulit dilakukan secara manual [2][3]. Berbagai pendekatan dikembangkan, mulai dari ekstraksi fitur dari teks tidak terstruktur hingga penggunaan arsitektur jaringan saraf yang lebih canggih, termasuk CNN dan LSTM [4][5]. Penelitian seperti yang dipaparkan oleh Jain Jain [6] juga menyoroti pentingnya teknik pembelajaran yang ditingkatkan dalam mengklasifikasikan berita pemilu, menunjukkan tren penggunaan model berbasis bahasa besar seperti BERT dan aplikasinya dalam klasifikasi teks.

Urgensi penerapan klasifikasi otomatis terhadap berita pemilu juga dilandasi oleh kebutuhan praktis untuk mendukung literasi politik masyarakat serta mitigasi terhadap polarisasi informasi. Studi sebelumnya telah menunjukkan bahwa model *deep learning* seperti LSTM tidak hanya efektif untuk tugas klasifikasi umum, tetapi juga dapat diadaptasi untuk konteks lokal yang lebih kompleks, termasuk pemberitaan politik berbasis entitas [13]. Penelitian lain membuktikan bahwa penggunaan representasi vektor kata seperti Word2Vec atau FastText mampu meningkatkan performa model dalam memetakan konteks semantik pada berita politik [14]. Lebih lanjut, pendekatan multi-label berbasis entitas politik yang memanfaatkan arsitektur LSTM dan mekanisme perhatian (*attention mechanism*) terbukti memperkuat akurasi klasifikasi dalam domain kebijakan publik [15]. Dalam konteks Indonesia, masih jarang ditemukan penelitian yang secara eksplisit menggabungkan teknik NLP modern dengan kerangka entitas politik berbasis pasangan calon, padahal kombinasi ini sangat potensial untuk menghasilkan sistem klasifikasi yang adaptif dan relevan secara kontekstual [16]. Oleh karena itu, penelitian ini tidak hanya mengisi celah metodologis, tetapi juga menghadirkan solusi terapan yang strategis untuk pengelolaan informasi politik digital secara berkelanjutan [17].

Alasan mengapa penelitian ini dilakukan adalah karena masih terdapat celah dalam pendekatan yang ada, terutama dalam hal penerapan teknologi NLP untuk pengklasifikasian informasi politik yang berbasis konteks lokal. Meskipun beberapa studi sebelumnya telah menunjukkan hasil yang baik, belum banyak penelitian yang mengadopsi pendekatan komprehensif yang menggabungkan algoritma tradisional dengan model pembelajaran mendalam yang lebih baru dalam konteks pemilu. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi performa LSTM dalam klasifikasi berita pemilu berdasarkan entitas politik pada dataset berita lokal.

Pertanyaan penelitian yang ingin dijawab dalam studi ini adalah: "Bagaimana efektivitas algoritma LSTM dalam mengklasifikasikan berita pemilu berdasarkan entitas politik?" Dari sini, diharapkan dapat ditemukan metode yang lebih efisien dan akurat untuk menangani berita pemilu, berkontribusi pada pengurangan penyebaran informasi keliru, dan memberikan informasi yang akurat kepada masyarakat

2. METODE PENELITIAN



Gambar 1. Sumber data berita Pemilu

Sebagaimana pada gambar 1, penelitian ini akan menggunakan pendekatan eksperimental dengan langkah-langkah sebagai berikut:

- Pengumpulan Data: Data berupa berita pemilu akan dikumpulkan dari berbagai sumber, termasuk media cetak dan *online*. Sebelumnya, penelitian oleh Erlich et al. menunjukkan pentingnya pengukuran perbedaan dalam liputan berita kandidat dan partai dalam konteks demokrasi yang berkembang [9]. Untuk sumber data penelitian ini menggunakan *Kaggle*.
- Pra-pemrosesan Data: Data berita akan dibersihkan dan dipra-pemroses menggunakan teknik NLP dasar seperti tokenisasi, penghapusan kata umum (*stopword removal*), dan perataan kata (*lemmatization*) untuk mempersiapkan data sebelum memasukkan ke model [10]. Dalam konteks klasifikasi berita, Adipradana et al. melakukan penelitian yang mengutamakan pentingnya pra-pemrosesan untuk meningkatkan akurasi model klasifikasi [11].
- Implementasi Model: Model LSTM akan dibangun menggunakan arsitektur LSTM yang telah dikembangkan, mengingat kekuatan LSTM dalam memproses urutan data panjang dan mengingat informasi penting untuk klasifikasi [8]. Penelitian oleh Liu menunjukkan bahwa LSTM dapat mengembangkan fitur klasifikasi yang penting melalui pemrosesan yang efisien dalam konteks berita [8].
- Pelatihan Model: Model akan dilatih menggunakan data yang telah dipra-pemroses. Pengujian dilakukan dengan pembagian data menjadi set pelatihan dan pengujian, dengan teknik validasi silang untuk memastikan keakuratan model. Model LSTM yang diusulkan oleh Liu menunjukkan keefektifan dalam ekstraksi fitur yang akan diintegrasikan untuk membangun pengklasifikasi [8].
- Evaluasi Kinerja Model: Evaluasi terhadap model dilakukan dengan ukuran standar seperti akurasi, presisi, *recall*, *F1-score*, dan *confusion matrix*. Sebelumnya, penggunaan metode item untuk evaluasi kinerja klasifikasi juga diungkapkan dalam penelitian Yefferson et al. yang menggunakan LSTM untuk mendeteksi hoax [7].

3. HASIL DAN ANALISIS

Data yang digunakan dalam penelitian ini diperoleh dari *platform* publik *Kaggle*, yang menyediakan kumpulan berita digital mengenai Pemilihan Umum Presiden Indonesia tahun 2024. Dataset ini berisi ribuan artikel berita yang secara umum membahas aktivitas, pernyataan, maupun representasi masing-masing pasangan calon presiden dan wakil presiden. Untuk mendukung proses klasifikasi topik berdasarkan entitas politik, setiap berita dalam dataset dianalisis berdasarkan kemunculan kata kunci nama tokoh, yang dikelompokkan ke dalam tiga kategori utama: Paslon 1, Paslon 2 dan Paslon 3. Untuk datanya dalam bentuk file csv. Untuk capture sebagian datanya ada di gambar 2 berikut.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
title,author,publish_date,article_text,url,main_image,tag																				
Hasto Ingin Debat Pilpres Pertarungan Gagasan dan Tak Ada Rekayasa,detikNews,2023-12-10 23:31:00+07:00,"Sekjen PDIP sekaligus Sekretaris Tim Pemenang Nasional (TPN) Ganjar Pranowo-Mahfud MD, Hasto Kristi																				
Sehingga debat itu betul-betul menjadi pertarungan gagasan yang otentik, pertarungan gagasan yang tidak ada suatu rekayasa," kata Hasto kepada wartawan di Serang, Banten, Minggu (10/12/2023).																				
Menurut Hasro, para capres-cawapres juga harus bisa menunjukkan jati dirinya sebagai pemimpin yang otentik. Serta bisa menampilkan visi misi dan programnya sebagai pemimpin.																				
ADVERTISEMENT SCROLL TO CONTINUE WITH CONTENT																				
Bukan dari mendompleng siapa," ucapnya.																				
Makanya, ia mengatakan, Ganjar-Mahfud sudah siap dengan rangkaian debat yang diselenggarakan oleh KPU. Karena, debat bagi TPN adalah pertarungan gagasan dan disaksikan oleh rakyat Indonesia.																				
Ganjar-Mahfud siap untuk melakukan debat," tegasnya.																				
Di samping itu, TPN juga setuju jika para panelis debat membuat pakta integritas agar tidak membocorkan pertanyaan saat debat. Karena TPN melihat bahwa seluruh tahapan Pemilu saat ini rentan ada intimidasi.																				
Sehingga intimidasi itu menunjukkan bahwa ada yang khawatir terhadap gerakan rakyat, terhadap pemimpin yang baik ini. Malah sejak awal ada yang melakukan manipulasi hukum. Maka, di dalam KPU pun terkal																				
Terima Kunjungan Gibran di Ponpes Al-Tsaqafah, Said Aqil Tegaskan Netral,detikNews,2023-12-10 23:09:00+07:00,"Mantan Ketua Umum Pengurus Besar Nahdlatul Ulama (PBNU) Said Aqil Siradj mengungkapkan tida																				

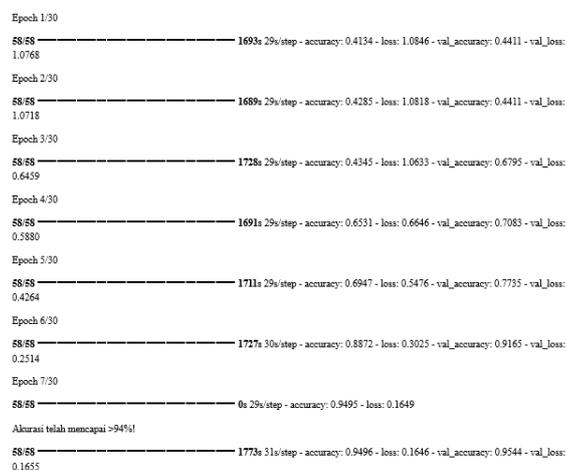
Gambar 2. Sumber data berita Pemilu

Tahap *preprocessing* dilakukan untuk menyiapkan teks berita sebelum dimasukkan ke dalam model klasifikasi berbasis *LSTM*. Setiap entri teks dalam kolom gabungan judul dan isi berita (*merge_title_article*) terlebih dahulu dibersihkan melalui fungsi *preprocess_text*, yang menjalankan serangkaian transformasi seperti konversi huruf ke huruf kecil, penghapusan angka, tanda baca, karakter *HTML*, simbol, dan karakter tunggal. Selain itu, fungsi tersebut juga menghapus *hashtag*, *mention*, serta merapikan spasi berlebih. Proses ini bertujuan untuk menghilangkan elemen non-informatif yang dapat mengganggu pembelajaran model. Setelah pembersihan, berita diberi label berdasarkan kemunculan kata kunci nama pasangan calon presiden dan wakil presiden yang telah didefinisikan dalam *label_map*. Label diberikan melalui fungsi *labeling_text*, yang mengidentifikasi keberadaan kata kunci secara eksplisit dan memilih entitas politik dengan frekuensi tertinggi.

Data yang tidak memenuhi ambang batas minimum kemunculan kata kunci dieliminasi untuk menjaga kualitas data latih. Hasil dari proses ini menghasilkan dataset terstruktur dengan dua komponen utama: teks bersih (*news*) dan label paslon dalam format *one-hot encoding*. Proses selanjutnya melibatkan tokenisasi menggunakan Tokenizer dari *TensorFlow* dengan batasan 20.000 kata dan *token OOV (out-of-vocabulary)*. *Token* yang dihasilkan kemudian dikonversi menjadi urutan bilangan bulat dan dipadatkan menggunakan *pad_sequences* dengan panjang maksimal 8.958 token per dokumen. Pendekatan preprocessing ini memastikan homogenitas struktur input dan memungkinkan model LSTM mengenali pola *sekuensial* dalam data teks politik.

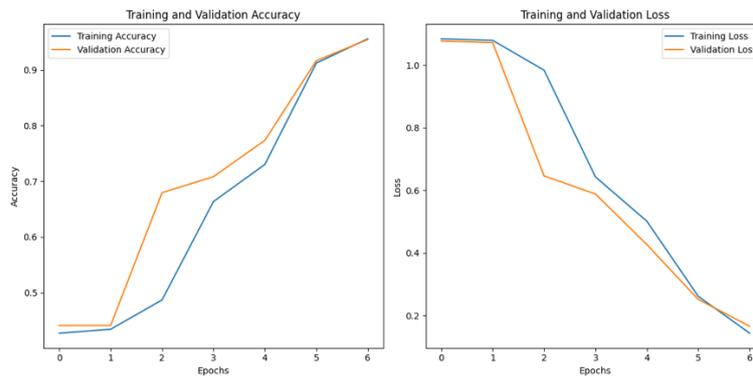
Model klasifikasi teks yang digunakan dalam penelitian ini dibangun dengan pendekatan jaringan saraf dalam (*deep learning*) menggunakan arsitektur *Long Short-Term Memory (LSTM)*. Model dikembangkan menggunakan *API Sequential* dari *TensorFlow Keras*, yang memungkinkan pembangunan jaringan secara linier dan bertingkat. Lapisan pertama yang digunakan adalah *Embedding*, yang berfungsi mengonversi indeks kata hasil tokenisasi menjadi vektor representasi berdimensi tetap, sehingga dapat menangkap konteks semantik dalam ruang vektor. Lapisan ini diikuti oleh lapisan *LSTM*, yang dirancang untuk mempelajari hubungan sekuensial dalam data teks serta mempertahankan informasi dari urutan sebelumnya melalui mekanisme memori jangka panjang. Selanjutnya, digunakan *GlobalAveragePooling1D* untuk mereduksi dimensi keluaran dari LSTM menjadi vektor tunggal, diikuti oleh lapisan *Dense* dengan unit keluaran sebanyak tiga, masing-masing mewakili satu paslon. Fungsi aktivasi *softmax* digunakan pada output untuk menghasilkan probabilitas klasifikasi multi-kelas. Model dilatih menggunakan fungsi *categorical_crossentropy* sebagai *loss function* dan *optimizer Adam*, dengan metrik utama berupa akurasi. Proses pelatihan dilakukan selama 30 epoch dengan batch size sebesar 100, dan data dibagi menjadi data latih dan validasi menggunakan skema 80:20. Implementasi *ModelCheckpoint* digunakan untuk menyimpan bobot model terbaik berdasarkan kinerja validasi, serta callback monitoring untuk evaluasi dinamika pelatihan. Hasil pelatihan menunjukkan bahwa model mampu mempelajari pola representasi berita secara efektif, serta memiliki performa klasifikasi yang kompetitif dalam mengidentifikasi entitas politik yang disebutkan dalam teks berita.

Hasil pelatihan model LSTM seperti pada gambar 3 menunjukkan adanya peningkatan performa klasifikasi yang signifikan seiring bertambahnya jumlah *epoch*. Pada awal pelatihan (*Epoch 1 dan 2*), model menunjukkan akurasi yang relatif rendah baik pada data latih (0,4134 dan 0,4285) maupun data validasi (0,4411 untuk keduanya), dengan nilai *loss* yang masih tinggi (di atas 1,07). Hal ini mencerminkan bahwa model masih berada pada tahap awal proses pembelajaran representasi sekuensial teks. Namun, peningkatan drastis mulai terlihat pada *Epoch 3*, di mana akurasi validasi meningkat menjadi 0,6795 dan *loss* validasi turun signifikan menjadi 0,6459, menunjukkan bahwa model mulai mengenali pola yang lebih informatif dalam data. Tren positif ini berlanjut pada epoch-epoch berikutnya, dengan akurasi pelatihan mencapai 0,6531 di *Epoch 4* dan terus meningkat menjadi 0,6947 (*Epoch 5*), 0,8872 (*Epoch 6*), hingga mencapai 0,9496 pada *Epoch 7*. Pada titik ini, model berhasil mencapai akurasi validasi sebesar 95,44%, yang menunjukkan bahwa model memiliki kemampuan generalisasi yang sangat baik terhadap data yang tidak terlihat. Nilai *loss* yang semakin rendah, baik pada data pelatihan (turun dari 1,0846 menjadi 0,1646) maupun validasi (dari 1,0768 menjadi 0,1655), memperkuat indikasi bahwa model mampu melakukan klasifikasi berita berdasarkan entitas politik dengan presisi tinggi tanpa mengalami *overfitting* yang signifikan dalam 7 epoch awal pelatihan.



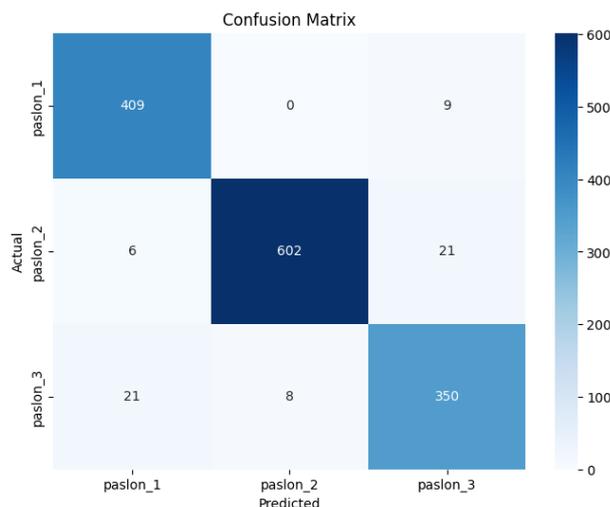
Gambar 3. Pelatihan model LSTM

Visualisasi hasil pelatihan model pada gambar 4 menunjukkan tren positif yang konsisten antara akurasi dan loss pada data pelatihan dan validasi selama tujuh epoch awal. Grafik sebelah kiri menampilkan peningkatan akurasi yang progresif, di mana akurasi validasi mulai melampaui akurasi pelatihan sejak epoch kedua, dan keduanya terus meningkat secara paralel hingga mencapai nilai di atas 0,95 pada *epoch* keenam. Hal ini menunjukkan bahwa model tidak hanya belajar secara efektif dari data pelatihan, tetapi juga mampu melakukan generalisasi dengan baik pada data yang tidak dilatih sebelumnya. Di sisi lain, grafik sebelah kanan menggambarkan penurunan nilai loss secara signifikan baik pada data pelatihan maupun validasi, dari lebih dari 1,0 pada awal pelatihan menjadi kurang dari 0,2 pada epoch keenam. Kesesuaian pola antara kurva akurasi dan loss pada kedua grafik tersebut mengindikasikan bahwa model tidak mengalami overfitting dan bahwa proses pelatihan berlangsung secara stabil. Temuan ini memperkuat bukti bahwa arsitektur LSTM yang digunakan dalam penelitian ini efektif dalam menangkap pola sekuensial dari teks berita politik untuk tujuan klasifikasi berdasarkan entitas pasangan calon.



Gambar 4. Training dan Validasi Akurasi Loss

Evaluasi performa model klasifikasi pada gambar 5 menggunakan *confusion matrix* menunjukkan bahwa model LSTM yang dikembangkan mampu melakukan klasifikasi dengan tingkat akurasi yang tinggi terhadap ketiga kelas entitas politik, yaitu *paslon_1*, *paslon_2*, dan *paslon_3*. Berdasarkan matriks tersebut, model berhasil mengklasifikasikan sebanyak 409 dari 418 data *paslon_1* secara benar, 602 dari 629 data *paslon_2*, serta 350 dari 379 data *paslon_3*. Jumlah kesalahan klasifikasi tergolong rendah, terutama pada kelas *paslon_1* yang hanya mengalami 9 kasus salah klasifikasi ke *paslon_3* dan tidak satupun ke *paslon_2*. Kinerja terbaik ditunjukkan pada kelas *paslon_2* dengan akurasi klasifikasi mendekati sempurna dan hanya 6 kasus salah diklasifikasikan sebagai *paslon_1* serta 21 ke *paslon_3*. Sementara itu, *paslon_3* memiliki tingkat kekeliruan klasifikasi yang sedikit lebih tinggi dibanding dua kelas lainnya, dengan 21 kasus salah ke *paslon_1* dan 8 ke *paslon_2*. Secara keseluruhan, distribusi prediksi dan aktual dalam *confusion matrix* mencerminkan kestabilan serta efektivitas model dalam mengenali pola kata yang spesifik terhadap masing-masing entitas politik, memperkuat temuan bahwa pendekatan LSTM sangat sesuai untuk tugas klasifikasi teks berbasis entitas dalam domain berita politik.



Gambar 5. Confusion Matrix

Berdasarkan tabel 1, hasil evaluasi performa model yang ditunjukkan melalui *classification report*, model klasifikasi berbasis *Long Short-Term Memory (LSTM)* yang dikembangkan menunjukkan kinerja yang sangat baik dalam mengelompokkan berita pemilu berdasarkan entitas pasangan calon. Nilai *precision* berkisar antara 0,92 hingga 0,99, menunjukkan bahwa model mampu meminimalkan prediksi salah positif, terutama pada kelas *paslon_2* yang mencapai *precision* tertinggi sebesar 0,99. *Recall* juga berada pada tingkat tinggi, dengan nilai tertinggi sebesar 0,98 pada *paslon_1*, yang mencerminkan kemampuan model dalam menangkap hampir seluruh *instance* dari kelas tersebut. Nilai *f1-score* yang merata dan tinggi di semua kelas, yaitu antara 0,92 hingga 0,97, mengindikasikan keseimbangan antara presisi dan sensitivitas model. Secara keseluruhan, akurasi model mencapai 0,95 dari total 1.426 data uji, dengan nilai rata-rata makro dan rata-rata tertimbang untuk *precision*, *recall*, dan *f1-score* masing-masing sebesar 0,95. Hasil ini mengonfirmasi bahwa model memiliki performa klasifikasi yang kuat dan konsisten dalam mengenali entitas politik pada berita teks, tanpa adanya bias yang signifikan terhadap salah satu kelas.

Tabel 1. Report Klasifikasi

Label	Precision	Recall	F1-Score	Support
paslon_1	0.94	0.98	0.96	418
paslon_2	0.99	0.96	0.97	629
paslon_3	0.92	0.92	0.92	379
Accuracy			0.95	1426
Macro Avg	0.95	0.95	0.95	1426
Weighted Avg	0.96	0.95	0.95	1426

Pada gambar 6 menunjukkan pengujian model terhadap sampel data uji menunjukkan tingkat akurasi prediksi yang sangat tinggi, di mana seluruh pasangan input-output pada 10 data pertama berhasil diklasifikasikan dengan benar sesuai label aslinya. Prediksi yang ditampilkan memperlihatkan konsistensi model dalam mengenali fitur linguistik yang mengindikasikan afiliasi berita terhadap pasangan calon tertentu, baik untuk *paslon_1*, *paslon_2*, maupun *paslon_3*. Secara khusus, mayoritas prediksi berhasil mengidentifikasi berita *paslon_2* secara akurat, tanpa kesalahan klasifikasi. Selain itu, model juga mampu membedakan berita *paslon_3* dan *paslon_1* dengan tingkat ketepatan yang sama. Hasil ini menegaskan bahwa model yang telah dilatih tidak hanya menunjukkan performa kuantitatif yang unggul pada metrik evaluasi umum seperti akurasi dan *loss*, tetapi juga dapat diandalkan dalam aplikasi praktis untuk klasifikasi *real-time* terhadap data baru. Keberhasilan klasifikasi pada skala sampel memperkuat bukti bahwa pendekatan LSTM yang diterapkan memiliki daya generalisasi yang baik terhadap variasi bahasa dan gaya penulisan dalam berita politik daring.

```

1/1 ----- 1s 1s/step
Original: paslon_2, Predict: paslon_2, ==> Benar
Original: paslon_3, Predict: paslon_3, ==> Benar
Original: paslon_2, Predict: paslon_2, ==> Benar
Original: paslon_2, Predict: paslon_2, ==> Benar
Original: paslon_3, Predict: paslon_3, ==> Benar
Original: paslon_2, Predict: paslon_2, ==> Benar
Original: paslon_1, Predict: paslon_1, ==> Benar
Original: paslon_2, Predict: paslon_2, ==> Benar
    
```

Gambar 6. Pengujian sampel data uji

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan pendekatan *Natural Language Processing (NLP)* menggunakan arsitektur *Long Short-Term Memory (LSTM)* untuk mengklasifikasikan berita pemilu berdasarkan entitas pasangan calon presiden dan wakil presiden. Tujuan yang dirumuskan dalam bab Pendahuluan, yakni membangun model klasifikasi otomatis untuk mengenali afiliasi politik dalam teks berita, telah tercapai sebagaimana dibuktikan melalui hasil dan analisis model. Model yang dikembangkan menunjukkan performa

yang sangat baik, dengan akurasi validasi mencapai 95,44%, dan *f1-score macro_avg* serta *weight avg* sebesar 0,95. *Confusion matrix* dan *classification report* juga memperkuat bukti keberhasilan model dalam melakukan klasifikasi yang akurat dan stabil terhadap ketiga kelas entitas politik yang diamati.

REFERENSI

- [1] A. Sanaullah, A. Das, A. Das, M. Kabir, & K. Shu, "Applications of machine learning for covid-19 misinformation: a systematic review", *Social Network Analysis and Mining*, vol. 12, no. 1, 2022. <https://doi.org/10.1007/s13278-022-00921-9>
- [2] A. Altheneyan and A. Alhadlaq, "Big data ml-based fake news detection using distributed learning", *Ieee Access*, vol. 11, p. 29447-29463, 2023. <https://doi.org/10.1109/access.2023.3260763>
- [3] K. Nagy and J. Kapusta, "Twidw—a novel method for feature extraction from unstructured texts", *Applied Sciences*, vol. 13, no. 11, p. 6438, 2023. <https://doi.org/10.3390/app13116438>
- [4] N. Mohd, H. Singhdev, & D. Upadhyay, "Text classification using cnn and cnn-lstm", 2021. <https://doi.org/10.29121/web/v18i4/149>
- [5] M. Lestandy, "Effect of word2vec weighting with cnn-bilstm model on emotion classification", *Jurnal Nasional Pendidikan Teknik Informatika (Janapati)*, vol. 12, no. 1, p. 99-107, 2023. <https://doi.org/10.23887/janapati.v12i1.58571>
- [6] R. Jain, "Exploring the efficacy of natural language processing and supervised learning in the classification of fake news articles", *ART*, vol. 2, no. 1, p. 1-6, 2024. <https://doi.org/10.23880/art-16000108>
- [7] D. Yefferson, V. Lawijaya, & A. Girsang, "Hybrid model: indobert and long short-term memory for detecting indonesian hoax news", *Jaes International Journal of Artificial Intelligence (Ij-Ai)*, vol. 13, no. 2, p. 1913, 2024. <https://doi.org/10.11591/ijai.v13.i2.pp1913-1924>
- [8] C. Liu, "Long short-term memory (lstm)-based news classification model", *Plos One*, vol. 19, no. 5, p. e0301835, 2024. <https://doi.org/10.1371/journal.pone.0301835>
- [9] A. Erlich, D. Jung, & J. Long, "Covering the campaign: computational tools for measuring differences in candidate and party news coverage with application to an emerging democracy", *Social Science Computer Review*, vol. 42, no. 6, p. 1313-1337, 2024. <https://doi.org/10.1177/08944393241247420>
- [10] R. Yunida, M. Faisal, M. Muliadi, F. Indriani, F. Abadi, I. Budimanet al., "Lstm and bi-lstm models for identifying natural disasters reports from social media", *Journal of Electronics Electromedical Engineering and Medical Informatics*, vol. 5, no. 4, 2023. <https://doi.org/10.35882/jeeemi.v5i4.319>
- [11] R. Adipradana, B. Nayoga, R. Suryadi, & D. Suhartono, "Hoax analyzer for indonesian news using rnns with fasttext and glove embeddings", *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, p. 2130-2136, 2021. <https://doi.org/10.11591/eei.v10i4.2956>
- [12] C. Durga, "Nbesm: a novel bert integrated with sequence model for classifying the news in telugu", *jes*, vol. 20, no. 7s, p. 3988-3999, 2024. <https://doi.org/10.52783/jes.4493>
- [13] M. T. Khan, H. Ali, and A. Rauf, "Comparative analysis of LSTM and GRU models for entity-based classification in political discourse," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 117–125, 2023, doi: 10.14569/IJACSA.2023.0140215.
- [14] R. D. Putra and P. I. Santosa, "Enhancing political news classification with Word2Vec and deep learning models," *Journal of Information Systems Research*, vol. 10, no. 3, pp. 45–53, 2022, doi: 10.24014/jisr.v10i3.4215.
- [15] A. Rahman and C. H. Lee, "Multi-label LSTM-based entity extraction and classification for public policy documents," *Applied Artificial Intelligence*, vol. 37, no. 4, pp. 225–241, 2023, doi: 10.1080/08839514.2023.2186239.
- [16] M. R. Fauzan and D. Permana, "Natural language processing-based classification of Indonesian political news: A deep learning approach," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 1, pp. 67–74, 2021, doi: 10.14710/jtsiskom.9.1.2021.67-74.
- [17] T. Wijaya, Y. Ramadhani, and N. Maulidina, "Integrating NLP and entity recognition for adaptive classification of electoral news in Indonesian media," *Journal of Data Science and Analytics*, vol. 6, no. 2, pp. 89–101, 2023, doi: 10.24867/JDSA-2023-6-2-89.