

## Optimasi *Hyperparameter* Convolutional Neural Network-Long Short-Term Memory dengan Fitur GloVe untuk Klasifikasi Berita Palsu

Winda Kurnia \Sari<sup>1</sup>, Iman Saladin B. Azhar<sup>2</sup>, Nurul Afifah<sup>3</sup>

<sup>1</sup> Departemen Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya  
<sup>2,3</sup> Departemen Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya  
e-mail: windakurniasari@unsri.ac.id, imansaladin@unsri.ac.id, nurul@unsri.ac.id

### Abstrak

Dalam upaya mengatasi penyebaran berita palsu di era digital yang semakin meresahkan, penelitian ini berfokus pada pengembangan metode untuk meningkatkan akurasi dalam mengidentifikasi dan menyaring berita palsu. Melalui penggunaan teknik klasifikasi teks dengan bantuan word embedding GloVe dan model Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM), penelitian ini berhasil mencapai akurasi sebesar 96.87% melalui tuning hyperparameter. Pendekatan ini mengintegrasikan analisis teks dengan kecerdasan buatan untuk memperkuat kemampuan membedakan antara berita palsu dan berita yang sah. Metode word embedding GloVe digunakan untuk memperkaya representasi kata dalam teks, sementara model CNN-LSTM memungkinkan pengambilan fitur-fitur penting dari teks dan memperhitungkan urutan kata-kata. Dengan hasil eksperimen yang menunjukkan potensi besar dari pendekatan ini, diharapkan bahwa teknik-teknik ini dapat membantu dalam menghadapi tantangan yang kompleks terkait penyebaran berita palsu di lingkungan digital. Dengan terus berkembangnya kecerdasan komputasional, upaya-upaya seperti ini diharapkan akan terus melahirkan solusi yang lebih efektif untuk menciptakan lingkungan online yang lebih aman dan terpercaya.

**Kata kunci:** Berita palsu, GloVe, Hyperparameter, CNN-LSTM

### Abstract

In an effort to combat the spread of fake news in the increasingly worrisome digital era, this research focuses on developing methods to enhance accuracy in identifying and filtering fake news. Through the use of text classification techniques aided by GloVe word embedding and Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) models, this study has achieved an accuracy of 96.87% through hyperparameter tuning. This approach integrates text analysis with artificial intelligence to strengthen the ability to distinguish between fake and genuine news. GloVe word embedding methods are utilized to enrich the representation of words in text, while the CNN-LSTM model allows for the extraction of important features from the text and consideration of word sequences. With experimental results showcasing the significant potential of this approach, it is hoped that these techniques can aid in addressing the complex challenges associated with the spread of fake news in the digital environment. With the continued advancement of computational intelligence, efforts like these are expected to continue evolving to create a safer and more trustworthy online environment.

**Keywords:** Fake news, GloVe, hyperparameter, CNN-LSTM

## 1. PENDAHULUAN

Penyebaran berita palsu telah menjadi permasalahan serius dalam era digital saat ini, mempengaruhi opini publik dan merusak kepercayaan masyarakat terhadap informasi. Fenomena ini memicu polarisasi dan dapat dengan mudah disebarkan melalui media sosial tanpa verifikasi kebenaran yang memadai. Dampak negatifnya dapat

dirasakan di berbagai sektor, dari politik hingga kesehatan masyarakat. Oleh karena itu, penelitian deteksi dan penanganan berita palsu menjadi sangat penting. Langkah-langkah dalam hal ini diperlukan untuk memperkuat integritas informasi di era digital [1].

Dalam penelitian [2] memperkenalkan model FNDNet, sebuah jaringan saraf konvolusi mendalam yang dirancang untuk deteksi berita palsu, mencapai akurasi yang mengesankan sebesar 98,36% pada dataset Kaggle. Sementara itu, [3] memeriksa kredibilitas berita yang terdeteksi melalui komentar pada item-item tertentu, mencapai akurasi masing-masing 90,4% dan 80,80% pada dataset PolitiFact dan GossipCop. Di sisi lain, [4] menemukan bahwa model WELFake mengategorikan berita sebagai nyata atau palsu dengan akurasi yang luar biasa sebesar 96,73%. Ini mewakili peningkatan dalam akurasi keseluruhan sebesar 1,31% dibandingkan dengan representasi enkoder berarah dari *transformer* (BERT) dan 4,25% dibandingkan dengan model Convolutional Neural Network (CNN). Selain itu, pendekatan berbasis frekuensi dan analisis yang difokuskan pada pola penulisan mereka melampaui metode berbasis prediksi yang menggunakan teknik *word embedding* (WE) Word2vec hingga 1,73%. Penelitian lain dengan dataset WELFake [5] menggunakan pembelajaran mesin dan pemrosesan bahasa alami untuk mengumpulkan berita dan mencoba menggunakan mesin pendukung vektor untuk menentukan apakah berita tersebut nyata atau palsu. Opsi yang diusulkan dibandingkan dengan model yang sudah ada. Versi yang diusulkan berfungsi dengan baik dan menentukan akurasi hingga 93,6%. Penelitian [6] untuk memastikan model memiliki akurasi setinggi model SOTA saat ini, tiga algoritma pembelajaran mesin diterapkan untuk klasifikasi teks. Analisis dilakukan terhadap Baseline Kelas Mayoritas, Klasifikasi Random Forest dengan kumpulan kata-kata, dan ensemble Model Empat Model Dua Lipat dari ensemble Lima Model Lima Lipat. Selanjutnya, penelitian ini menguji data secara berurutan dengan Model BERT yang Disesuaikan Secara Manual.

Convolutional Neural Network (CNN) memiliki keunggulan dalam mengekstraksi fitur spasial dari data, yang sangat cocok untuk pengolahan citra dan pola spasial dalam teks [7],[8]. Kemampuan ini memungkinkan CNN untuk secara efisien menangkap pola-pola lokal dalam teks yang berkontribusi pada klasifikasi. Long Short-Term Memory (LSTM), di sisi lain, sangat baik dalam menangani urutan data, seperti teks, karena kemampuannya untuk mengingat informasi jangka panjang dari urutan data serta mengatasi masalah *vanishing gradient* pada jaringan rekursif tradisional [9],[10]. Ketika digabungkan dalam model CNN-LSTM, kelebihan keduanya dapat dioptimalkan. CNN digunakan untuk mengekstraksi fitur dari representasi teks spasialnya, sementara LSTM digunakan untuk memahami konteks dari urutan kata-kata yang diberikan. Dengan cara ini, CNN-LSTM memungkinkan pengambilan fitur-fitur penting dari teks serta memperhitungkan urutan kata-kata, sehingga meningkatkan akurasi klasifikasi teks [11]. Gabungan ini memungkinkan model untuk mengatasi tantangan dalam klasifikasi teks yang melibatkan pemahaman makna konteks dan pola urutan kata-kata, sambil mempertahankan keunggulan dalam menangkap fitur spasial dan lokal dari teks. Ini membuat CNN-LSTM menjadi pilihan yang kuat untuk kasus klasifikasi teks, terutama di dalam bidang seperti deteksi berita palsu di media sosial.

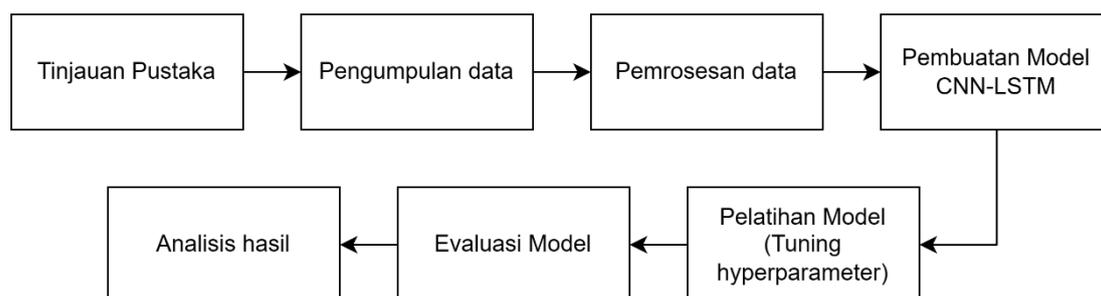
Penelitian ini mengaplikasikan model CNN-LSTM yang dioptimalkan dengan *word embedding* GloVe dan *tuning hyperparameter* untuk meningkatkan akurasi klasifikasi berita palsu. Dengan memanfaatkan fitur-fitur unggulan CNN-LSTM,

penggunaan GloVe sebagai representasi vektor kata, dan penyesuaian *hyperparameter*, penelitian ini berpotensi memberikan kontribusi penting dalam deteksi berita palsu di era digital.

## 2. METODOLOGI PENELITIAN

### 2.1. Metodologi yang diusulkan

Dalam metodologi penelitian, langkah-langkah yang diterapkan melibatkan penggunaan kerangka kerja terstruktur untuk memandu setiap tahap studi. Kerangka kerja penelitian yang digunakan, seperti yang digambarkan dalam Gambar 1, terdiri dari beberapa fase. Langkah awal melibatkan tinjauan pustaka untuk mengevaluasi studi terbaru dalam rentang waktu 1 hingga 5 tahun terakhir. Selanjutnya, dataset WELFake dengan lebih dari 72.000 sampel data digunakan dalam tahap pengumpulan data. Setelah itu, dilakukan *embedding* dengan fitur GloVe 300d, serta tokenisasi. Tahap pembuatan model CNN-LSTM kemudian dilakukan, diikuti dengan pelatihan model dengan *tuning hyperparameter*. Setelah proses pelatihan, model dievaluasi menggunakan nilai presisi, recall, dan f1-score. Akhirnya, hasil evaluasi dianalisis untuk mendapatkan kesimpulan dari penelitian ini.



Gambar 1. Metodologi Penelitian yang diusulkan

### 2.2. Pengumpulan data

Dataset WELFake, yang terdiri dari 72.134 artikel berita, memiliki 35.028 berita nyata dan 37.106 berita palsu. Untuk meminimalkan risiko overfitting dan meningkatkan kualitas pelatihan model machine learning, para peneliti menggabungkan empat dataset berita terkenal seperti Kaggle, McIntire, Reuters, dan BuzzFeed Political [4]. Dataset ini mencakup empat kolom utama: nomor seri, judul berita, isi berita, dan label yang menandakan apakah berita tersebut nyata (label 1) atau palsu (label 0).

### 2.3. Pemrosesan data

Pemrosesan pada penelitian ini melibatkan penggunaan word *embedding* GloVe dengan 300 dimensi. GloVe, singkatan dari *Global Vectors for Word Representation*, adalah teknik *embedding* kata yang menggunakan model pembelajaran mesin untuk memetakan kata ke dalam ruang vektor yang disematkan. Dalam representasi vektor GloVe, kata-kata dengan makna yang serupa cenderung memiliki vektor yang lebih dekat secara geometris, sehingga memungkinkan algoritma untuk memahami konteks dan makna kata secara lebih baik dalam pemrosesan teks.

Tabel 1. Spesifikasi Dataset

Dataset	Berita asli	Berita palsu
Kaggle	10387	10413
McIntire	3171	3164
Reuters	21417	23481
BuzzFeed Politicial	53	48
<b>WELFake dataset</b>	<b>35028</b>	<b>37106</b>

#### 2.4. Pembuatan Model CNN-LSTM

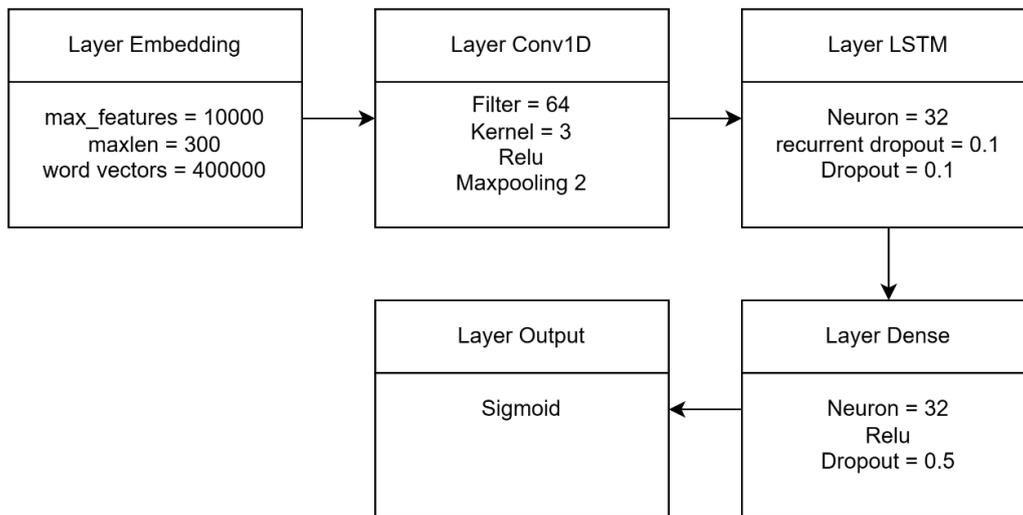
Setelah tahap pra-pemrosesan data dan pembagian data selesai, langkah berikutnya adalah pembuatan model CNN-LSTM. Proses ini melibatkan tokenisasi, di mana teks dari data disegmentasi menjadi token-token yang merepresentasikan unit-unit semantik seperti kata-kata atau frasa. Dengan memanfaatkan model CNN-LSTM, sistem dapat memperoleh pemahaman yang lebih baik tentang struktur dan hubungan antar token dalam teks, memungkinkan pengambilan fitur-fitur penting dan pemodelan urutan kata-kata secara efektif untuk tujuan klasifikasi atau analisis teks.

#### 2.5. Pelatihan Model (*Tuning hyperparameter*)

Setelah pembentukan model CNN-LSTM, langkah selanjutnya adalah melatih data dengan *tuning hyperparameter* untuk mencapai akurasi yang optimal. Hal ini melibatkan penyesuaian berbagai parameter model seperti learning rate, jumlah neuron dalam lapisan-lapisan jaringan, serta ukuran fitur word *embedding* yang digunakan. Proses *tuning hyperparameter* bertujuan untuk menemukan kombinasi parameter terbaik yang dapat meningkatkan kinerja model dalam mengklasifikasikan atau menganalisis data teks dengan akurasi yang tinggi. Dengan penyesuaian yang tepat, model CNN-LSTM dapat dioptimalkan untuk menangani berbagai macam tugas pemrosesan bahasa alami dengan hasil yang memuaskan. Arsitektur CNN-LSTM yang diusulkan dapat dilihat pada Gambar 2.

#### 2.6. Convolutional Neural Network-Long Short-Term Memory

Dalam konteks pengembangan model klasifikasi teks, CNN-LSTM muncul sebagai varian yang menjanjikan, menggabungkan keunggulan CNN dalam mengekstraksi fitur spasial dengan kemampuan LSTM dalam memahami urutan data. Dilatarbelakangi oleh masalah gradien yang hilang atau meledak dalam RNN tradisional, LSTM memperkenalkan struktur gerbang yang memungkinkan pemrosesan informasi yang lebih efisien dari konteks waktu yang lebih panjang. Selain itu, penelitian yang menggabungkan CNN dan LSTM, seperti model C-LSTM, telah membuktikan kemampuannya dalam mengatasi tugas-tugas klasifikasi teks dengan hasil yang memuaskan, melebihi kinerja model CNN atau LSTM tunggal. Oleh karena itu, dalam menghadapi tantangan kompleks dalam pengolahan teks, pendekatan gabungan antara CNN dan LSTM menjanjikan hasil yang lebih baik dan dapat diandalkan [12], [13].



Gambar 2. Arsitektur CNN-LSTM yang diusulkan

### 3. HASIL DAN ANALISIS

Tabel 2 menampilkan hasil dari pelatihan model CNN-LSTM yang telah *dituning* hyperparameter, menghasilkan akurasi sebesar 96.87%. Akurasi ini menggambarkan tingkat keakuratan model dalam mengklasifikasikan data, dengan akurasi pada data latih mencapai 98.93% dan data uji sebesar 96.88%. Hasil yang diperoleh menunjukkan kemampuan model dalam menggeneralisasi pola dari data latih ke data uji, menunjukkan konsistensi dan kehandalan dalam memberikan prediksi yang akurat.

Tabel 2. Hasil Akurasi Model

	Hasil
Akurasi model latih	98.93%
Akurasi model tes	96.88%
Akurasi model	<b>96.87%</b>

Validasi menggunakan *confusion matrix* juga menunjukkan hasil yang memuaskan dengan nilai presisi, recall, dan F1-score yang tinggi, sebagaimana terlihat pada Gambar 3. Hasil ini memberikan indikasi bahwa model CNN-LSTM yang dikembangkan dalam penelitian ini mampu melakukan klasifikasi pada dataset berita palsu dengan akurasi yang tinggi, serta mampu mengidentifikasi berbagai kelas dengan baik.

Dalam perbandingan dengan penelitian terdahulu, penelitian ini berhasil meningkatkan tingkat akurasi menjadi 96.87% dengan menggunakan model CNN-LSTM dan fitur GloVe, dibandingkan dengan akurasi sebesar 96.73% yang diperoleh dari penelitian sebelumnya yang menggunakan *Word Embedding Word2Vec* [4]. Hal ini menunjukkan bahwa pendekatan yang diusulkan dalam penelitian ini mampu memberikan peningkatan yang signifikan dalam pengklasifikasian berita palsu, serta

meningkatkan kemampuan model untuk memahami dan memproses informasi teks dengan lebih baik.

Model report :				
	precision	recall	f1-score	support
0	0.96	0.97	0.97	8792
1	0.97	0.97	0.97	9242
accuracy			0.97	18034
macro avg	0.97	0.97	0.97	18034
weighted avg	0.97	0.97	0.97	18034

Gambar 3. Hasil *Confusion matrix*

#### 4. KESIMPULAN

Dalam penelitian ini, kami berhasil mengembangkan model CNN-LSTM yang efektif dalam mengklasifikasikan berita palsu dengan akurasi yang tinggi. Dengan melakukan *tuning hyperparameter* dan menggunakan fitur GloVe, kami mencapai akurasi sebesar 96.87%, menunjukkan kemampuan model untuk membedakan antara berita palsu dan berita yang sah dengan sangat baik. Validasi menggunakan *confusion matrix* juga menunjukkan hasil yang memuaskan, dengan nilai presisi, recall, dan F1-score yang tinggi. Temuan ini mengindikasikan bahwa model CNN-LSTM yang dikembangkan dapat menjadi alat yang berguna dalam mengatasi penyebaran berita palsu di media sosial, meningkatkan kualitas informasi yang diterima oleh masyarakat secara keseluruhan.

Untuk penelitian selanjutnya, perlu dilakukan eksplorasi lebih lanjut dalam beberapa aspek. Pertama, dapat dilakukan peningkatan pada model dengan memperluas dataset atau menggunakan teknik pembelajaran yang lebih canggih untuk meningkatkan akurasi klasifikasi. Selain itu, penelitian dapat diarahkan untuk memperdalam pemahaman tentang karakteristik dan pola dari berita palsu itu sendiri, serta mengembangkan strategi yang lebih efektif dalam mengatasinya. Dengan pendekatan yang lebih holistik dan inovatif, diharapkan penelitian selanjutnya dapat memberikan kontribusi yang lebih besar dalam menanggulangi masalah berita palsu di era digital yang semakin kompleks ini.

#### REFERENCES

- [1] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, Sep. 2020.
- [2] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet – A deep convolutional neural network for fake news detection," *Cogn. Syst. Res.*, vol. 61, pp. 32–44, 2020.
- [3] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "dEFEND: Explainable Fake News Detection," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 395–405.
- [4] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "WELFake: Word

- Embedding Over Linguistic Features for Fake News Detection,” IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 4, pp. 881–893, 2021.
- [5] K. V Mahijith, A. U. Sunil, C. A., A. Sreelathan, and M. V Thomas, “A Novel Fake News Detection Approach Using Machine Learning,” in *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICT)*, 2022, pp. 448–454.
- [6] C. She and C. Greenberg, “High Performing Explanatory Fake News Classification on Longer Texts,” *J. Student Res.*, vol. 12, no. 4 SE-, Nov. 2023.
- [7] Z. Liu, H. Huang, C. Lu, and S. Lyu, “Multichannel CNN with Attention for Text Classification.” 2020.
- [8] J. Zhang, F. Liu, W. Xu, and H. Yu, “Feature Fusion Text Classification Model Combining CNN and BiGRU with Multi-Attention Mechanism,” *Future Internet*, vol. 11, no. 11. 2019.
- [9] W. K. Sari, D. P. Rini, R. F. Malik, and I. S. B. Azhar, “Multilabel Text Classification in News Articles Using Long-Term Memory with Word2Vec.” *Jurnal Resti*, pp. 276–285, 2020.
- [10] J. Dai, C. Chen, and Y. Li, “A Backdoor Attack Against LSTM-Based Text Classification Systems,” *IEEE Access*, vol. 7, pp. 138872–138878, 2019.
- [11] Y. Zhang, Z. Jia, J. Yuru, H. Gaijuan, and C. Ruoyu, “A Text Sentiment Classification Modeling Method Based on Coordinated CNN-LSTM-Attention Model\*.” *Chinese Journal of Electronics*, vol. 28, no. 1, pp. 120–126, 2019.
- [12] S. Liang, B. Zhu, Y. Zhang, S. Cheng, and J. Jin, “A Double Channel CNN-LSTM Model for Text Classification,” in *2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2020, pp. 1316–1321.
- [13] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, and P. Hellinckx, “CNN-LSTM architecture for predictive indoor temperature modeling,” *Build. Environ.*, vol. 206, p. 108327, 2021.