

Enhancing Hoax News Detection Performance through Modified Convolutional Neural Network Architecture

Siti Mujilahwati^{1*}, Danang Bagus Reknadi²

^{1,2} Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Lamongan, Lamongan, 62213, Indonesia

Informasi Artikel

Diterima : 18 Mei 2025
Revisi : 02 Juni 2025
Publikasi : 20 Juni 2025

Kata Kunci:

Deteksi Hoaks
CNN
Optimalisasi
Word2vec
TurnBackHoax

ABSTRAK

Penyebaran berita palsu (hoaks) di internet semakin meluas seiring meningkatnya penggunaan internet di Indonesia. Untuk mendeteksi hoaks secara otomatis, penelitian ini mengusulkan model Convolutional Neural Network (CNN) berbasis data dari situs Detik.com dan TurnBackHoax.id. Data meliputi berbagai kategori seperti politik, agama, dan teknologi. Model dikembangkan melalui beberapa tahapan optimasi: tokenisasi dan padding teks, pemanfaatan Word2Vec pre-trained sebagai bobot embedding, pembagian data secara stratified, penanganan ketidakseimbangan kelas dengan class weight, serta penerapan arsitektur CNN bertingkat dengan dropout. Pelatihan dilakukan menggunakan Adam optimizer dengan learning rate kecil dan teknik early stopping. Hasil evaluasi menunjukkan model usulan mencapai akurasi, presisi, recall, dan F1-score sebesar 96%, dengan waktu pelatihan tercepat 100.42 detik. Model juga dievaluasi dengan model ROC_AUC dengan score 98.91%. Performa ini mengungguli model CNN-1D dan Augmentasi-CNN. Pendekatan ini terbukti efektif dan efisien dalam mendeteksi berita hoaks Bahasa Indonesia.

ABSTRACT

The spread of fake news (hoaxes) on the internet is increasingly widespread, along with the increasing use of the internet in Indonesia. To detect hoaxes automatically, this study proposes a Convolutional Neural Network (CNN) model based on data from the Detik.com and TurnBackHoax.id sites. The data includes various categories such as politics, religion, and technology. The model was developed through several optimization stages: tokenization and text padding, utilization of pre-trained Word2Vec as embedding weights, stratified data division, handling class imbalance with class weights, and implementing a multilevel CNN architecture with dropout. Training was carried out using the Adam optimizer with a small learning rate and an early stopping technique. The evaluation results showed that the proposed model achieved an accuracy, precision, recall, and F1-score of 96%, with the fastest training time of 100.42 seconds. The model was also evaluated with the ROC_AUC model, with a score of 98.91%. This performance outperformed the CNN-1D and Augmentasi-CNN models. This approach has proven to be effective and efficient in detecting Indonesian hoax news.

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



*Penulis Koresponden

Email: moedjee@unisla.ac.id

Cara sitasi IEEE::

S. Mujilawati, dan D. B. Reknadi, "Enhancing Hoax News Detection Performance through Modified Convolutional Neural Network Architecture," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 2, p. 642-650, Juni 2025. doi:10.30811/jaise.v5i2.6953

1. PENDAHULUAN

Seiring perkembangan penggunaan internet semakin meluas, baik dari kalangan perkotaan maupun pedesaan, orang tua, remaja bahkan anak-anak, membawa dampak yang signifikan pada penyebaran berita secara cepat, bahkan tanpa filter. Di Indonesia secara sistematis belum dapat melakukan penyaringan apakah berita palsu atau fakta yang disebar di internet. Dampak pada penyebaran berita palsu ini dapat menimbulkan kebencian yang berlebihan terhadap sesuatu. Oleh sebab itu berbagai cara telah dilakukan untuk mendeteksi berita palsu[1].

Salah satu upaya dari pencegahan berita palsu ini adalah adanya komunitas online yang bertujuan memerangi informasi palsu di internet yaitu dengan nama TurnBackHoax yang beralamatkan situs online di Turnbackhoax.id[2], [3]. Situs ini berisi arsip rangkuman berita palsu dan juga diskusi dari forum anti fitna, hasut, dan hoax (FAFHH). Data dari situs ini memiliki kategori yang beragam, mulai dari politik, olah raga, agama, teknologi dan sebagainya. Dari situs tersebut, maka didapatkan data rangkuman berita hoax yang dapat dijadikan acuan pada deteksi berita hoax dari internet[4].

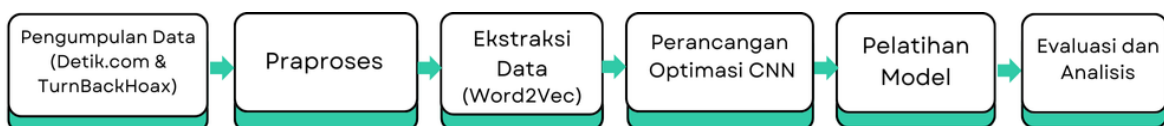
Beberapa penelitian sebelumnya yang mendeteksi berita palsu terutama dalam Bahasa Indonesia telah banyak memanfaatkan berbagai metode mesin *learning* dan juga *deep learning*. penelitian dari Noor et al.2023 telah melakukan augmentasi data untuk deteksi hoax dengan menggunakan metode *Convolutional Neural Network* (CNN). Metode yang digunakan pada penelitian mereka menggunakan model arsitektur CNN-1D yang sebelumnya dilakukan model pengubahan kata yang sama dan melakukan *Random Insertion* (RI), serta menggunakan *Random Deletion* (RD). pengujian yang dilakukan dengan model CNN-1D mendapat akurasi sebesar 82.81%, Precision 83.66%, Recall 98.37% dan F1-Score 90.42% [2]. Meskipun penelitian mereka menunjukkan hasil performa yang cukup baik, namun pendekatan augmentasi yang dilakukan seperti *Random Insertion* dan *Random Deletion* belum tentu menghasilkan representasi semantik yang relevan, sehingga hasil klasifikasi dapat terganggu jika tidak disesuaikan dengan konteks Bahasa Indonesia.

Penelitian Lainnya juga melakukan penelitian serupa deteksi berita hoax dengan menggunakan model *deep learning* yang di kembangkan, Teknik yang di usulkan pada penelitian mereka setelah dilakukan pengujian model dan di evaluasi dengan model ROC_AUC sebesar 99.9%, skor F1 dan akurasi 98.8% [5]. Penelitian lainnya dari Arianto dkk, 2021 juga mencoba menggunakan metode K-Means dan *Support Vector Machine* (SVM) dengan berbagai macam kernel serta *multilayer perceptron* untuk mendapatkan hasil optimasi klasifikasi berita. Pengujian model yang diusulkan mendapatkan hasil score akurasi rata-rata 60% dan akurasi maksimal sebesar 80% [6]. Penelitian terbaru untuk mendeteksi berita palsu dengan menggunakan *Fine Tuning* IndoBERT, pertama dilakukan preproses, selanjutnya dilakukan pelatihan dan evaluasi dengan model yang diusulkan IndoBERT dapat melakukan deteksi berita palsu tentang politik dengan score akurasi sebesar 94.1% dan ROC_AUC 99.1% [7]. Penelitian ini menggunakan model transformer memerlukan sumber daya komputasi yang lebih besar serta waktu pelatihan yang lebih lama, sehingga kurang efisien untuk skenario implementasi nyata di lingkungan dengan keterbatasan sumber daya.

Oleh karena itu, dibutuhkan pendekatan alternatif yang tetap mengedepankan akurasi dan efektivitas, namun lebih ringan dan efisien untuk diterapkan, seperti optimasi arsitektur CNN dengan teknik tambahan yang mampu meningkatkan kinerja klasifikasi. Dengan pendekatan ini, diharapkan hasil penelitian dapat memberikan kontribusi pada pengembangan sistem deteksi berita hoaks yang lebih ringan, efisien, dan akurat dalam konteks Bahasa Indonesia, serta dapat digunakan secara luas dalam aplikasi dunia nyata.

2. METODE

Penelitian ini bertujuan untuk mengembangkan sistem deteksi berita hoaks berbahasa Indonesia yang efisien dan akurat dengan pendekatan optimasi arsitektur *Convolutional Neural Network* (CNN). Metodologi yang digunakan dalam penelitian ini terdiri dari beberapa tahapan utama dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Data berita hoaks diperoleh dari situs berita Detik.com dan TurnBackHoax.id, yaitu platform komunitas daring yang menyediakan arsip verifikasi berita palsu dalam berbagai kategori seperti politik, sosial, teknologi, dan agama. Dari kedua sumber data tersebut dikumpulkan menjadi dataset sebanyak 4.312 data berita. Dataset ini akan diekstraksi dan dikategorikan menjadi dua kelas: hoaks (0) dan fakta (1). Berikut contoh data yang telah dikumpulkan.

Tabel 1. Contoh Dataset

Id Berita	Isi Berita	Kategori	Kelas
1	Pandangan jauh ke depan (foresight) perlu dimiliki oleh para Capres-Cawapres, karena bangsa kita berada pada situasi dunia yang semakin rentan dan kompleks.	Politik	1
2	MENYESATKAN, isi video BUKAN tentang pelantikan Adian Napitupulu menjadi Menteri. FAKTA: video sebenarnya berisi pembacaan yang mentranskrip artikel berisi tulisan opini dengan potongan-potongan video yang [...]	Politik	0
3	Minyak biji Sacha Inchi diprediksi akan menjadi tren dalam beberapa tahun ke depan. Empat mahasiswa UNS kembangkan bahan ini untuk pencegahan stunting.	Teknologi	1
4	hasil periksa fakta Rahmah a n (UIN Sunan Ampel Surabaya). Faktanya, lokasi tersebut adalah taman hiburan di Cina Barat Daya. Pengunjung dapat memindai kode QR [...]	Teknologi	0
5	Atlet para atletik Saptoyogo Purnomo mengaku puas setelah mencatatkan hasil gemilang di multievent Asian Para Games. Kini, ia membidik target lain. Apa itu?	Olahraga	1
6	"Hasil Periksa Fakta Luthfiah Oktari Jasmien (Institut Agama Islam Negeri Surakarta). Informasi tersebut tidak benar. Faktanya, Kepala Bagian Humas Pemkot Surabaya telah memastikan informasi tersebut [...]	Olahraga	0

2.2 Pra-pemrosesan Teks

Sebelum dilakukan pelatihan model, data teks akan melalui tahapan pra-pemrosesan untuk meningkatkan kualitas input. Tahapan yang digunakan antara lain [8], [9], [10] Pembersihan teks (menghapus simbol, HTML tag, URL, angka, dan tanda baca yang tidak relevan), *Case folding* (mengubah huruf menjadi huruf kecil semua), Tokenisasi, *Stopword removal*, *Stemming* (menggunakan stemmer Bahasa Indonesia yaitu *Sastrawi*). Seluruh proses yang dilakukan pada eksperimen dalam penelitian ini dengan menggunakan kode python[11].

2.3 Ekstraksi Fitur

Teks yang telah diproses akan dikonversi menjadi representasi numerik menggunakan metode *embedding* Word2Vec [12], [13] yang telah dilatih pada korpus Bahasa Indonesia. Representasi ini akan digunakan sebagai input ke dalam arsitektur CNN-1D. berikut adalah parameter yang akan digunakan pada penelitian ini.

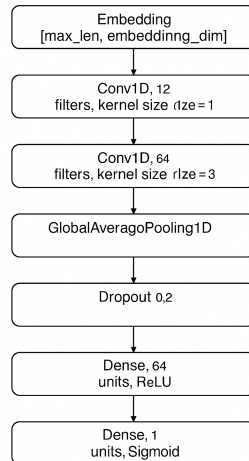
Tabel 2. Parameter Word2vec

No	Parameter	Nilai Umum (default / contoh)	Penjelasan
1	sentences	Tokenized text (list of list)	Kalimat yang sudah dipisah per kata
2	vector_size	100	Ukuran dimensi embedding vector (juga disebut embedding_dim)
3	window	5	Ukuran jendela konteks (berapa kata kiri-kanan diperhatikan)
4	min_count	1 atau 5	Minimum frekuensi kata agar dimasukkan ke dalam model
5	workers	4 (contoh)	Jumlah thread untuk paralelisme training
6	sg	1	1 untuk Skip-gram, 0 untuk CBOW
7	epochs	5 atau lebih	Jumlah iterasi training pada seluruh data
8	seed	42 (opsional)	Untuk reproduibilitas
9	negative	5 atau 10	Jumlah negative sampling
10	sentences	Tokenized text (list of list)	Kalimat yang sudah dipisah per kata

2.4 Perancangan dan Optimasi Model CNN

Arsitektur CNN-1D akan dikembangkan dan dioptimasi melalui eksplorasi beberapa konfigurasi, antara lain [14], [15] Variasi jumlah filter, kernel size, dan stride. Penambahan layer dropout untuk mencegah overfitting. Penambahan layer pooling (max-pooling atau global average pooling). Eksperimen kombinasi

jumlah convolutional layers. Penggunaan activation function (ReLU) dan teknik early stopping. Berikut adalah Gambar 2 desain arsitektur CNN yang dioptimasi dalam penelitian ini.



Gambar 2. Desain Arsitektur CNN yang dioptimasi

2.5 Pelatihan dan Evaluasi Model

Model CNN dilatih dengan data pelatihan menggunakan algoritma *Adam Optimizer* dan fungsi binary cross-entropy. Pembagian data pelatihan, test dan validasi adalah sebagai berikut.

Tabel 3. *Splitting* Data Pelatihan, Pengujian dan Validasi

No	Data	Presentasi	Jumlah
1	Pelatihan	80%	3.449
2	Pengujian	20%	863
3	Validasi	20% dari Pelatihan	690

Evaluasi dilakukan pada data uji menggunakan metrik [16], [17]:

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

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

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

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

Keterangan:

- TP: True Positive
- TN: True Negative
- FP: False Positive
- FN: False Negative

AUC dihitung secara numerik, tidak ada satu rumus eksak, tetapi secara konseptual merupakan probabilitas bahwa model akan memberikan skor lebih tinggi untuk sampel positif daripada sampel negatif [18]. Hasil model yang paling optimal akan dibandingkan dengan baseline model dari penelitian sebelumnya (seperti CNN dengan augmentasi sederhana) untuk menunjukkan peningkatan performa dan efisiensi.

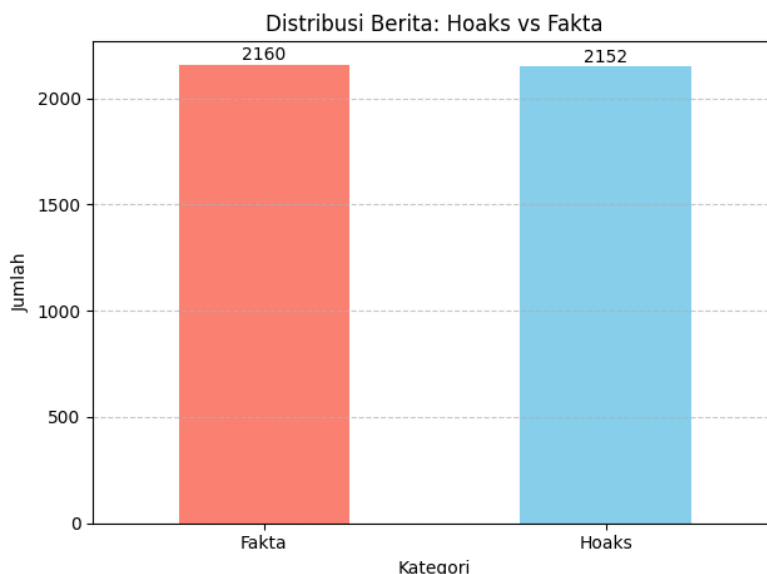
2.6 Analisis dan Interpretasi Hasil

Hasil evaluasi dianalisis untuk mengetahui sejauh mana optimasi CNN mampu meningkatkan akurasi dan efisiensi model dalam mendeteksi hoaks berbahasa Indonesia. Performa juga dievaluasi dari segi kompleksitas model, waktu pelatihan, dan kemungkinan implementasi pada sistem nyata dengan keterbatasan sumber daya [19], [20].

3. HASIL DAN PEMBAHASAN

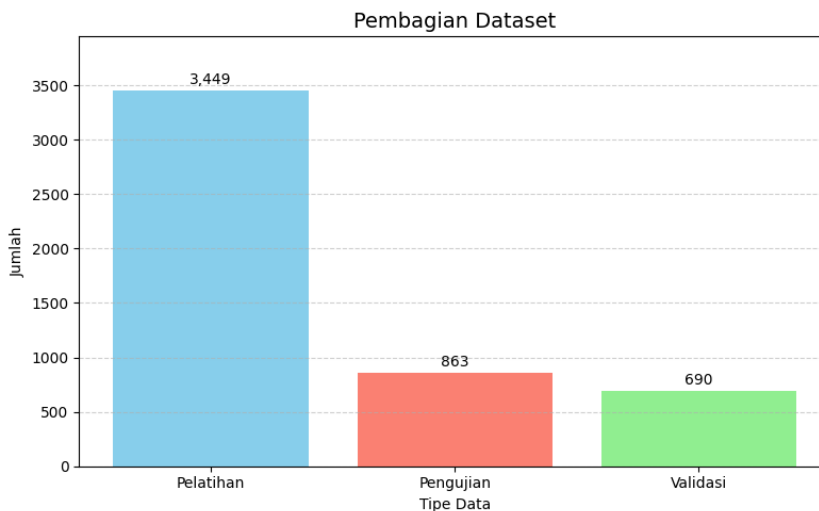
3.1 Hasil Pembagian Data

Dataset yang digunakan hasil dari pengumpulan data sebanyak 4.312 dari proses persiapan data mulai menggabungkan dari dua sumber data yaitu dari detik.com dan turnbackhoax.com. Dimana dalam hal ini data dari sumber turnbackhoax diberi label 0, artinya adalah kelas hoaks. Sedangkan data dari detik.com diberikan label 1, artinya kelas fakta. Dapat dilihat jumlah data dari kedua sumber tersebut pada Gambar 3 berikut.



Gambar 3. Jumlah Dataset Hoaks vs Fakta

Dari gambar terlihat bahwa jumlah berita hoaks sebanyak 2.152 dan berita fakta sebanyak 2.160. Dari jumlah data ini selanjutnya dilakukan *splitting* atau pembagian data pelatihan, pengujian dan juga validasi. Distribusi data ini ditampilkan seperti pada Gambar 4 berikut ini.



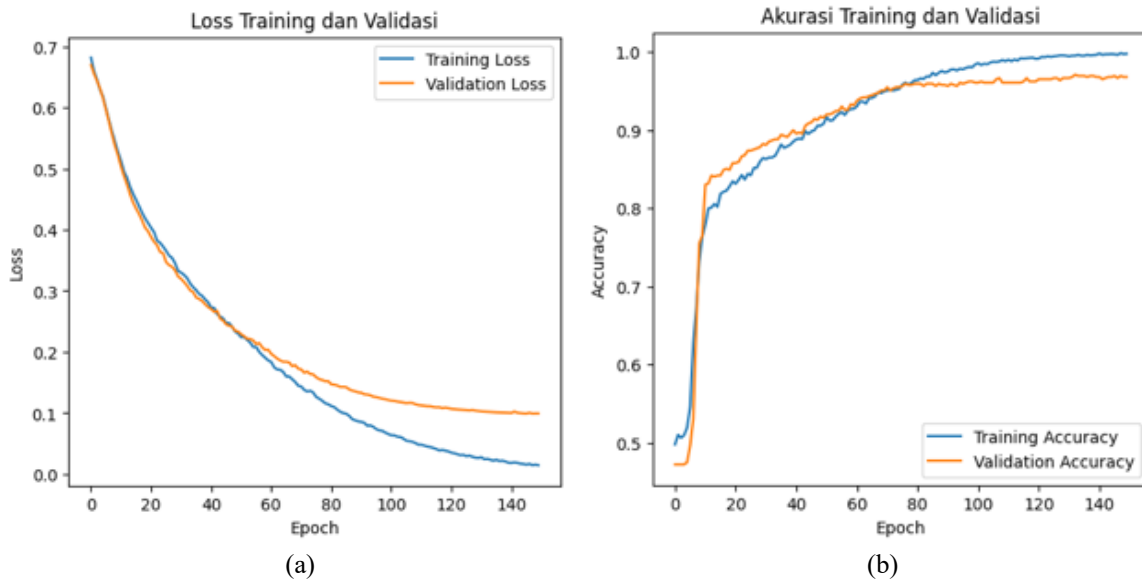
Gambar 4. Pembagian Jumlah Data

Dalam proses pengembangan model pembelajaran mesin, pembagian dataset merupakan langkah krusial untuk memastikan bahwa model yang dibangun mampu belajar dengan baik sekaligus melakukan generalisasi terhadap data baru. Pada skenario ini, dataset dibagi ke dalam tiga bagian utama, yaitu data pelatihan (*training*), data pengujian (*testing*), dan data validasi (*validation*). Sebagian besar data, yaitu sebesar 80% dari total dataset atau sebanyak 3.449 data, dialokasikan untuk proses pelatihan. Tujuan dari data pelatihan ini adalah untuk memungkinkan model mempelajari pola-pola atau fitur penting dari data sehingga dapat melakukan prediksi yang akurat.

Sementara itu, sebanyak 20% dari keseluruhan data, yaitu 863 data, digunakan sebagai data pengujian. Data pengujian ini bersifat independen dan tidak dilibatkan dalam proses pelatihan, sehingga dapat memberikan gambaran objektif mengenai performa akhir model terhadap data yang belum pernah dilihat sebelumnya. Penggunaan data pengujian sangat penting untuk menilai kemampuan generalisasi model dan menghindari bias evaluasi.

Dari total data pelatihan, dilakukan lagi pembagian sebesar 20% untuk keperluan validasi, sehingga diperoleh 690 data yang berfungsi sebagai data validasi. Data validasi digunakan selama proses pelatihan untuk

model. Selain itu, digunakan arsitektur CNN bertingkat dengan dua lapisan konvolusi dan *pooling* yang diikuti oleh lapisan *Dropout* untuk mengurangi risiko *overfitting*.



Gambar 6. Grafik (a) *Loss Training dan Validasi*, (b) *Akurasi Training dan Validasi*

Model juga dikompilasi menggunakan *optimizer Adam* dengan nilai *learning rate* yang kecil (0.00001), agar proses pelatihan berlangsung secara halus dan stabil. Terakhir, diterapkan teknik *EarlyStopping* untuk menghentikan pelatihan secara otomatis ketika *validation loss* tidak menunjukkan perbaikan selama 150 *epoch*. Seluruh strategi tersebut secara kolektif berkontribusi pada hasil pelatihan yang optimal dan kemampuan generalisasi model yang baik terhadap data uji. Tabel 4 berikut adalah hasil evaluasi pengujian dari model yang telah dilatih.

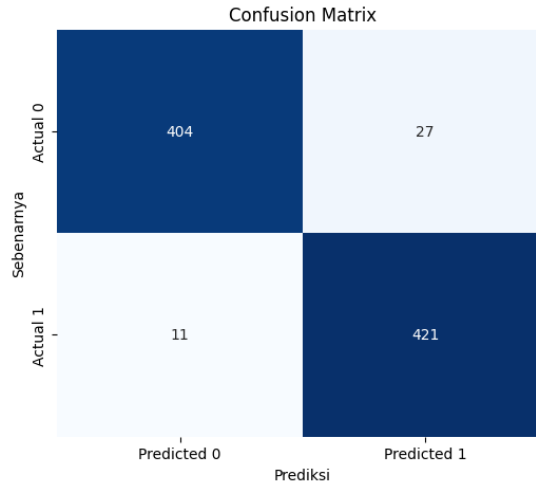
Tabel 4. Hasil Pengujian Model

No	Evaluasi	Kelas	Presentase
1	Akurasi		0.9560
2	Loss		0.1253
3	Recall	0	0.94
		1	0.97
4	Presisi	0	0.97
		1	0.94
5	F1-Score	0	0.96
		1	0.96
6	AUC_Score	-	0.9891

Evaluasi akhir terhadap model CNN menunjukkan hasil yang sangat memuaskan. Model berhasil mencapai nilai *test accuracy* sebesar 95.60% dan *test loss* yang rendah sebesar 0.1253, menandakan bahwa model mampu melakukan generalisasi dengan baik pada data uji. Hasil ini juga tercermin dalam *confusion matrix* (Lihat Gambar 7), di mana dari total 863 sampel uji, model mampu mengklasifikasikan 816 sampel secara benar, dengan hanya 38 kesalahan klasifikasi. Pada kelas 0, model memperoleh *precision* 0.97, *recall* 0.94, dan *F1-score* 0.96, sedangkan pada kelas 1, model menghasilkan *precision* 0.94, *recall* 0.97, dan *F1-score* 0.96. Nilai rata-rata tertimbang (*weighted average*) untuk seluruh metrik evaluasi berada pada angka 0.96, menunjukkan performa yang seimbang antar kelas dan minim bias terhadap kelas mayoritas maupun minoritas. Hasil ini menguatkan bahwa strategi optimasi yang diterapkan, termasuk penggunaan *pretrained embedding*, penyesuaian *class weight*, arsitektur CNN bertingkat, serta teknik regulasi seperti *dropout* dan *early stopping*, berkontribusi signifikan dalam mencapai kinerja model yang optimal dan stabil. Dibuktikan dengan AUC score mencapai 98.91%.

Selain menunjukkan performa klasifikasi yang tinggi, model CNN yang digunakan juga terbukti efisien secara komputasional. Arsitektur yang relatif ringan, dengan hanya dua lapisan konvolusi dan tanpa penggunaan model sekuensial seperti LSTM, memungkinkan proses pelatihan berjalan lebih cepat dan meminimalkan kebutuhan sumber daya komputasi. Penggunaan *GlobalAveragePooling1D* alih-alih *flattening layer* juga membantu mengurangi jumlah parameter tanpa mengorbankan kualitas representasi fitur. Hal ini

memungkinkan model untuk mencapai akurasi tinggi dengan risiko *overfitting* yang rendah dan waktu inferensi yang singkat. Dengan demikian, model ini tidak hanya efektif secara performa, tetapi juga efisien dan layak diterapkan dalam skenario nyata yang memerlukan prediksi cepat dan konsumsi memori yang rendah.



Gambar 7. *Confusion Matrix* Pengujian Model

3.3 Evaluasi dengan Model CNN Sederhana

Pada tahap ini, dilakukan evaluasi terhadap performa model CNN sederhana yang digunakan untuk melakukan klasifikasi teks. CNN dikenal efektif dalam menangkap fitur spasial dari data, termasuk dalam konteks teks di mana pola-pola lokal seperti frasa atau *n-gram* dapat ditangkap melalui lapisan konvolusinya. Evaluasi ini bertujuan untuk menilai sejauh mana model CNN sederhana dapat mengidentifikasi dan mengklasifikasikan data teks berdasarkan fitur yang telah diproses sebelumnya. Hasil evaluasi ini akan menjadi dasar perbandingan dengan model lain serta menjadi tolok ukur efektivitas penggunaan CNN dalam konteks analisis sentimen atau klasifikasi teks yang sedang diteliti. Tabel 5 berikut berbedaan hasil pengujian dengan *baseline*. Dengan menggunakan data yang sama, dan model arsitektur yang berbeda.

Tabel 5. Evaluasi Baseline

Model	Akurasi %	Presisi %	Recall %	F1-Score %	Time Train / detik
CNN-1D [14]	90	88	90	91	165.25
Augmentasi-CNN [2]	83	84	98	91	1825.85
Optimasi CNN (Model yang di usulkan)	96	96	96	96	100.42

Berdasarkan hasil pengujian tiga model dengan parameter yang sama (batch size 32, validasi 0.2, dan 150 epoch), model yang diusulkan, yaitu Optimasi CNN, menunjukkan performa yang paling unggul dibandingkan model sebelumnya, baik dari segi akurasi maupun efisiensi waktu. Model ini mencapai akurasi, presisi, *recall*, dan *F1-score* yang sama tinggi, yaitu sebesar 96%, yang menandakan kemampuan klasifikasi yang sangat baik dan seimbang. Dibandingkan dengan model CNN-1D [14] yang memiliki akurasi 90% dan *Augmentasi-CNN* [2] dengan akurasi 83%, model usulan ini jelas memberikan peningkatan yang signifikan. Menariknya, meskipun memberikan hasil terbaik, waktu pelatihan model Optimasi CNN justru paling singkat, yakni hanya 100.42 detik, lebih cepat dibandingkan CNN-1D (165.25 detik) dan *Augmentasi-CNN* (1825.85 detik). Hal ini menunjukkan bahwa optimasi yang dilakukan pada model sangat efektif dalam meningkatkan performa sekaligus efisiensi. Tanpa membutuhkan teknik *augmentasi* tambahan, model ini mampu menghasilkan klasifikasi yang lebih akurat dan efisien, menjadikannya solusi yang lebih baik dalam tugas deteksi hoaks yang diuji.

4. KESIMPULAN

Berdasarkan hasil evaluasi terhadap tiga arsitektur model CNN yang diuji dengan parameter yang sama, dapat disimpulkan bahwa model Optimasi CNN yang diusulkan memiliki performa terbaik dalam tugas klasifikasi teks, khususnya dalam mendeteksi hoaks. Model ini tidak hanya menunjukkan peningkatan signifikan dalam hal akurasi, presisi, *recall*, dan *F1-score*, yang seluruhnya mencapai nilai tertinggi sebesar 96%. Tetapi juga menunjukkan efisiensi waktu pelatihan terbaik, yaitu hanya 100.42 detik. Hasil ini mengindikasikan bahwa pendekatan optimasi yang dilakukan pada arsitektur CNN mampu meningkatkan efektivitas model secara keseluruhan tanpa memerlukan proses tambahan seperti *augmentasi* data. Dengan

demikian, model yang diusulkan terbukti lebih efisien dan akurat, serta layak dijadikan acuan dalam pengembangan sistem klasifikasi teks berbasis CNN di masa mendatang.

REFERENSI

- [1] B. P. Putra, B. Irawan, C. Setianingsih, F. T. Elektro, U. Telkom, And D. Learning, "Deteksi Ujaran Kebencian Dengan Menggunakan Algoritma Convolutional Neural Network Pada Gambar Hatespeech Detection Using Convolutional Neural Network Algorithm Based On Image," In *E-Proceeding Of Engineering*, Aug. 2018, Pp. 2395–2402.
- [2] A. Z. M. Noor, R. Gernowo, And O. D. Nurhayati, "Data Augmentation For Hoax Detection Through The Method Of Convolutional Neural Network In Indonesian News," *Jurnal Penelitian Pendidikan Ipa*, Vol. 9, No. 7, 2023, Doi: 10.29303/Jppipa.V9i7.4214.
- [3] A. B. Prasetyo, R. R. Isnanto, D. Eridani, Y. A. A. Soetrisno, M. Arfan, And A. Sofwan, "Hoax Detection System On Indonesian News Sites Based On Text Classification Using Svm And Sgd," In *Proceedings - 2017 4th International Conference On Information Technology, Computer, And Electrical Engineering, Icitacee 2017*, 2017. Doi: 10.1109/Icitacee.2017.8257673.
- [4] M. Rifky Maulana *Et Al.*, "Penerapan Metode Lstm Dalam Pembuatan Sistem Pendeteksi Berita Palsu Berbahasa Indonesia," Indonesia, Oct. 2023. Doi: <https://doi.org/10.35706/Syji.V12i02.9924>.
- [5] L. Bodhi Wijaya, Y. Nuraga Wicaksana, S. S. Widhiyari, And A. Saptawijaya, "Pengembangan Model Deteksi Hoaks Berbahasa Indonesia Menggunakan Kombinasi Indobert Dan Bilstm," *Electronics*, Vol. 2, Pp. 12–16, Apr. 2024. Accessed: Jun. 02, 2025. [Online]. Available: <https://bulettingemastik.id/index.php/bg/article/view/202404.003/13>
- [6] R. Arianto, S. W. Harco Leslie, Y. Heryadi, And E. Abdurachman, "Fake News Detection Model Based On Credibility Measurement For Indonesian Online News," *J Theor Appl Inf Technol*, Vol. 99, No. 7, 2021.
- [7] C. Jocelynne, L. Tobing, I. Lanang Wijayakusuma, L. Putu, And I. Harini, "Detection Of Political Hoax News Using Fine-Tuning Indobert," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/jaic>
- [8] R. Feldman And J. Sanger, "Text Mining Preprocessing Techniques," In *The Text Mining Handbook*, Cambridge University Press, 2006, Pp. 57–63. Doi: 10.1017/Cbo9780511546914.004.
- [9] S. Khairunnisa, A. Adiwijaya, And S. Al Faraby, "Pengaruh Text Preprocessing Terhadap Analisis Sentimen Komentar Masyarakat Pada Media Sosial Twitter (Studi Kasus Pandemi Covid-19)," *Jurnal Media Informatika Budidarma*, Vol. 5, No. 2, P. 406, 2021, Doi: 10.30865/Mib.V5i2.2835.
- [10] S. Mujilawati, "Pre-Processing Text Mining Pada Data Twitter," *Seminar Nasional Teknologi Informasi Dan Komunikasi*, Vol. 2016, No. Sentika, Pp. 2089–9815, 2016.
- [11] S. Mujilawati, M. Sholihin, R. Wardhani, And M. R. Zamroni, "Python Based Machine Learning Text Classification," In *Journal Of Physics: Conference Series*, 2022. Doi: 10.1088/1742-6596/2394/1/012015.
- [12] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, And Z. Abidin, "Perbandingan Kinerja Word Embedding Word2vec, Glove, Dan Fasttext Pada Klasifikasi Teks," *Jurnal Tekno Kompak*, Vol. 14, No. 2, P. 74, 2020, Doi: 10.33365/Jtk.V14i2.732.
- [13] M. A. Fauzi, "Word2vec Model For Sentiment Analysis Of Product Reviews In Indonesian Language," *International Journal Of Electrical And Computer Engineering (Ijece)*, Vol. 9, No. 1, P. 525, 2019, Doi: 10.11591/Ijece.V9i1.Pp525-530.
- [14] E. Y. Hidayat And D. Handayani, "Penerapan 1d-Cnn Untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female Daily Review," *Jurnal Nasional Teknologi Dan Sistem Informasi*, Vol. 8, No. 3, Pp. 153–163, Jan. 2023, Doi: 10.25077/Teknosi.V8i3.2022.153-163.
- [15] Moch. A. Nasichuddin, T. B. Adji, And W. Widyawan, "Performance Improvement Using Cnn For Sentiment Analysis," *Ijitee (International Journal Of Information Technology And Electrical Engineering)*, Vol. 2, No. 1, 2018, Doi: 10.22146/Ijitee.36642.
- [16] L. Bodhi Wijaya, Y. Nuraga Wicaksana, S. S. Widhiyari, And A. Saptawijaya, "Pengembangan Model Deteksi Hoaks Berbahasa Indonesia Menggunakan Kombinasi Indobert Dan Bilstm," 2024.
- [17] L. Triyono, R. Gernowo, Prayitno, M. Rahaman, And T. R. Yudiantoro, "Indonesian Fake News Detection Using Various Machine Learning Technique," *International Journal On Informatics Visualization*, Vol. 7, No. 3, 2023, Doi: 10.30630/Joiv.7.3.1243.
- [18] N. Rai, D. Kumar, N. Kaushik, C. Raj, And A. Ali, "Fake News Classification Using Transformer Based Enhanced Lstm And Bert," *International Journal Of Cognitive Computing In Engineering*, Vol. 3, Pp. 98–105, Jun. 2022, Doi: 10.1016/J.Ijce.2022.03.003.
- [19] I. K. Sastrawan, I. P. A. Bayupati, And D. M. S. Arsa, "Detection Of Fake News Using Deep Learning Cnn–Rnn Based Methods," *Ict Express*, Vol. 8, No. 3, 2022, Doi: 10.1016/J.Icte.2021.10.003.
- [20] T. Crippa Praha And M. Nugraheni, "Indonesian Fake News Classification Using Transfer Learning In Cnn And Lstm," Indonesia, Sep. 2024. Doi: <http://dx.doi.org/10.62527/Joiv.8.2.2126>.