IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR

Penerapan Bi-LSTM Untuk Named Entity Recognition Pada Teks Bahasa Indonesia

Akmal Hisyam Pradhana[1, a)], Erna Daniati[2, a)], Muhammad Najibulloh Muzaki[3, a)]

[a)] Sistem Informasi
Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri

Author Emails
[2)] Corresponding author: ernadaniati@unpkediri.ac.id
[1)]akmalhisyampradhana@gmail.com,
[3)] m.n.muzaki@gmail.com

Abstract. *This study aims to develop and evaluate a Named Entity Recognition (NER) model based on the Bidirectional Long Short-Term Memory (Bi-LSTM) architecture, capable of automatically identifying entities in Indonesian-language texts. The urgency of this research lies in the limited availability of effective NER systems for the Indonesian language, particularly for non-formal texts that exhibit unique structures and vocabulary. The main issue addressed is the low accuracy of entity extraction due to the limitations of previous NER models in understanding the complex and non-standard nature of the Indonesian language. The dataset used in this study was compiled from annotated Indonesian text corpora in the BIO (Beginning-Inside-Outside) format and categorized into several entity types, including Person, Location, Organization, Quantity, and Time. The process involved preprocessing (tokenization, BIO tagging, and padding), constructing the Bi-LSTM architecture, training the model using an 80:20 train-test split technique, and evaluating its performance using Precision, Recall, F1-Score, and confusion matrix metrics. The results indicate that the Bi-LSTM model achieved an overall accuracy of 99% and an F1-Score of 0.99, with the highest performance observed in recognizing ORGANIZATION and PERSON entities. This research contributes to the development of culturally adaptive NER systems and has the potential to be applied in education, cultural preservation, and context-aware information retrieval in the Indonesian language.*

Keywords **:**
BIO, BiLSTM, Indonesian Language, Named Entity Recognition, NLP.

Abstraksi. Penelitian ini bertujuan untuk membangun dan mengevaluasi model Named Entity Recognition (NER) berbasis arsitektur Bidirectional Long Short-Term Memory (Bi-LSTM) yang mampu mengenali entitas secara otomatis dalam teks berbahasa Indonesia. Urgensi penelitian ini terletak pada masih minimnya sistem NER yang efektif untuk bahasa Indonesia, terutama pada teks non-formal yang memiliki struktur dan kosakata unik. Permasalahan utama yang diangkat adalah rendahnya akurasi ekstraksi entitas akibat keterbatasan model-model NER sebelumnya dalam memahami konteks bahasa Indonesia yang kompleks dan tidak baku. Data dikumpulkan dari korpus teks Indonesia yang telah dianotasi format BIO (Beginning-Inside-Outside) dan diklasifikasikan dalam jenis entitas seperti Person, Location, Organization, Quantity, dan Time. Proses melibatkan preprocessing (tokenisasi, pelabelan BIO, dan padding), pembangunan arsitektur Bi-LSTM, pelatihan model teknik train-test split (80:20), serta evaluasi menggunakan metrik Precision, Recall, F1-Score, dan confusion matrix. Hasil penelitian menunjukkan model Bi-LSTM berhasil mencapai akurasi keseluruhan sebesar 99% dan F1-Score sebesar 0.99, dengan performa terbaik pada entitas ORGANIZATION dan PERSON. Penelitian ini berkontribusi pada pengembangan NER berbasis budaya lokal serta potensial diterapkan dalam pendidikan, pelestarian budaya, dan pencarian informasi kontekstual berbahasa Indonesia.

Kata Kunci : Bahasa Indonesia, BIO, BiLSTM, Named Entity Recognition, NLP.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR

PENDAHULUAN

Pendahuluan Perkembangan teknologi telah membawa dampak besar pada peradaban, seiring dengan bertumbuhnya kompleksitas interaksi budaya dalam Masyarakat [1]. Natural Language Processing (NLP) telah menjadi salah satu bidang penting dalam kemajuan teknologi masa kini berbasis kecerdasan buatan, khususnya dalam pemahaman teks berbahasa manusia [2]. Kemajuan teknologi yang pesat telah memungkinkan manusia untuk menyelesaikan pekerjaan dengan cara yang lebih efektif dan dalam waktu yang lebih singkat [3]. Salah satu cabang penting dalam NLP adalah Named Entity Recognition (NER), yaitu proses identifikasi dan klasifikasi entitas bernama seperti nama orang, lokasi, organisasi, dan lainnya dalam sebuah teks secara otomatis [4]. Penerapan NER telah banyak digunakan dalam sistem pencarian informasi, ekstraksi data otomatis, hingga sistem tanya jawab cerdas [5]. Pencarian informasi dari data berbasis teks merupakan elemen inti dalam NLP, yang memiliki peran krusial dalam mendeteksi serta mengelompokkan berbagai entitas secara otomatis [6]. Namun demikian, penelitian NER dalam konteks bahasa Indonesia masih terbatas.

Cerita indonesia merupakan bagian dari kekayaan budaya Indonesia yang sering digunakan dalam pendidikan, hiburan, pelestarian local hingga mengandung nilai-nilai moral [7]. Narasi-narasi tersebut diambil dari kumpulan data sastra bereputasi, arsip yang tersedia untuk umum, serta himpunan karya yang telah diterbitkan dalam bentuk antologi [8]. Teks bahasa indonesia memiliki karakteristik khusus, seperti gaya bahasa naratif, penggunaan istilah lokal, dan struktur kalimat yang tidak selalu mengikuti pola baku [9]. Hal ini menyebabkan pendekatan NER yang umum digunakan pada teks formal seperti berita atau dokumen resmi menjadi kurang efektif bila langsung diterapkan. Masalah yang muncul adalah belum tersedianya sistem NER yang optimal dan adaptif terhadap struktur bahasa Indonesia, sehingga proses ekstraksi informasi dari teks tersebut masih dilakukan secara manual dan tidak efisien. Penelitian ini hadir sebagai upaya untuk menjawab permasalahan tersebut dengan menerapkan pendekatan Bidirectional Long Short-Term Memory (BiLSTM) sebagai model NER.

Model BiLSTM dipilih karena kemampuannya dalam memahami konteks kata dari dua arah (kiri dan kanan) [10]. Kebaruan (novelty) dari penelitian ini terletak pada penggunaan metode NER berbasis BiLSTM untuk mencari entitas pada bahasa Indonesia yang sebelumnya belum banyak dieksplorasi. Selain itu, penelitian ini juga menyusun proses pelabelan otomatis menggunakan format BIO (Beginning, Inside, Outside) dengan entitas dalam seperti Person, Location, Organization, Quantity dan Time [11].

Dalam tinjauan pustaka, sejumlah penelitian sebelumnya telah menunjukkan efektivitas BiLSTM dalam tugas NER untuk teks formal seperti korpus berita dan data medis. Namun, hingga saat ini masih sangat sedikit kajian yang menguji model ini pada korpus dengan gaya bahasa Indonesia. Oleh karena itu, penelitian ini mencoba mengisi kekosongan tersebut sebagai kontribusi terhadap pengembangan NLP yang lebih kontekstual terhadap budaya lokal. Tujuan dari penelitian ini adalah membangun dan mengevaluasi model NER berbasis BiLSTM yang mampu mengenali entitas secara otomatis dan akurat dalam bahasa Indonesia. Melalui proses pelatihan dan evaluasi model, penelitian ini diharapkan dapat menghasilkan sistem yang adaptif terhadap bentuk bahasa naratif dan memberikan akurasi yang baik dalam klasifikasi entitas. Membuka peluang pemanfaatan dalam pengembangan cerita otomatis, media pembelajaran, serta bidang humaniora berbasis teknologi digital [12]. Hasil dari penelitian ini diharapkan dapat menjadi pijakan awal bagi pengembangan sistem ekstraksi informasi otomatis, serta mendukung pelestarian budaya melalui digitalisasi dan analisis teks bahasa indonesia. Selain itu, penerapan model NER dalam domain ini juga berpotensi diaplikasikan pada berbagai bidang seperti pendidikan, pengarsipan literatur daerah, dan pengembangan sistem pencarian informasi berbasis konteks budaya Indonesia.

TINJAUAN PUSTAKA

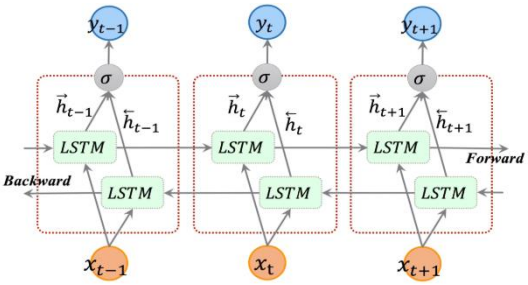
Penelitian mengenai NER pada teks berbahasa Indonesia telah mengalami perkembangan, khususnya dengan pemanfaatan arsitektur LSTM dan turunannya. Hidayatullah mengembangkan model Bi-LSTM untuk mengenali entitas dalam ulasan wisata berbahasa Indonesia dan berhasil mencapai F1-Score sebesar 94,3% [13]. Keunggulan dari penelitian ini terletak pada kemampuannya memproses teks non-formal tanpa memerlukan fitur linguistik tambahan, meskipun keterbatasan domain (ulasan wisata) membatasi generalisasinya ke jenis teks naratif. Pada tahun 2021 Santoso turut menerapkan model Bi-LSTM end-to-end untuk mengekstrak entitas dalam teks berita dan meraih F1-Score 83,18%, namun tanpa mempertimbangkan penyeimbangan data antar kelas entitas [14]. Sementara itu, Mustofa pada tahun 2024 memperluas cakupan entitas dalam domain informasi bencana dan berhasil menangani

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR

masalah distribusi data tidak seimbang melalui teknik random oversampling, yang menghasilkan F1-Score sebesar 87,5% [15]. Penelitian ini menjadi referensi penting terkait pentingnya penanganan data tidak seimbang pada entitas minoritas. Widhiyasana pada tahun 2021 menggabungkan arsitektur CNN dan LSTM (Conv-LSTM) dalam klasifikasi berita dan menemukan bahwa kombinasi ini efektif dalam menangkap fitur penting dari teks, meskipun belum secara eksplisit digunakan untuk NER [16]. Selain pendekatan berbasis LSTM, Koto di tahun 2021 mengembangkan IndoBERTweet, sebuah model pralatih berbasis transformer yang dirancang khusus untuk teks Twitter berbahasa Indonesia [17]. Meskipun tidak menggunakan LSTM, penelitian ini menunjukkan bahwa pretrained language model dengan domain khusus dapat meningkatkan performa NER secara signifikan pada teks informal. Berdasarkan studi-studi tersebut, dapat disimpulkan bahwa penerapan LSTM dalam NER untuk bahasa Indonesia menunjukkan performa tinggi terutama pada teks informal dan domain khusus. Namun, belum ditemukan studi yang secara eksplisit menerapkan pendekatan LSTM pada yang memiliki gaya bahasa Indonesia naratif khas. Kekosongan ini menunjukkan bahwa penelitian yang dilakukan pada teks Indonesia berpotensi menyempurnakan kajian sebelumnya dengan memperluas domain aplikasi NER berbasis LSTM. Selain itu, hasil-hasil terdahulu juga menegaskan pentingnya penanganan distribusi data yang tidak seimbang dan optimalisasi hyperparameter, yang keduanya menjadi aspek penting dalam pengembangan model yang andal.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun dan mengevaluasi model NER berbasis algoritma BiLSTM. Dalam arsitektur Bidirectional Long Short Term Memory (BiLSTM), terdapat dua buah jaringan LSTM yang bekerja secara berlawanan arah. Jaringan pertama bertugas memproses data masukan secara berurutan dari arah depan (forward), sedangkan jaringan kedua memproses data dari arah belakang (backward). Hasil keluaran dari kedua jaringan tersebut kemudian digabungkan pada setiap langkah waktu. Pendekatan dua arah ini memungkinkan model untuk menangkap konteks dari informasi sebelumnya maupun yang akan datang dalam setiap urutan input. Struktur BiLSTM yang digunakan mengacu pada referensi [15] dan divisualisasikan pada Gambar 1.



GAMBAR 1. Struktur BiLSTM

Dalam struktur jaringan BiLSTM, pemrosesan oleh LSTM ke arah depan (*forward*) dinyatakan dengan persamaan $\overrightarrow{h_t} = LSTM(x_t, h_{t-1})$. Sementara itu, untuk arah sebaliknya (*backward*), prosesnya dilambangkan dengan $\overleftarrow{h_t} = LSTM(x_t, h_{t+1})$. Hasil akhir yang diperoleh dari model BiLSTM merupakan hasil penggabungan kedua arah tersebut, yang diformulasikan sebagai $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$. Sehingga Tujuan utama dari penelitian ini adalah untuk mengembangkan sistem Bi-LSTM secara otomatis yang mampu mengenali kategori entitas seperti pada Tabel 1.

TABEL 1. Jenis Entitas

Label	Jenis Entitas	Penjelasan
ORGANIZATION	Organisasi	Nama organisasi, seperti Perusahaan/ institusi.
LOCATION	Lokasi	Nama tempat atau lokasi geografis seperti kota, negara, atau gedung.
PERSON	Nama Orang	Nama orang.
TIME	Waktu	Waktu seperti tanggal, hari, atau jam.
QUANTITY	Kuantitas	Informasi berupa jumlah atau satuan.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR

Dataset yang digunakan dalam penelitian ini disimpan dalam format pickle dan berisi pasangan data antara teks dan anotasi entitas. Setiap data terdiri dari sebuah kalimat dan anotasi yang menunjukkan entitas-entitas yang terdapat di dalam kalimat tersebut. Pada Gambar 2. Anotasi ditulis dalam bentuk tuple yang terdiri dari tiga elemen, yaitu posisi karakter awal, posisi karakter akhir, dan label jenis entitas, seperti PERSON atau ORGANIZATION.

```
(
    'Pengamat politik dari Universitas Gadjah Mada Arie Sudjito '
    'menilai, keinginan Ketua Umum Partai Golkar Aburizal Bakrie '
    'untuk maju kembali sebagai ketua umum merupakan pemaksaan kehendak.',
    {
        'entities': [
            (22, 45, 'ORGANIZATION'),
            (46, 58, 'PERSON'),
            (89, 102, 'ORGANIZATION'),
            (103, 118, 'PERSON')
        ]
    }
)
```

GAMBAR 2. Struktur Dataset Named Entity Recognition

Data Preprocessing

Tahapan Data yang digunakan berupa kumpulan teks beranotasi yang disimpan dalam format pickle. Setiap sampel terdiri dari teks dan anotasi entitas yang mencakup label serta posisi karakter. Proses preprocessing dilakukan untuk memastikan bahwa data sesuai untuk dimasukkan ke dalam model pembelajaran mesin. Tahapan preprocessing meliputi :

a). Tokenisasi: Memecah teks menjadi token kata untuk memudahkan pemetaan ke dalam bentuk vektor. Proses ini penting agar model dapat mengenali struktur dan konteks kalimat. Tokenisasi telah terbukti sebagai langkah awal krusial dalam tugas pemrosesan bahasa alami seperti Named Entity Recognition (NER) [18].

b). Labelisasi BIO: Setiap token diberi label menggunakan skema BIO (Beginning, Inside, Outside) berdasarkan keberadaan entitas dalam teks. Skema BIO memungkinkan model untuk membedakan awal dan bagian dalam suatu entitas, sehingga meningkatkan akurasi identifikasi entitas [19].

c). Mapping ke Indeks Numerik: Token dan label diubah menjadi indeks numerik menggunakan kamus word2idx dan tag2idx, Sehingga hasilnya Seperti pada Gambar 3.

```
Contoh Mapping Kata ke Indeks (word2idx):
1000: 1
separatis: 2
produk: 3
kacamata: 4
Dede: 5
bor: 6
invasif: 7
bertahan: 8
tujuan-tujuan: 9
digitalnya: 10

Mapping Tag ke Indeks (tag2idx):
I-ORGANIZATION: 0
B-LOCATION: 1
I-TIME: 2
B-ORGANIZATION: 3
O: 4
I-QUANTITY: 5
B-PERSON: 6
I-LOCATION: 7
B-QUANTITY: 8
I-PERSON: 9
B-TIME: 10
```

GAMBAR 3. Mapping Kata Ke Indeks

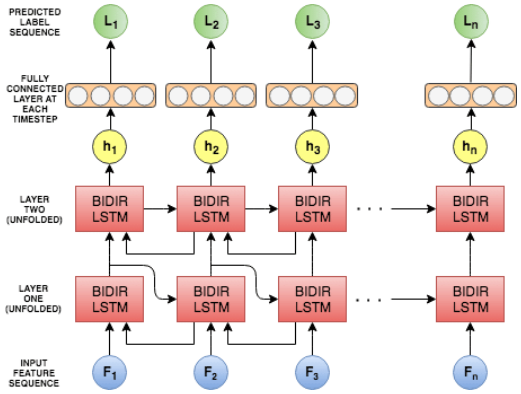
Representasi numerik ini diperlukan agar data dapat diproses oleh model neural network. Proses ini sejalan dengan praktik standar dalam pelatihan model deep learning [20].

d). Padding: Semua urutan token dipanjangkan atau dipotong agar memiliki panjang yang seragam (max_len). Padding penting untuk memfasilitasi pelatihan dalam batch tetap, terutama pada model berbasis RNN atau Transformer. Penyesuaian panjang input telah menjadi praktik umum untuk menghindari kesalahan dimensi pada pemrosesan batch [21].

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR

Arsitektur Model

Model yang digunakan dalam penelitian ini adalah jaringan neural Bi-LSTM, yang dirancang khusus untuk menangani permasalahan sequence labeling seperti Named Entity Recognition (NER).



GAMBAR 4. Arsitektur Bi-LSTM Untuk NER

Arsitektur model terdiri dari beberapa lapisan seperti:

a. Embedding Layer

Lapisan embedding berfungsi untuk merepresentasikan kata ke dalam bentuk vektor berdimensi 50 yang memiliki makna semantik. Representasi ini memungkinkan model menangkap hubungan antar kata dalam ruang vektor. Embedding word-level masih menjadi pendekatan yang relevan dan digunakan secara luas dalam implementasi model NER modern [22].

b. Bidirectional LSTM Layer

Lapisan BiLSTM dengan 100 unit neuron digunakan untuk menangkap konteks dari dua arah (maju dan mundur) dalam satu urutan teks. BiLSTM memberikan keuntungan karena memahami makna kata tidak hanya berdasarkan konteks sebelumnya tetapi juga sesudahnya, yang sangat penting dalam mendeteksi entitas dengan benar. Studi terbaru menunjukkan bahwa BiLSTM tetap kompetitif dalam berbagai tugas sekuensial, termasuk NER, terutama dalam pengaturan low-resource [23].

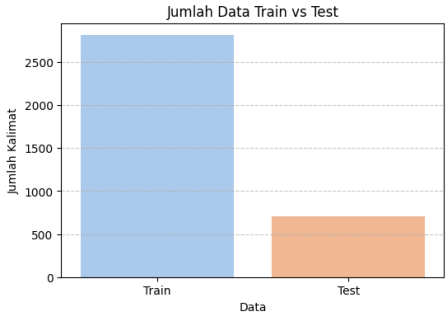
c. Time Distributed Dense Layer

Lapisan ini digunakan untuk menerapkan layer Dense secara independen pada tiap waktu/token. Tujuannya adalah memetakan keluaran dari LSTM ke prediksi label tag untuk setiap token secara paralel. Struktur seperti ini umum digunakan dalam implementasi NER modern dan memungkinkan efisiensi pemrosesan dalam model deep learning [24]. Model ini diimplementasikan menggunakan TensorFlow/Keras, dengan fungsi aktivasi softmax pada lapisan output untuk menghasilkan distribusi probabilitas atas kelas label. Digunakan juga fungsi categorical_crossentropy sebagai loss function, sesuai dengan pendekatan klasifikasi multi-kelas dalam model sequence labeling. Pemilihan fungsi ini mengikuti praktik umum dalam pelatihan model neural NER modern [25].

Pelatihan Model

Data Data yang digunakan dalam penelitian ini dibagi menjadi dua bagian utama, yaitu data latih (training data) dan data uji (testing data), guna memfasilitasi proses pembelajaran mesin secara optimal. Proses pembagian data dilakukan dengan menggunakan teknik train-test split, yang merupakan metode umum dalam pembelajaran mesin untuk mengevaluasi kinerja model. Dalam implementasinya pada Gambar 5. rasio pembagian data yang digunakan adalah 80:20. Artinya, sebanyak 80% dari total data digunakan sebagai data latih untuk melatih model agar dapat mengenali pola dan karakteristik dari dataset, sementara sisanya sebesar 20% digunakan sebagai data uji untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR



GAMBAR 5. Pembagian Data Latih Dan Uji

Model dilatih selama 5 epoch, yaitu lima kali proses pelatihan penuh terhadap seluruh data latih. Jumlah epoch ini dipilih untuk memberikan kesempatan kepada model agar dapat mempelajari pola-pola penting dalam data secara bertahap tanpa mengalami overfitting. Selama proses pelatihan, digunakan ukuran batch sebesar 1, yang berarti bahwa pembaruan bobot dilakukan setelah setiap sampel data diproses. Ukuran batch kecil seperti ini memungkinkan model untuk menyesuaikan bobotnya secara lebih cepat terhadap setiap perubahan kecil dalam data, meskipun cenderung menghasilkan pembelajaran yang lebih fluktuatif.

```
history = model.fit(X_train, y_train, batch_size=1,
                    epochs=5, validation_data=(X_test, y_test), verbose=1)
```

GAMBAR 6. Pelatihan model menggunakan epoch sebanyak 5 kali

Untuk mengoptimalkan proses pembelajaran, digunakan optimizer Adam (Adaptive Moment Estimation), yang dikenal efektif dalam mempercepat konvergensi dan stabil dalam berbagai jenis masalah pembelajaran mesin. Evaluasi kinerja model dilakukan dengan menggunakan metrik akurasi, yaitu persentase jumlah prediksi yang benar dibandingkan dengan total jumlah prediksi. Metrik ini dipilih karena sederhana dan cukup representatif untuk kasus klasifikasi yang digunakan dalam penelitian ini.

Evaluasi Kinerja

Evaluasi kinerja model dilakukan menggunakan beberapa metrik dan visualisasi yang umum digunakan dalam tugas NER. Pendekatan ini bertujuan untuk memahami kemampuan model dalam mengidentifikasi entitas secara akurat serta mengamati kesalahan klasifikasi antar label. Tahap evaluasi menjadi langkah krusial untuk menilai keakuratan sekaligus menjamin validitas temuan yang dihasilkan dalam proses identifikasi entitas [26]. Evaluasi dilakukan dengan menghitung metrik Precision, Recall, dan F1-Score untuk setiap label entitas [27]. Accuracy menunjukkan proporsi prediksi yang benar dari seluruh prediksi yang dilakukan, dihitung dengan membagi jumlah prediksi benar dengan total data uji [28]. Precision mengukur ketepatan model dalam memprediksi kelas positif, yakni rasio antara prediksi positif yang benar dengan seluruh prediksi positif. Recall menilai kemampuan model dalam menemukan semua data positif yang sebenarnya, yaitu jumlah prediksi positif yang benar dibanding total data positif aktual. Recall mengindikasikan sejauh mana model mampu mengenali seluruh kemunculan suatu kelas secara menyeluruh dalam dataset [29]. F1-score merupakan rata-rata harmonis dari precision dan recall, berguna untuk menyeimbangkan keduanya, terutama saat dibutuhkan kompromi antara ketepatan dan kelengkapan.

$$\text{Precision} = \frac{TP}{TP + FP} \quad 1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad 2)$$

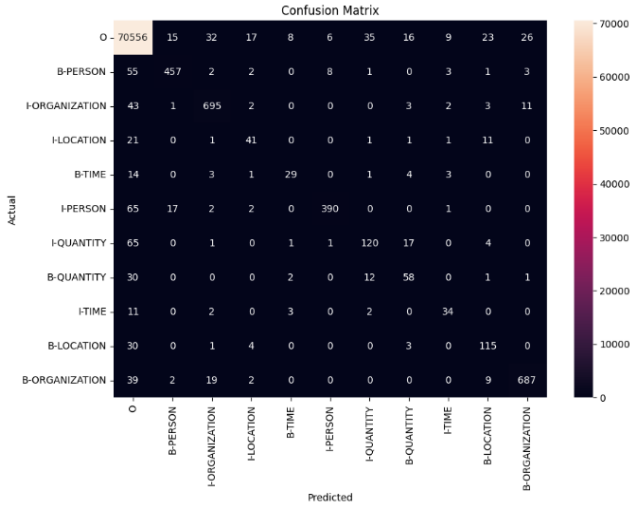
IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad 3)$$

Metrik ini memberikan gambaran menyeluruh mengenai performa model dalam mengenali entitas, khususnya untuk label-label yang tidak seimbang secara frekuensi. Studi oleh Li [30] menekankan pentingnya F1-score dalam pengukuran kinerja NER karena menyeimbangkan precision dan recall dalam satu nilai agregat. Lalu Confusion matrix digunakan untuk menganalisis kesalahan klasifikasi antar label entitas. Matriks ini memungkinkan identifikasi spesifik terhadap kelas yang paling sering tertukar, serta memberikan informasi visual tentang distribusi prediksi dan kesalahan model. Dalam penelitian oleh Rajbhandari et al. [31], visualisasi confusion matrix terbukti efektif untuk diagnosis model dalam konteks NLP, termasuk klasifikasi teks dan ekstraksi informasi.

HASIL DAN PEMBAHASAN

Model Bi-LSTM yang diterapkan dalam penelitian ini diuji menggunakan dataset bahasa indonesia untuk tugas NER. Evaluasi performa model dilakukan menggunakan Confusion Matrix yang menggambarkan hubungan antara label aktual dan label yang diprediksi oleh model. Gambar 7. menunjukkan confusion matrix hasil prediksi terhadap 11 label entitas, yaitu: O, B-PERSON, I-ORGANIZATION, I-LOCATION, B-TIME, I-PERSON, I-QUANTITY, B-QUANTITY, I-TIME, B-LOCATION, B-ORGANIZATION. Dari Confusion Matrix, label O (bukan entitas) memiliki jumlah prediksi terbanyak, yaitu sebesar 70.556, yang menunjukkan dominasi token non-entitas dalam korpus. Prediksi model terhadap label O cukup akurat, meskipun masih terjadi sejumlah false positive dan false negative terhadap kelas entitas seperti B-PERSON dan I-ORGANIZATION. Untuk entitas bernama (named entity), performa model bervariasi. Kategori I-ORGANIZATION dan B-ORGANIZATION menunjukkan performa yang sangat baik dengan jumlah prediksi benar masing-masing sebanyak 695 dan 687, jauh lebih tinggi dibanding kesalahan prediksinya. Ini menunjukkan bahwa model Bi-LSTM berhasil mengenali pola-pola spesifik untuk organisasi baik dalam posisi awal (B-) maupun lanjutan (I-). Sementara itu, prediksi terhadap entitas B-PERSON dan I-PERSON cukup akurat dengan hasil benar masing-masing 457 dan 390, namun tetap terdapat confusion yang signifikan dengan kelas O dan sesama entitas, misalnya I-PERSON terkadang diprediksi sebagai O (65 kali) atau B-PERSON (17 kali). Hal ini menunjukkan bahwa model mengalami kesulitan dalam membedakan transisi antar posisi entitas. Entitas yang lebih jarang muncul seperti I-QUANTITY, B-QUANTITY, B-TIME, dan I-TIME memiliki akurasi yang relatif rendah, yang diduga karena ketidakseimbangan jumlah data pelatihan. Contohnya, I-QUANTITY hanya diprediksi benar sebanyak 120 kali, tetapi keliru diklasifikasikan sebagai O sebanyak 65 kali. Demikian pula B-QUANTITY memiliki 58 prediksi benar tetapi 30 token salah diklasifikasikan sebagai O.



GAMBAR 7. Hasil Confusion Matrix 11 Entitas

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 4, No. 2 Juli 2025
https://subset.id/index.php/IJCSR

Model Bi-LSTM efektif dalam mengenali entitas utama seperti ORGANIZATION dan PERSON, terutama jika didukung oleh jumlah sampel yang mencukupi. Kesalahan prediksi terbesar terjadi antara kelas O dan entitas, yang mengindikasikan bahwa model kadang tidak yakin apakah sebuah token merupakan bagian dari entitas atau bukan. Entitas minor yang jarang muncul dalam data pelatihan, seperti TIME dan QUANTITY, memiliki akurasi yang lebih rendah, yang mengarah pada perlunya teknik penyeimbangan data seperti oversampling atau augmentasi data. Transisi antar tag posisi (B- ke I-) masih menjadi tantangan, misalnya kesalahan prediksi dari I-ORGANIZATION ke O atau B-ORGANIZATION menunjukkan bahwa model belum sepenuhnya memahami struktur sekuensial BIO tagging. Hasil ini memperkuat temuan dari penelitian sebelumnya bahwa LSTM sangat efektif untuk tugas-tugas sekuensial seperti NER, namun tetap memiliki keterbatasan dalam menangani entitas jarang dan transisi posisi antar entitas. Model NER yang dikembangkan menggunakan arsitektur Bi-LSTM juga telah dievaluasi menggunakan metrik Precision, Recall, dan F1-Score. Evaluasi dilakukan terhadap 13 label entitas termasuk label "O" (non-entitas). Gambar 8. menunjukkan hasil evaluasi performa model terhadap data uji:

	precision	recall	f1-score	support
O	0.99	1.00	1.00	70743
B-PERSON	0.93	0.86	0.89	532
I-ORGANIZATION	0.92	0.91	0.92	760
I-LOCATION	0.58	0.53	0.55	77
B-TIME	0.67	0.53	0.59	55
I-PERSON	0.96	0.82	0.88	477
I-QUANTITY	0.70	0.57	0.63	209
B-QUANTITY	0.57	0.56	0.56	104
I-TIME	0.64	0.65	0.65	52
B-LOCATION	0.69	0.75	0.72	153
B-ORGANIZATION	0.94	0.91	0.92	758
accuracy			0.99	73920
macro avg	0.78	0.74	0.76	73920
weighted avg	0.99	0.99	0.99	73920

GAMBAR 8. Hasil Classification Report 11 Entitas

Secara keseluruhan, model mencapai akurasi sebesar 99% dengan rata-rata F1-Score berbobot (Weighted Avg) sebesar 0.99, yang menunjukkan bahwa model memiliki kinerja yang sangat baik dalam mengidentifikasi entitas secara keseluruhan, terutama pada kelas mayoritas. Hasil evaluasi menunjukkan bahwa model Bi-LSTM sangat baik dalam mengenali token dengan label "O", yang mendominasi data dengan dukungan (support) sebanyak 70.743. Precision dan Recall pada label ini sangat tinggi, yakni 0.99 dan 1.00, sehingga menghasilkan F1-Score sempurna (1.00). Namun, performa tinggi ini tidak serta-merta menunjukkan kemampuan model dalam mengenali entitas karena ketidakseimbangan kelas yang cukup signifikan. Beberapa entitas utama seperti B-PERSON, I-PERSON, B-ORGANIZATION, dan I-ORGANIZATION menunjukkan performa yang sangat baik dengan f1-score > 0.85. Hal ini mengindikasikan bahwa model mampu mengidentifikasi nama orang dan organisasi dengan akurasi tinggi. Hal ini dapat dikaitkan dengan frekuensi kemunculan label-label tersebut dalam data pelatihan serta konsistensi pola leksikal dari nama entitas tersebut.

Sebaliknya, entitas seperti I-LOCATION, B-TIME, dan B-QUANTITY menunjukkan performa yang relatif rendah dengan F1-Score di bawah 0.60. Hal ini dapat disebabkan oleh beberapa factor yaitu Jumlah data pelatihan yang terbatas untuk label tersebut (sebagaimana terlihat dari nilai support), Konteks linguistik yang bervariasi dan tidak konsisten, Kemungkinan ambiguitas leksikal, terutama pada entitas kuantitas dan waktu, yang bisa menyerupai token non-entitas. Nilai Macro Average F1-Score sebesar 0.76 mencerminkan performa rata-rata model terhadap semua kelas tanpa memperhatikan proporsi jumlah token. Ini menunjukkan bahwa meskipun akurasi keseluruhan tinggi, model masih memiliki kelemahan dalam mengenali beberapa entitas minor, terutama yang underrepresented.

KESIMPULAN

Penelitian berhasil membangun model Named Entity Recognition (NER) berbasis arsitektur Long Short-Term Memory (LSTM) yang mampu mengenali entitas dalam teks berbahasa Indonesia dengan akurasi tinggi. Model menunjukkan kinerja yang sangat baik dalam mengidentifikasi entitas mayoritas seperti ORGANIZATION dan PERSON, terutama pada kelas dengan jumlah data pelatihan yang mencukupi. Keunggulan utama model terletak pada

kemampuannya memahami konteks sekuensial dan menangani gaya bahasa naratif. Namun demikian, model masih memiliki kelemahan dalam mengenali entitas minor seperti TIME dan QUANTITY, yang disebabkan oleh ketidakseimbangan data dan kompleksitas konteks linguistik. Transisi antar tag posisi entitas juga masih menjadi tantangan, sebagaimana terlihat dari kesalahan prediksi pada kelas I- dan B-. Dengan demikian, penelitian ini memberikan jawaban atas permasalahan kurang optimalnya sistem NER dalam domain teks non-formal, serta membuktikan bahwa pendekatan LSTM dapat memberikan performa tinggi dengan catatan perlunya penanganan terhadap distribusi data yang tidak seimbang. Penelitian lanjutan disarankan untuk menerapkan teknik penyeimbangan data seperti oversampling atau augmentasi data untuk meningkatkan akurasi pada entitas yang jarang muncul. Selain itu, eksplorasi arsitektur lanjutan seperti BiLSTM-CRF atau penggunaan pre-trained language models seperti BERT yang disesuaikan untuk bahasa Indonesia dapat menjadi arah pengembangan berikutnya. Terakhir, penelitian ke depan juga disarankan pula untuk memperluas cakupan dataset agar mencakup lebih banyak variasi bahasa dari berbagai daerah untuk meningkatkan generalisasi model.

DAFTAR PUSTAKA

[1] Sucipto *et al.*, "Hidden Treasures of Kediri's Medicinal Plants: A Collaborative Effort to Map and Validate Authentic Information Using Innovative QR Code Security and Cryptography," *IOP Conf Ser Earth Environ Sci*, vol. 1242, no. 1, p. 012036, Sep. 2023, doi: 10.1088/1755-1315/1242/1/012036.

[2] A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavaf, and E. A. Fox, "Natural Language Processing Advancements By Deep Learning: A Survey," *CoRR*, vol. abs/2003.01200, 2020, [Online]. Available: https://arxiv.org/abs/2003.01200

[3] Sucipto, A. Ristyawan, D. Harini, W. I. Zaman, M. N. Muzaki, and M. N. A. Abdulnazar, "Integrating Cryptographic Security Features in Information System Barcodes for Self-Service Systems," *Advance Sustainable Science Engineering and Technology*, vol. 6, no. 4, p. 02404012, Sep. 2024, doi: 10.26877/asset.v6i4.850.

[4] B. Jehangir, S. Radhakrishnan, and R. Agarwal, "A survey on Named Entity Recognition — datasets, tools, and methodologies," *Natural Language Processing Journal*, vol. 3, p. 100017, Jun. 2023, doi: 10.1016/j.nlp.2023.100017.

[5] D. Moldovan and M. Surdeanu, "On the Role of Information Retrieval and Information Extraction in Question Answering Systems," 2003, pp. 129–147. doi: 10.1007/978-3-540-45092-4_6.

[6] E. Daniati, A. P. Wibawa, and W. S. G. Irianto, "Event Extraction in Narrative Texts: A Zero-Shot Approach Using Bert and Bi-LSTM on Andersen's Fairy Tales," in *2025 17th International Conference on Knowledge and Smart Technology (KST)*, IEEE, Feb. 2025, pp. 208–213. doi: 10.1109/KST65016.2025.11003342.

[7] E. Daniati, A. P. Wibawa, and W. S. G. Irianto, "The Relevance of Andersen's Children's Stories to the Growth and Development of Indonesian Children," 2025, pp. 108–124. doi: 10.2991/978-2-38476-352-8_10.

[8] E. Daniati, A. P. Wibawa, W. S. G. Irianto, A. Ghosh, and L. Hernandez, "Analyzing event relationships in Andersen's Fairy Tales with BERT and Graph Convolutional Network (GCN)," *Science in Information Technology Letters*, vol. 5, no. 1, pp. 40–60, May 2024, doi: 10.31763/sitech.v5i1.1810.

[9] N. Novianti, "Indonesian Folk Narratives: On the Interstices of National Identity, National Values, and Character Education," *Journal of Ethnology and Folkloristics*, vol. 16, no. 1, pp. 99–116, Jun. 2022, doi: 10.2478/jef-2022-0006.

[10] A. Pogiatzis and G. Samakovitis, "Using BiLSTM Networks for Context-Aware Deep Sensitivity Labelling on Conversational Data," *Applied Sciences*, vol. 10, no. 24, p. 8924, Dec. 2020, doi: 10.3390/app10248924.

[11] T. Kato, K. Abe, H. Ouchi, S. Miyawaki, J. Suzuki, and K. Inui, "Embeddings of Label Components for Sequence Labeling: A Case Study of Fine-grained Named Entity Recognition," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 222–229. doi: 10.18653/v1/2020.acl-srw.30.

[12] E. Daniati, A. P. Wibawa, and W. S. G. Irianto, "Extracting Narrative Events in Andersen's Fairy Tales Using a Hybrid BERT-LSTM Model," 2025, pp. 355–372. doi: 10.1007/978-981-96-4613-5_26.

[13] A. F. Hidayatullah, M. F. D. A. Putra, A. P. Wibowo, and K. R. Nastiti, "Named Entity Recognition on Tourist Destinations Reviews in the Indonesian Language," *Jurnal Linguistik Komputasional (JLK)*, vol. 6, no. 1, pp. 30–35, Apr. 2023, doi: 10.26418/jlk.v6i1.89.

[14] J. Santoso, E. I. Setiawan, C. N. Purwanto, E. M. Yuniarno, M. Hariadi, and M. H. Purnomo, "Named entity recognition for extracting concept in ontology building on Indonesian language using end-to-end bidirectional long short term memory," *Expert Syst Appl*, vol. 176, p. 114856, Aug. 2021, doi: 10.1016/j.eswa.2021.114856.

[15] G. F. Shidik *et al.*, "Indonesian disaster named entity recognition from multi source information using bidirectional LSTM (BiLSTM)," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 3, p. 100358, Sep. 2024, doi: 10.1016/j.joitmc.2024.100358.

[16] Yudi Widhiyasana, Transmissia Semiawan, Ilham Gibran Achmad Mudzakir, and Muhammad Randi Noor, "Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 4, pp. 354–361, Nov. 2021, doi: 10.22146/jnteti.v10i4.2438.

[17] F. Koto, J. H. Lau, and T. Baldwin, "IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 10660–10668. doi: 10.18653/v1/2021.emnlp-main.833.

[18] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural Architectures for Named Entity Recognition," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 260–270. doi: 10.18653/v1/N16-1030.

[19] L. Ratinov and D. Roth, "Design challenges and misconceptions in named entity recognition," in *Proceedings of the Thirteenth Conference on Computational Natural Language Learning - CoNLL '09*, Morristown, NJ, USA: Association for Computational Linguistics, 2009, p. 147. doi: 10.3115/1596374.1596399.

[20] X. Ma and E. Hovy, "End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 1064–1074. doi: 10.18653/v1/P16-1101.

[21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[22] R. Thirukovalluru, N. Monath, K. Shridhar, M. Zaheer, M. Sachan, and A. McCallum, "Scaling Within Document Coreference to Long Texts," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 3921–3931. doi: 10.18653/v1/2021.findings-acl.343.

[23] C. van Niekerk *et al.*, "Uncertainty Measures in Neural Belief Tracking and the Effects on Dialogue Policy Performance," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 7901–7914. doi: 10.18653/v1/2021.emnlp-main.623.

[24] J. Li, A. Sun, J. Han, and C. Li, "A Survey on Deep Learning for Named Entity Recognition," *IEEE Trans Knowl Data Eng*, vol. 34, no. 1, pp. 50–70, Jan. 2022, doi: 10.1109/TKDE.2020.2981314.

[25] Z. Sha *et al.*, "Low I/O Intensity-aware Partial GC Scheduling to Reduce Long-tail Latency in SSDs," *ACM Transactions on Architecture and Code Optimization*, vol. 18, no. 4, pp. 1–25, Dec. 2021, doi: 10.1145/3460433.

[26] E. Daniati, R. Firliana, A. S. Wardani, and A. C. Zarkasi, "Evaluation Framework for Decision Making Based On Sentiment Analysis in Social Media," in *2021 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, IEEE, Dec. 2021, pp. 47–51. doi: 10.1109/ICAMIMIA54022.2021.9807790.

[27] E. Daniati and H. Utama, "ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER," *Journal of Information System Management (JOISM)*, vol. 4, no. 2, pp. 125–131, Jan. 2023, doi: 10.24076/joism.2023v4i2.973.

[28] M. Ikhsan, "Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Long Short-Term Memory," *The Indonesian Journal of Computer Science Research*, vol. 2, no. 1, pp. 31–41, Feb. 2023, doi: 10.59095/ijcsr.v2i1.29.

[29] H. Utama, E. Daniati, and A. Masruro, "Weak Supervision Dengan Pendekatan Labeling Function Untuk Analisis Sentimen Pada Twitter," *The Indonesian Journal of Computer Science Research*, vol. 3, no. 1, pp. 49–57, Jan. 2024, doi: 10.59095/ijcsr.v3i1.93.

[30] J. Li, A. Sun, J. Han, and C. Li, "A Survey on Deep Learning for Named Entity Recognition," *IEEE Trans Knowl Data Eng*, vol. 34, no. 1, pp. 50–70, Jan. 2022, doi: 10.1109/TKDE.2020.2981314.

[31] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Federated Query processing for Big Data in Data Science," in *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2019, pp. 6145–6147. doi: 10.1109/BigData47090.2019.9005530.