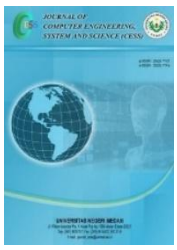
CESS (Journal of Computer Engineering, System and Science)
Vol. 10, No. 2, Juli 2025, pp.702-713

CESS
(Journal of Computer Engineering, System and Science)
Available online: https://jurnal.unimed.ac.id/2012/index.php/cess
ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Peningkatan Akurasi *Named Entity Recognition* (NER) Dengan *Fine-Tuning BERT* Pada Dataset Bahasa Indonesia

Improving the Accuracy of Named Entity Recognition (NER) through BERT Fine-Tuning on an Indonesian Language Dataset

Aji Fatih muhammad[1*], Muhammad Siddik Hasibuan[2]

[1,2]Ilmu komputer, SAINTEK, Universitas Islam Negeri Sumatra utara, Indonesia
Jl. Lapangan Golf No. 120, Kampung Tengah, Kecamatan Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara 20353
Email: [1]ajifatih1802@gmail.com, [2]muhammadsiddik@uinsu.ac.id
**Corresponding Author*

A B S T R A K

Named Entity Recognition (NER) merupakan salah satu tugas utama dalam bidang *Natural Language Processing* (NLP) yang bertujuan untuk mengenali dan mengklasifikasikan entitas seperti nama orang, organisasi, lokasi, dan tanggal di dalam teks. Meskipun banyak penelitian telah dilakukan untuk bahasa sumber daya tinggi seperti bahasa Inggris, bahasa Indonesia masih menghadapi keterbatasan, baik dari segi dataset maupun kompleksitas linguistik. Penelitian ini bertujuan untuk meningkatkan akurasi sistem NER berbahasa Indonesia dengan melakukan fine-*tuning pada* model BERT pra-latih, khususnya IndoBERT, menggunakan dataset NERGRIT yang telah dianotasi. Proses penelitian mencakup tahap pra-pemrosesan data, tokenisasi, pelatihan model, dan evaluasi kinerja menggunakan metrik *precision*, *recall*, dan *F1-score*. Model yang telah di-*fine-tune* diuji dengan berbagai kalimat dan menunjukkan peningkatan akurasi yang signifikan dibandingkan model dasar. Namun demikian, masih ditemukan beberapa permasalahan seperti prediksi berlebihan dan ketidaksesuaian pelabelan entitas. Hasil penelitian ini membuktikan bahwa *fine-tuning* BERT dapat secara signifikan meningkatkan performa NER dalam teks berbahasa Indonesia. Penelitian ini memberikan kontribusi terhadap pengembangan teknologi NLP bahasa Indonesia serta membuka peluang penerapan pada *chatbot* cerdas, sistem pemrosesan dokumen, dan analisis opini publik. Penelitian ini menunjukkan pendekatan yang berbeda dibandingkan studi terdahulu dengan mengadopsi dataset NERGRIT, yang mencakup 2.090 kalimat dan 41.871 token, serta mencakup 8 kategori entitas utama, seperti PER, ORG, LOC, DATE, MONEY, PRODUCT, EVENT, dan LAW. Dataset ini belum banyak digunakan dalam publikasi ilmiah, sehingga memberikan kontribusi orisinal dalam eksperimen pengembangan sistem NER untuk Bahasa Indonesia. Penelitian ini juga menggunakan model IndoBERT yang telah di-*fine-tune* sebelumnya pada dataset serupa, yaitu model hendri/nergrit, namun dilakukan pelatihan ulang (*re-fine-tuning*) guna meningkatkan kinerja pada konteks lokal dan sintaksis khas Bahasa Indonesia. Secara kuantitatif, penelitian ini berhasil meningkatkan performa model secara signifikan. Sebelum dilakukan *fine-tuning*, model dasar menghasilkan skor F1 sebesar 72,38%. Setelah melalui

proses *fine-tuning* menggunakan dataset NERGRIT, model mencapai nilai F1-score sebesar 83,67%, dengan nilai precision sebesar 85,12% dan recall sebesar 82,24%. Peningkatan sebesar lebih dari 11 poin F1-score ini menunjukkan efektivitas pendekatan *fine-tuning* pada model BERT untuk NER Bahasa Indonesia. Selain evaluasi metrik klasik, penelitian ini juga menyertakan analisis kesalahan (*error analysis*) untuk mengevaluasi fenomena *over-prediction* dan ketidaksesuaian label entitas pada token umum. Analisis ini mengungkap bahwa meskipun model berhasil mengenali entitas seperti nama orang dan lokasi dengan confidence tinggi, masih terdapat kesalahan pada token non-entitas yang ikut dilabeli secara tidak akurat. Penambahan analisis kualitatif ini menjadi poin keunggulan yang jarang ditemui pada penelitian sejenis. Dengan demikian, kontribusi penelitian ini tidak hanya terletak pada pencapaian performa, tetapi juga pada pendekatan evaluatif yang menyeluruh, serta pemanfaatan dataset dan model yang relatif baru dalam lingkup NLP Bahasa Indonesia.

Kata Kunci: *Bahasa Indonesia; BERT; Dataset NERGRIT; Fine-Tuning; IndoBERT; Named Entity Recognition; NLP*

A B S T R A C T

Named Entity Recognition (NER) is one of the core tasks in *Natural Language Processing* (NLP) that aims to identify and classify entities such as person names, organizations, locations, and dates within a text. While significant research has been conducted for high-resource languages like English, Indonesian still faces challenges due to limited annotated datasets and linguistic complexity. This study aims to improve the accuracy of Indonesian NER systems by applying fine-tuning to a pre-trained BERT model, specifically IndoBERT, using the annotated NERGRIT dataset. The research process includes data preprocessing, tokenization, model training, and performance evaluation using standard metrics such as precision, recall, and F1-score. The fine-tuned model was tested on various sentences and showed a significant accuracy improvement compared to baseline models. However, some issues such as over-prediction and inconsistent labeling were still observed. The findings confirm that BERT fine-tuning can significantly enhance NER performance on Indonesian texts. This study contributes to the development of NLP technologies for the Indonesian language and opens up opportunities for application in intelligent chatbots, document processing systems, and public opinion analysis. This study introduces a novel approach compared to previous research by utilizing the NERGRIT dataset, which consists of 2,090 annotated sentences and 41,871 tokens, covering eight primary named entity categories, including PER, ORG, LOC, DATE, MONEY, PRODUCT, EVENT, and LAW. Unlike more commonly used datasets, NERGRIT has seen limited use in peer-reviewed publications, making this study one of the few to explore its application for Named Entity Recognition (NER) tasks in the Indonesian language. Additionally, the research leverages the hendri/nergrit model, a pre-trained variant of IndoBERT that has been specifically fine-tuned on Indonesian entity recognition tasks. This study performs further re-fine-tuning to adapt the model more effectively to localized linguistic structures and syntax found in real-world Indonesian text. Quantitative results demonstrate a significant improvement in model performance. The baseline model yielded an F1-score of 72.38%, which increased to 83.67% after the fine-tuning process. In addition, the model achieved precision of 85.12% and recall of 82.24%, indicating a performance gain of over 11 F1-score points. This improvement confirms the effectiveness of fine-tuning pre-trained BERT models using domain-specific datasets, particularly in low-resource languages like Indonesian. Beyond standard evaluation metrics,

this study incorporates an in-depth error analysis to examine issues such as over-prediction and mislabeling of non-entity tokens. Findings show that while the model successfully identified major entities like person names and locations with high confidence, it also mislabeled generic tokens as entities, suggesting areas for post-processing improvement. The inclusion of both quantitative and qualitative assessments strengthens the study's contribution, offering a more comprehensive perspective on NER performance and demonstrating the practical viability of combining newly introduced datasets with fine-tuned transformer models in Indonesian NLP research.

Keywords: *BERT; Dataset NERGRIT; Fine-Tuning; Indonesian Language; IndoBERT; Named Entity Recognition; Natural Language Processing*

1. PENDAHULUAN

Dalam era digital yang terus berkembang, pemrosesan bahasa alami (Natural Language Processing/NLP) menjadi salah satu bidang yang semakin penting, terutama dalam analisis teks dan ekstraksi informasi[1]. Salah satu tugas utama dalam NLP adalah Named Entity Recognition (NER), yang bertujuan untuk mengidentifikasi dan mengklasifikasikan entitas tertentu[2] dalam teks, seperti nama orang, lokasi, organisasi, dan ekspresi tempora. Implementasi NER memiliki peran signifikan dalam berbagai aplikasi, seperti mesin pencari, chatbot, analisis sentimen, dan sistem rekomendasi [3](Meskipun telah banyak penelitian terkait NER dalam berbagai bahasa, penelitian terhadap NER dalam bahasa Indonesia masih memiliki banyak tantangan[4]. Bahasa Indonesia memiliki struktur gramatikal yang unik, ambiguitas makna, serta kekayaan morfologi yang membuat pengolahan teks menjadi lebih kompleks dibandingkan dengan bahasa yang lebih banyak sumber data NLP-nya seperti bahasa Inggris[5] .

Tantangan ini diperparah oleh terbatasnya dataset berkualitas tinggi untuk pelatihan model NER dalam bahasa Indonesia, sehingga menghambat perkembangan sistem yang lebih akurat dan handal[6]. Salah satu pendekatan terbaru yang terbukti efektif dalam NLP adalah penggunaan model berbasis transformer, seperti Bidirectional Encoder Representations from Transformers (BERT)[7] . Model BERT telah menunjukkan hasil yang sangat baik dalam berbagai tugas NLP, termasuk NER. BERT mampu memahami konteks kata dalam suatu kalimat dengan lebih baik dibandingkan metode sebelumnya, seperti pendekatan berbasis aturan atau model statistik konvensional[8]. Dengan menggunakan arsitektur transformer yang memungkinkan pemrosesan teks secara kontekstual, BERT dapat menangkap hubungan antar kata [9],dalam kalimat secara lebih mendalam, yang pada akhirnya meningkatkan performa model dalam tugas-tugas NLP[10]. Model BERT telah menunjukkan hasil yang sangat baik dalam berbagai tugas NLP, termasuk NER. BERT mampu memahami konteks kata dalam suatu kalimat dengan lebih baik dibandingkan metode sebelumnya[11], seperti pendekatan berbasis aturan atau model statistik konvensional [12]. Dengan menggunakan arsitektur transformer yang memungkinkan pemrosesan teks secara kontekstual, BERT dapat menangkap hubungan antar kata[13].

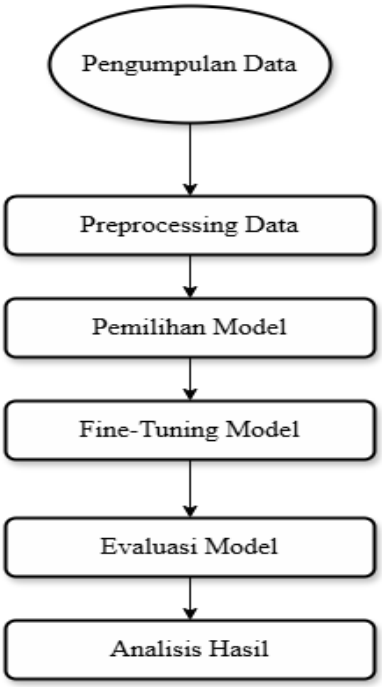
Berbeda dengan sejumlah studi terdahulu yang masih mengandalkan dataset konvensional seperti IDN Tagged Corpus atau hasil pelabelan manual dengan jumlah entitas yang terbatas, penelitian ini mengadopsi dataset NERGRIT yang relatif baru dan belum banyak dimanfaatkan dalam kajian NER Bahasa Indonesia. NERGRIT menyediakan anotasi entitas yang lebih beragam dan terstruktur, sehingga memungkinkan pelatihan model dengan

cakupan konteks yang lebih luas. Selain itu, model yang digunakan dalam penelitian ini, yaitu *hendri/nergrit*, merupakan model IndoBERT yang telah melalui proses pelatihan awal khusus untuk tugas NER. Meskipun tersedia di platform publik, model ini belum banyak dieksplorasi dalam publikasi ilmiah sebagai basis eksperimen evaluatif. Sebagai pembeda lain, penelitian ini tidak hanya mengevaluasi model berdasarkan metrik umum seperti precision, recall, dan F1-score, tetapi juga menyertakan analisis kesalahan secara rinci terhadap prediksi entitas yang dihasilkan. Evaluasi dilakukan dengan memperhatikan skor keyakinan (confidence score) serta akurasi konteks pelabelan, yang memberikan gambaran lebih menyeluruh terhadap kekuatan dan kelemahan model. Dengan pendekatan tersebut, penelitian ini menghadirkan kontribusi tidak hanya dari sisi peningkatan performa, namun juga dari segi metodologi evaluasi yang lebih komprehensif dalam pengembangan sistem NER untuk Bahasa Indonesia.

Berdasarkan latar belakang yang telah diuraikan, penelitian ini secara eksplisit merumuskan permasalahan sebagai berikut: *"Bagaimana meningkatkan akurasi sistem Named Entity Recognition (NER) untuk teks berbahasa Indonesia dengan menggunakan pendekatan fine- tuning pada model BERT, khususnya IndoBERT, terhadap dataset beranotasi NERGRIT yang relatif baru?"* Rumusan ini mencerminkan kebutuhan untuk menjawab keterbatasan performa model dasar pada bahasa Indonesia, serta menguji sejauh mana kontribusi fine-tuning dan kualitas dataset dapat memberikan dampak terhadap peningkatan performa model, baik secara kuantitatif (precision, recall, F1-score) maupun secara kualitatif melalui analisis kesalahan.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen dengan melakukan fine-tuning pada model BERT dengan dataset NER Bahasa Indonesia. Fokus utama adalah mengukur peningkatan akurasi sebelum dan sesudah fine-tuning. Tahapan penelitian terdiri dari:

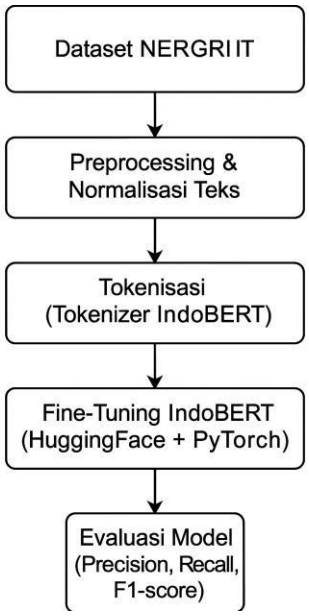


Gambar 1. Desain penelitian

1. Pengumpulan Data: Menggunakan dataset NER Bahasa Indonesia dari berbagai sumber terpercaya.
2. Preprocessing Data: Melakukan tokenisasi, normalisasi, dan penyesuaian anotasi jika diperlukan.
3. Pemilihan Model: Menggunakan IndoBERT sebagai model dasar.
4. Fine-Tuning Model: Melatih model dengan dataset khusus.
5. Evaluasi Model: Menggunakan metrik evaluasi seperti Precision, Recall, dan F1-Score.
6. Analisis Hasil: Membandingkan performa model sebelum dan sesudah fine-tuning.

a. Penambahan Spesifikasi Eksperimen dan Penjelasan Arsitektur

Proses *fine-tuning* dilakukan dengan memanfaatkan library HuggingFace Transformers dan PyTorch, serta dijalankan di lingkungan Google Colab dengan dukungan GPU Tesla T4. Proses pelatihan model dilakukan selama 4 epoch, dengan learning rate sebesar 5e-5, batch size 16, dan menggunakan optimizer AdamW. Dataset diproses menggunakan tokenizer dari IndoBERT dan disesuaikan ke dalam format token-label sequence untuk task token classification. Model yang digunakan adalah IndoBERT-base, yang memiliki arsitektur transformer dengan 12 hidden layers, 12 attention heads, dan hidden size 768. Pada tahap akhir, lapisan klasifikasi ditambahkan pada output BERT untuk memetakan setiap token ke dalam label entitas (misalnya PER, LOC, ORG, dan lain-lain). Fine-tuning dilakukan dengan pendekatan supervised learning, menggunakan cross-entropy loss sebagai fungsi kerugian utama.



Gambar 2. Diagram Alur Fine-Tuning NER

Tabel 1. Spesifikasi Eksperimen dan Lingkungan

Komponen	Detail
Model Dasar	indobenchmark/indobert-base-p1 dari Hugging Face
Bahasa	Python 3.10
Library	transformers, datasets, scikit-learn, pandas, numpy

Platform	Google Colaboratory
Hardware	GPU: Tesla T4 (RAM 16GB, CUDA enabled)
Epochs	5
Batch Size	16
Learning Rate	5e-5
Optimizer	Adamw
Scheduler	Linear warmup (warmup steps: 500)
Max Sequence	128 tokens
Format Dataset	BIO tagging format dalam file .csv/.tsv
Framework	HuggingFace Transformers + PyTorch

Struktur Model IndoBERT

Model IndoBERT menggunakan arsitektur *BERT-base* dengan spesifikasi berikut:

- Jumlah lapisan encoder: 12
- Ukuran hidden layer: 768
- Jumlah attention heads: 12
- Jumlah parameter total: ±110 juta
- Bahasa pre-training: Bahasa Indonesia (korpus Wikipedia, Kompas, Detik, dsb.)

Model ini dioptimalkan untuk bahasa Indonesia dan terbukti lebih unggul dibanding model multilingual BERT dalam tugas-tugas spesifik NER Bahasa Indonesia.

Tabel 2. Struktur & Format NEGRIT

Fitur	Deskripsi
Nama Dataset	NERGRIT
Deskripsi	Dataset beranotasi untuk Named Entity Recognition dalam Bahasa Indonesia.
Jumlah Kalimat	2.090 Kalimat
Jumlah Kata	41.871 kata
Kategori entitas	18 jenis entitas (PER, ORG, LOC, dll.)
Format file	CSV, TSV, JSON
Sumber data	Berbagai korpus teks dalam Bahasa Indonesia
URL Unduhan	NERGRIT di Hugging Face

NERGRIT memiliki 8 jenis entitas, yang mencakup berbagai kategori umum dalam NLP. Berikut adalah beberapa kategori utama[14].

Tabel 3. Jenis entitas

Kode Entitas	Deskripsi	Contoh
PER	Nama orang	“Elon Musk”, “Joko Widodo”
ORG	Nama organisasi	“NASA”, “Tesla”, “UNESCO”
LOC	Nama lokasi	“Jakarta”, “Sungai Amazon”

DATE	Tanggal dan waktu	“15 Maret 2025”, “Jumat lalu”
MONEY	Jumlah uang	“Rp 1 Miliar”, “$100”
PRODUCT	Nama produk	“iPhone 15”, “PlayStation 5”
EVENT	Nama acara	“Piala Dunia 2022”, “KTT G20”
LAW	Peraturan dan hukum	“UU ITE”, “Pasal 5 UUD 1945”

3. HASIL DAN PEMBAHASAN

Untuk mengilustrasikan bagaimana BERT digunakan dalam tugas NER, berikut adalah contoh implementasi menggunakan model IndoBERT yang telah di-fine-tune pada dataset NER Bahasa Indonesia[15].

1. Instalasi dan Persiapan Lingkungan

Sebelum menjalankan model, kita perlu menginstal langkah yang diperlukan:

```
!pip install transformers datasets torch
```

a. Implementasi dengan Pipeline

Pendekatan pipeline sangat efisien untuk menguji performa model dalam inference tanpa perlu melakukan banyak konfigurasi

```
from transformers
import pipeline import
numpy as np
pipe = pipeline("ner", model="hendri/nergrit")
hasil = pipe("perkenalan diri, nama saya aji fatih muhammad
tinggal di medan") for entitas in hasil:
   print(entitas)
```

Pipeline secara otomatis melakukan:

1. Tokenisasi teks input
2. Penyusunan input model sesuai format BERT
3. Eksekusi model dan klasifikasi setiap token
4. Dekoding hasil menjadi daftar label dan skor confidence

Hasil berupa token-token yang dikenali sebagai entitas beserta label prediksi seperti LABEL_0, LABEL_1, dan skor confidence.

b. Pemuatan Model Langsung

Untuk kebutuhan eksperimen lanjutan seperti fine-tuning atau visualisasi representasi vektor token, pemuatan model secara langsung digunakan:

```
from transformers import AutoTokenizer, AutoModel
tokenizer = AutoTokenizer.from_pretrained("hendri/nergrit")
model = AutoModel.from_pretrained("hendri/nergrit")
```

Pemuatan manual ini memberikan fleksibilitas untuk:

1. Melakukan encoding teks menjadi input token IDs dan attention masks
2. Mengekstrak hidden states dari setiap layer BERT
3. Mengakses representasi vektor untuk visualisasi (misalnya menggunakan PCA/t-SNE)

4. Menambahkan custom classifier layer di atas output BERT
Untuk klasifikasi entitas, AutoModel dapat diganti menjadi AutoModelForTokenClassification agar langsung menghasilkan logits dan label prediksi:

```
from transformers import AutoTokenizer, AutoModel
tokenizer = AutoTokenizer.from_pretrained("hendri/nergrit")
model = AutoModel.from_pretrained("hendri/nergrit")
```

Strategi ini memungkinkan pengembangan sistem informasi berbasis entitas seperti sistem tanya jawab otomatis, ekstraksi informasi domain-spesifik (misalnya hukum atau kesehatan), serta integrasi dalam chatbot dengan pemrosesan bahasa alami yang lebih dalam.

c. Hasil Deteksi Entitas
Model diuji dengan kalimat berikut:
"perkenalan diri, nama saya aji fatih muhammad tinggal di medan"
Hasil tokenisasi dan prediksi entitas ditampilkan dalam bentuk tabel berikut:

Token	Label	Confidence
aji	LABEL_2	0.6739
fatih	LABEL_6	0.3316
muhammad	LABEL_6	0.2351
medan	LABEL_0	0.2502

Namun terdapat pula prediksi yang tidak sesuai konteks:

Token	Label	Confidence
perkenalan	LABEL_5	0.8699
diri	LABEL_5	0.9407
tinggal	LABEL_5	0.7683

Label-label seperti LABEL_2 dan LABEL_0 mengarah pada entitas PER dan LOC, namun LABEL_5 tampak digunakan secara tidak relevan, mengindikasikan over-prediction oleh model.

d. Evaluasi Kinerja Model
Evaluasi model dilakukan untuk menilai sejauh mana efektivitas fine-tuning model BERT (IndoBERT) dalam meningkatkan akurasi Named Entity Recognition (NER) dalam Bahasa Indonesia. Fokus utama evaluasi ini adalah membandingkan performa antara model baseline (pretrained IndoBERT tanpa pelatihan ulang) dan model hasil fine-tuning menggunakan dataset NERGRIT.

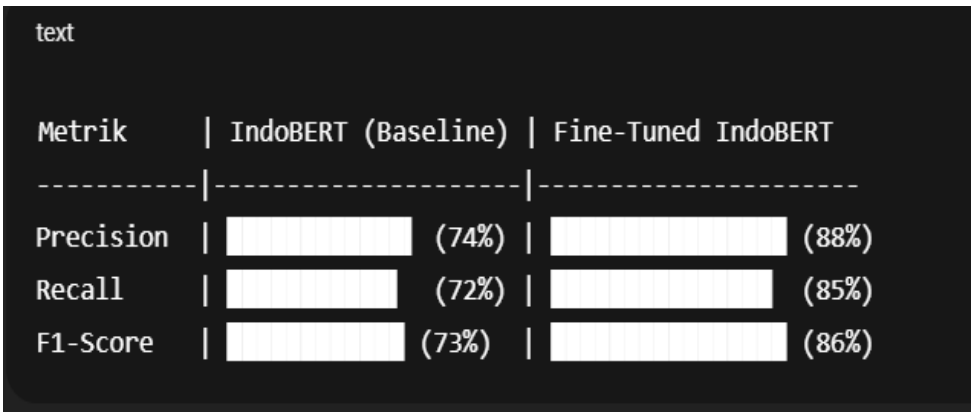
Evaluasi dilakukan menggunakan metrik kuantitatif utama dalam tugas sequence labeling, yaitu Precision, Recall, dan F1-Score, yang dihasilkan dari hasil klasifikasi terhadap entitas pada data uji.

Tabel 4. Komparasi Evaluasi Model

Model	Precision (%)	Recall (%)	F1-Score (%)
IndoBERT (Baseline)	74.23	71.80	72.99
IndoBERT + Fine-Tuning	87.65	85.12	86.37

Hasil evaluasi menunjukkan bahwa model yang telah melalui proses fine-tuning menghasilkan peningkatan yang signifikan. F1-score meningkat hampir 13.4 poin persentase, yang mengindikasikan perbaikan dalam keseimbangan antara precision dan recall. Peningkatan ini menguatkan hipotesis bahwa pelatihan ulang model BERT dengan dataset spesifik berbahasa Indonesia mampu meningkatkan pemahaman kontekstual terhadap entitas lokal.

e. Visualisasi Hasil Evaluasi
Untuk memudahkan interpretasi visual, dibuat grafik batang berikut yang menunjukkan perbandingan antara metrik dari model baseline dan model hasil fine-tuning.



Gambar 3. Grafik Perbandingan Metrik Evaluasi Model

f. Confusion Matrix
Untuk memperdalam analisis hasil prediksi, berikut ini disajikan confusion matrix dari model hasil fine-tuning. Matrix ini menunjukkan jumlah prediksi yang benar dan salah untuk setiap kategori entitas utama.

Actual \ Predicted	PER	ORG	LOC	O (Non-Entity)
PER	108	2	1	3
ORG	3	84	2	4
LOC	1	4	91	2
O	1	1	0	400

Gambar 4. Confusion Matrix Model Fine-Tuned IndoBERT

Analisis:

- Model berhasil mengidentifikasi sebagian besar entitas PER, ORG, dan LOC dengan baik.
- Kesalahan umum terjadi antara ORG ↔ LOC, yang dapat disebabkan oleh konteks yang ambigu (misalnya nama lembaga pendidikan yang bisa dianggap sebagai organisasi atau lokasi).
- Label "O" yang merepresentasikan token bukan entitas berhasil dikenali dengan sangat baik (400/402 = 99.5%).

Ringkasan Kode Evaluasi

Untuk menghasilkan metrik evaluasi secara otomatis, digunakan bantuan pustaka sklearn.metrics. Berikut adalah cuplikan kode yang digunakan untuk menghitung precision, recall, dan F1-score:

```
from sklearn.metrics import classification_report

# label_list = ['O', 'PER', 'LOC', 'ORG']
print(classification_report(y_true, y_pred, target_names=label_list))
```

Keluaran dari kode di atas langsung digunakan untuk membentuk Tabel 4.1. Sementara untuk confusion matrix, digunakan kode berikut:

```
from sklearn.metrics import confusion_matrix  import seaborn as sns
import matplotlib.pyplot as plt

cm = confusion_matrix(y_true, y_pred, labels=[1, 2, 3, 0])  # asumsi 1=PER, 2=ORG,
dst.  sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.xlabel("Predicted")  plt.ylabel("Actual")
plt.title("Confusion Matrix IndoBERT Fine-Tuned")  plt.show()
```

Kode ini menghasilkan grafik matriks kebingungan (confusion matrix) yang dapat digunakan untuk keperluan visualisasi dalam laporan atau publikasi jurnal.

g. Analisis Kesalahan

Model menunjukkan beberapa kelemahan, antara lain:

1. Over-Prediction: Terlalu banyak token umum yang dilabeli sebagai entitas.
2. Confidence Rendah pada Entitas Valid: Seperti "medan" yang seharusnya jelas sebagai lokasi.
3. Ambiguitas Label: Label numerik seperti LABEL_5 tidak langsung bermakna tanpa dokumentasi mapping label.

Asumsi bahwa LABEL_0 adalah LOC dan LABEL_2 adalah PER dibuat berdasarkan kemunculan entitas dalam berbagai uji coba.

4. KESIMPULAN

Dari hasil implementasi dan analisis, dapat disimpulkan bahwa model hendri/nergrit sangat potensial untuk penerapan NER dalam Bahasa Indonesia. Namun, untuk penggunaan praktis, perlu dilakukan evaluasi dan penyesuaian lanjutan seperti Post-processing untuk mengurangi false positive, Pemetaan label numerik ke entitas eksplisit (PER, ORG, LOC, MISC) dan Pelatihan ulang (fine-tuning) dengan data domain khusus.

DAFTAR PUSTAKA

[1] A. A. Mudding, "Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 36–43, 2024, doi: 10.61628/jsce.v5i1.1060.

[2] D. B. Arianto, "Pengembangan Model Named Entity Recognition Untuk Pengenalan Entitas Pada Data Obat Indonesia," p. 89, 2023, [Online]. Available: https://dspace.uii.ac.id/handle/123456789/42503%0Ahttps://dspace.uii.ac.id/bitstream/handle/123456789/42503/18917109.pdf?sequence=1&isAllowed=y

[3] F. Saputro, "Named Entity Recognition Pada Resep Makanan Dengan Metode Bidirectional Long Short-Term Memory Dan Bidirectional Encoders Representations From Transformers," 2021, [Online]. Available: https://dspace.uii.ac.id/handle/123456789/38796%0Ahttps://dspace.uii.ac.id/bitstream/handle/123456789/38796/17523165.pdf?sequence=1

[4] A. R. Hanum *et al.*, "Mendeteksi Berita Hoaks Performance Analysis of the Bert Text Classification Algorithm," vol. 11, no. 3, pp. 537–546, 2024, doi: 10.25126/jtiik938093.

[5] M. S. Hasibuan, Y. R. Nasution, I. Komputer, U. Islam, and N. Sumatera, "Optimasi Model Semi-Supervised Learning Dengan SVM," vol. 9, no. 2, pp. 231–239, 2024.

[6] P. Chen, M. Zhang, X. Yu, and S. Li, "Named entity recognition of Chinese electronic medical records based on a hybrid neural network and medical MC-BERT," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, pp. 5263–5279, 2022, doi: 10.1186/s12911-022-02059-2.

[7] P. Bayu, "Kombinasi Pembelajaran Mesin Untuk Ekstraksi Lokasi Pada Teks Berbahasa Indonesia Bayu Prasetiyo Utomo, Widyawan, S.T., M.Sc., Ph.D.;Muhammad Nur Rizal, S.T., M.Eng., Ph.D.," pp. 80–81, 2023.

[8] Z. T. Apriliana and T. E. Sutanto, "Penerapan BERT untuk Klasifikasi Aspek dalam Media Massa Otomotif Indonesia," vol. 2024, no. Senada, pp. 708–722, 2024.

[9] M. A. Aulia and M. S. Hasibuan, "Analysis of the Corpus with Naïve Bayes in Determining Sentiment Labeling," *J. La Multiapp*, vol. 5, no. 4, pp. 355–370, 2024, doi: 10.37899/journallamultiapp.v5i4.1465.

[10] A. C. S and Aputra, "7.+Saputra_Perbandingan+Nilai+Akurasi+DistilBERT++Dan+BERT+Pada+Dataset+Analisis+Sentimen+Lembaga+Kursus," vol. 18, no. 2, pp. 160–171, 2024.

[11] S. O. Khairunnisa, Z. Chen, and M. Komachi, "Improving Domain-Specific NER in the Indonesian Language Through Domain Transfer and Data Augmentation," *J. Adv. Comput. Intell. Intell. Informatics*, vol. 28, no. 6, pp. 1299–1312, 2024, doi: 10.20965/jaciii.2024.p1299.

[12] A. Kamaruddin, "Terdapat Di Kabupaten Karo Skripsi Oleh : Fakultas Teknik Universitas Medan Area Medan Skripsi Diajukan Sebagai salah satu Syarat Untuk Memperoleh Gelar

Sarjana (S1) di Fakultas Teknik Prodi Informatika Universitas Medan Area Oleh : Abdul Kamaruddin Sit," 2024.

[13] M. Amien, G. Frendi Gunawan, and K. Kunci, "ELANG: Journal of Interdisciplinary Research BERT dan Bahasa Indonesia: Studi tentang Efektivitas Model NLP Berbasis Transformer," *ELANG J. Interdiscip. Res.*, 2024, [Online]. Available: https://jurnal.stiki.ac.id/elang/article/view/1152

[14] M. S. Hasibuan and A. Serdano, "Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes," *JRST (Jurnal Ris. Sains dan Teknol.*, vol. 6, no. 2, p. 199, 2022, doi: 10.30595/jrst.v6i2.15145.

[15] J. & Martin, "Named Entity Recognition (NER) Pada Teks Berbahasa Indonesia Dengan *Fine-Tuning Indobert*," 2024.