

**PENERAPAN *NAMED ENTITY RECOGNITION (NER)* MENGGUNAKAN
INDOBERT PADA CERITA RAKYAT “HIKAYAT WAYANG ARJUNA &
PURUSARA”**

SKRIPSI

Diajukan Untuk Penelitian Skripsi Guna Memenuhi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer (S.Kom.)

Program Studi Sistem Informasi UN PGRI Kediri



OLEH :

DANDA BAGASKORO

NPM : 2213030083

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI

2026

DANDA BAGASKORO

NPM : 2213030083

Judul:

**PENERAPAN *NAMED ENTITY RECOGNITION (NER)*
MENGUNAKAN INDOBERT PADA CERITA RAKYAT “HIKAYAT
WAYANG ARJUNA & PURUSARA”**

**Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri**

Pembimbing I



Dr. Erna Daniati, M.Kom
NIDN. 0723058501

Pembimbing II



Rina Firliana, M.Kom.
NIDN. 0731087703

Skripsi oleh:

DANDA BAGASKORO

NPM : 2213030083

Judul

**PENERAPAN *NAMED ENTITY RECOGNITION (NER)* MENGGUNAKAN
INDOBERT PADA CERITA RAKYAT “HIKAYAT WAYANG ARJUNA &
PURUSARA”**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

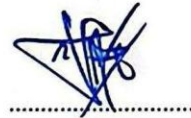
Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 26 Juni 2026

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua : Dr. Erna Daniati, M.Kom



2. Penguji 1 : Dwi Harini, S.Si., M.M



3. Penguji 2 : Rina Firliana, M.Kom



Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer

Dr. Sulistiono, M.Si

NIDN. 0007076801

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Danda Bagaskoro
Jenis Kelamin : Laki-Laki
Tempat/tgl. lahir : Nganjuk/ 17 April 2003
NPM : 2213030083
Fak/Prodi. : FTIK/ S1-Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 26 Juni 2026

Yang Menyatakan



DANDA BAGASKORO

NPM: 2213030083

MOTTO

HET LEVEN IS GEEN LOLLETJE

“Hidup ini bukan kesenangan tapi penuh suka dan duka serta perjuangan”

(H.R. DJIMAT HENDRO SUWARNO)

“TAT TWAM ASI”

Aku adalah kamu dan kamu adalah ku. Menolong orang lain berarti menolong diri sendiri dan menyakiti orang lain berarti menyakiti diri sendiri

ABSTRAK

Danda Bagaskoro : *Penerapan Named Entity Recognition (NER) Menggunakan IndoBERT pada Cerita Rakyat Hikayat Wayang Arjuna dan Purasara*, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2026..

Kata Kunci : *Named Entity Recognition, IndoBERT, Natural Language Processing, Cerita Rakyat, Hikayat Wayang.*

Perkembangan teknologi pengolahan bahasa alami (*Natural Language Processing* atau NLP) telah membuka peluang dalam pengelolaan dan ekstraksi informasi dari data teks secara otomatis. Salah satu permasalahan yang sering ditemukan pada dokumen teks adalah sulitnya mengidentifikasi informasi penting seperti nama tokoh, nama tempat, maupun nama organisasi secara cepat dan akurat. Permasalahan tersebut juga ditemukan pada cerita rakyat Indonesia yang umumnya memiliki banyak tokoh, lokasi, dan unsur budaya yang tersebar di dalam teks. Oleh karena itu, diperlukan suatu sistem yang mampu melakukan identifikasi entitas secara otomatis untuk memudahkan proses pengolahan informasi pada cerita rakyat.

Penelitian ini bertujuan untuk menerapkan metode *Named Entity Recognition* (NER) menggunakan model IndoBERT dalam mengenali entitas pada cerita rakyat Hikayat Wayang Arjuna dan Purasara. Entitas yang menjadi fokus penelitian meliputi *Person* (PER), *Location* (LOC), dan *Organization* (ORG). Dataset yang digunakan merupakan dataset NER Bahasa Indonesia yang telah diberi label menggunakan skema BIO (*Beginning, Inside, Outside*). Tahapan penelitian meliputi pembacaan dataset, pelabelan data, tokenisasi menggunakan tokenizer IndoBERT, proses *fine-tuning* model, evaluasi model, serta pengujian menggunakan kalimat-kalimat yang diambil dari cerita Hikayat Wayang Arjuna dan Purasara.

Metodologi penelitian yang digunakan mengacu pada tahapan pengembangan sistem pembelajaran mesin yang terdiri dari pengumpulan data, persiapan data, pemodelan, evaluasi, dan pengujian. Model IndoBERT dilatih menggunakan 1.464 data pelatihan, 367 data validasi, dan 509 data pengujian. Evaluasi model dilakukan menggunakan metrik *Precision*, *Recall*, dan *F1-Score* untuk mengukur kemampuan model dalam mengenali entitas yang terdapat pada teks.

Hasil penelitian menunjukkan bahwa model IndoBERT mampu melakukan identifikasi entitas pada teks berbahasa Indonesia dengan nilai *Precision* sebesar

56,30%, *Recall* sebesar 60,55%, dan *F1-Score* sebesar 58,35%. Berdasarkan hasil evaluasi per kategori entitas, kategori *Location* memperoleh performa terbaik dengan nilai *F1-Score* sebesar 76,07%, diikuti kategori *Person* sebesar 64,67%, dan kategori *Organization* sebesar 44,96%. Hasil tersebut menunjukkan bahwa model IndoBERT cukup efektif dalam mengenali entitas pada teks cerita rakyat Indonesia, khususnya untuk kategori lokasi dan tokoh.

Penelitian ini diharapkan dapat menjadi dasar dalam pengembangan sistem ekstraksi informasi otomatis pada cerita rakyat Indonesia serta mendukung upaya digitalisasi dan pelestarian warisan budaya melalui pemanfaatan teknologi *Natural Language Processing*.

KATA PENGANTAR

Puji Syukur Kami panjatkan Kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul “Penerapapan *Named Entity Recognition (NER)* Menggunakan *Indobert* pada cerita rakyat hikayat wayang arjuna & purusara”:ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri. Pada kesempatan ini diucapkan terima kasih dan penghargaan yang setulus-tulusnya kepada :

1. **Dr. Zainal Afandi, M.Pd.** selaku Rektor UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
2. **Dr. Sulistiono, M.Si.** selaku Dekan Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
3. **Dr. Sucipto, M.Kom.** selaku Ka Prodi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
4. **Dr. Erna Daniati, M.Kom** dan **Rina Firliana, M.Kom** selaku Pembimbing skripsi yang dengan sabar membimbing dan memotivasi dalam menyelesaikan skripsi ini.
5. Bapak Ibu dosen Program Studi Sistem Informasi atas segala ilmu, bimbingan, dan dedikasi yang telah diberikan selama masa studi.
6. Orang Tua dan Keluarga, yang telah mendukung dalam do’a, dana dan motivasi dalam penyelesaian skripsi ini.
7. Teman-teman yang selalu membantu memberi semangat dalam penyelesaian skripsi ini.

Kediri, 26 Juni 2026



DANDA BAGASKORO

NPM : 221303008

DAFTAR ISI

Halaman Pengesahan	iii
Halaman Pernyataan	iv
Halaman Motto.....	v
ABSTRAK	vi
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Batasan Masalah	3
C. Rumusan Masalah	4
D. Tujuan Penelitian	4
E. Manfaat Penelitian	4
BAB II KAJIAN PUSTAKA	8
A. Kajian Teori	8
B. Kajian Hasil Penelitian Terdahulu.....	13
C. Kerangka Berpikir.....	19
BAB III METODE PENELITIAN.....	22
A. Desain Penelitian.....	22
B. Objek Penelitian	22
C. Dataset Penelitian.....	23
D. Pelabelan Entitas	24
E. Arsitektur indoBERT	25
F. Persiapan Data.....	25
G. Implementasi Model IndoBERT	25
H. Tokenisasi dan Penyelarasan Label	26
I. Evaluasi Model.....	27
J. Pengujian Model	28
K. Alur Penelitian	28

BAB IV HASIL DAN PEMBAHASAN	35
A. Gambaran Umum Hasil Penelitian	35
B. Hasil Pembacaan Dataset	35
C. Hasil Pelabelan Entitas.....	36
D. Hasil Implementasi Model IndoBERT	37
E. Hasil Tokenisasi dan Alignment Label.....	37
F. Hasil Peatihan Model	38
G. Hasil Evluasi Model	39
H. Analisis Per Kategori Entitas	41
I. Entitas Person (PER)	42
J. Entitas Organization (ORG	43
K. Hasil Pengujian pada Cerita Hikayat Wayang.....	43
L. Pembahasan.....	44
BAB V PENUTUP.....	46
A. Kesimpulan.....	46
B. Saran.....	47
DAFTAR PUSTAKA	48

DAFTAR GAMBAR

Gambar 2.1 Kerangka Berpikir	19
Gambar 3.1 Pembacaan Dataset	23
Gambar 3.2 Implemenrasi Indobert.....	26
Gambar 3.3 Proses Tokenisasi Pelatihan Model	26
Gambar 3.4 Alur Penelitian Model NER Menggunakan Indobert	29
Gambar 4.1 Pembacaan Dataset	35
Gambar 4.2 Hasil Pelabelan Entitas	36
Gambar 4.3 Implemenrasi Model Indobert	37
Gambar 4.4 Proses Tokenisasi.....	37
Gambar 4.5 Hasil Pelatihan Model	38
Gambar 4.6 Grafik Precision, Recall, dan F1-Score	39
Gambar 4.7 Grafik Loss Model.....	40
Gambar 4.8 Grafik F-1 Score Tiap Entitas.....	41
Gambar 4.9 Grafik Hasil Pengujian	43

DAFTAR TABEL

Tabel 2.1 Kajian Hasil Penelitian Terdahulu.....	19
Tabel 3.1 Dataset Penelitian	23
Tabel 3.2 Pelabelan Entitas	24
Tabel 3.3 Hasil Pelabelan	24
Tabel 3.4 Persiapan Data	25
Tabel 3.5 Hasil Parameter	27
Tabel 3.6 Spesifikasi Perangkat Keras	32
Tabel 3.7 Spesifikasi Perangkat Lunak	33
Tabel 3.8 Spesifikasi Model Indobert	33
Tabel 3.9 Konfigurasi Pelatihan Model	34
Tabel 4.1 Jumlah Data.....	36
Tabel 4.2 Hasil Pelabelan Entitas.....	36
Tabel 4.3 Hasil Pelatihan Model	38
Tabel 4.4 Hasil Perkembangan F-1 Score	38
Tabel 4.5 Hasil Evaluasi Model	39
Tabel 4.6 Hasil Evaluasi.....	42
Tabel 4.7 Entitas Person.....	42
Tabel 4.8 Entitas Organization	43
Tabel 4.9 Contoh Hasil Pengujian pada Kalimat Hikayat Wayang	44

DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan Skripsi	65
Lampiran 2 Surat Keterangan Bebas Similarity	66
Lampiran 3 Halaman Awal Cek Similarity	67
Lampiran 4 Kartu Bukti Submit Artikel.....	67

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence / AI*) telah mendorong kemajuan pesat dalam bidang pengolahan bahasa alami atau *Natural Language Processing* (NLP). Sebagai salah satu cabang utama AI, NLP memungkinkan komputer untuk memahami, menafsirkan, dan memanipulasi bahasa manusia. Salah satu tugas fundamental dan menantang dalam NLP adalah *Named Entity Recognition* (NER), yaitu proses otomatis untuk mengidentifikasi dan mengklasifikasikan entitas bernama dalam teks menjadi kategori yang telah ditentukan, seperti nama orang (*Person*), lokasi (*Location*), organisasi (*Organization*), dan istilah khusus lainnya. Kemampuan NER menjadi tulang punggung bagi berbagai aplikasi canggih seperti sistem temu balik informasi (*information retrieval*), penambangan teks (*text mining*), analisis sentimen, chatbots cerdas, dan yang terpenting, digitalisasi serta pelestarian dokumen bersejarah dan sastra (Daniati, Prasetya, Sakti, Irianto, dan Hernandez 2025).

Dalam konteks linguistik Indonesia, pengembangan sistem NER yang robust menghadapi tantangan multidimensi. (Daniati, Prasetya, Sakti, Irianto, dan Nafalski 2025) Bahasa Indonesia, dengan struktur morfologis yang kompleks, tingkat polisemi yang tinggi, keberagaman dialek, dan pengaruh dari banyak bahasa daerah serta bahasa asing (seperti Sanskerta, Arab, Belanda, dan Inggris), menyajikan landscape yang unik. Tantangan ini semakin mengemuka ketika dihadapkan pada teks sastra klasik atau *literatur tradisional*, yang sering kali ditulis dalam gaya bahasa yang arkais, menggunakan kosakata yang sudah jarang dipakai, dan mengandung struktur naratif serta ungkapan budaya yang khas. Teks-teks semacam ini tidak hanya berfungsi sebagai karya sastra, tetapi juga sebagai repositori kearifan lokal, nilai-nilai filosofis, dan identitas budaya Nusantara (Ding et al. 2021).

Sayangnya, sebagian besar khazanah sastra klasik Indonesia, termasuk berbagai hikayat, babad, dan cerita rakyat, masih tersimpan dalam bentuk naskah fisik, manuskrip tulisan tangan, atau dokumen digital tanpa struktur dan metadata yang

memadai. Proses digitalisasi yang hanya sekadar mengubah bentuk fisik menjadi teks digital (*plain text*) tanpa dilengkapi dengan analisis semantik dan enkapsulasi informasi menjadikan naskah-naskah tersebut kurang dapat diakses, dicari, dan dimanfaatkan secara maksimal dalam era digital (Daniati et al. 2024). Di sinilah NER berperan sebagai teknologi kunci untuk mengekstrak informasi berharga, mengidentifikasi entitas-entitas penting, dan pada akhirnya mengubah teks mentah menjadi data terstruktur yang kaya pengetahuan (Pradhana, Daniati, dan Najibulloh 2025).

Di antara sekian banyak cerita rakyat Indonesia, "Hikayat Wayang Arjuna & Purusara" dipilih sebagai objek penelitian. Hikayat ini merupakan bagian dari tradisi wayang yang mendalam, tidak hanya menceritakan kisah epik tokoh-tokoh seperti Arjuna, tetapi juga sarat dengan istilah budaya, nama tempat mitologis, dan konsep filosofis Jawa. Analisis manual terhadap teks sepanjang ini untuk mengidentifikasi semua entitas yang relevan adalah pekerjaan yang sangat memakan waktu dan rentan terhadap subjektivitas. Oleh karena itu, otomatisasi melalui NER menjadi sebuah kebutuhan (Susanty 2021).

Perkembangan terbaru dalam bidang NLP diwarnai oleh kehadiran model berbasis *transformer*, dengan BERT (*Bidirectional Encoder Representations from Transformers*) sebagai pionir (Daniati, Prasetya, Sakti, Irianto, dan Hernandez 2025). Model ini merevolusi pemahaman kontekstual bahasa berkat mekanisme *self-attention* yang memungkinkan pemrosesan kata dengan mempertimbangkan seluruh konteks kalimat dari kedua arah. Untuk bahasa Indonesia, IndoBERT hadir sebagai model *pretrained* yang khusus dilatih dengan korpus besar berbahasa Indonesia, sehingga memiliki pemahaman awal yang baik terhadap sintaksis dan semantik bahasa tersebut. Kinerja unggul IndoBERT dalam berbagai tugas NLP dasar menjadikannya kandidat yang potensial untuk di-*fine-tune* pada tugas spesifik seperti NER untuk domain teks klasik (Ferdiana et al. 2024).

Namun, penerapan IndoBERT untuk NER pada teks sastra klasik seperti "Hikayat Wayang Arjuna & Purusara" bukan tanpa tantangan. Model ini pada awalnya dilatih dengan data bahasa Indonesia kontemporer (berita, artikel web, dll.), sehingga kemampuannya dalam memahami bahasa lama, istilah khusus wayang, dan struktur naratif kuno masih perlu diuji dan ditingkatkan. Selain itu, belum tersedianya dataset

teranotasi untuk entitas budaya dalam teks klasik menjadi kendala utama dalam proses pelatihan dan evaluasi model.

Berdasarkan uraian di atas, penelitian ini hadir untuk menjawab tantangan tersebut. Dengan menerapkan teknik NER menggunakan model IndoBERT pada teks "Hikayat Wayang Arjuna & Purusara".

Dengan demikian, penelitian ini diharapkan tidak hanya memberikan sumbangan akademis pada perkembangan NLP untuk bahasa Indonesia, khususnya di domain teks klasik, tetapi juga memiliki dampak praktis dalam menjaga dan mempromosikan kekayaan sastra dan budaya Nusantara di dunia digital.

B. Batasan Masalah

Agar penelitian ini dapat dilaksanakan dengan fokus dan terukur, maka ruang lingkupnya dibatasi pada hal-hal berikut:

1. Pengarang Naskah: Naskah *Hikayat Wayang Arjuna & Purusara* yang digunakan dalam penelitian ini ditulis oleh Nikmah Sunardjo, Muhamad Fanani, Sri Sayekti, Putri Minerva Mutiara, dan Nurul Ainin.
2. Sumber Data: Teks yang digunakan hanya berasal dari satu naskah, yaitu "Hikayat Wayang Arjuna & Purusara".
3. Jenis Entitas: Klasifikasi entitas dibatasi hanya pada tiga kategori utama: PER (Nama Orang/Tokoh), LOC (Nama Lokasi/Tempat), dan CULT (Istilah Budaya/Khusus). Kategori istilah budaya mencakup nama senjata, pusaka, ajaran, atau istilah adat yang khas.
4. Model yang Digunakan: Penelitian akan memanfaatkan model *pretrained* IndoBERT (atau variannya seperti IndoBERT-lite) tanpa melakukan modifikasi signifikan terhadap arsitektur dasar model. Fokus utama adalah pada proses *fine-tuning* dan evaluasi.
5. Format Anotasi: Pelabelan data akan menggunakan skema BIO (Begin, Inside, Outside) atau IOB2, yang merupakan standar dalam tugas NER.
6. Metrik Evaluasi: Kinerja model akan diukur menggunakan metrik standar dalam evaluasi NER, yaitu Precision (Ketepatan), Recall (Lingkup Temuan), dan F1-Score (Rata-rata Harmonik Precision dan Recall).

C. Rumusan Masalah

Berdasarkan identifikasi dan batasan masalah di atas, rumusan masalah penelitian ini adalah:

1. Bagaimana karakteristik linguistik dan budaya dalam teks “*Hikayat Wayang Arjuna & Purusara*” yang memengaruhi proses identifikasi entitas bernama?
2. Tantangan apa saja yang muncul dalam penerapan *Named Entity Recognition* (NER) menggunakan IndoBERT pada teks sastra klasik Indonesia?
3. Bagaimana model IndoBERT beradaptasi dengan bahasa arkais dan istilah budaya dalam teks klasik selama proses *fine-tuning*?

D. Tujuan Penelitian

Sesuai dengan rumusan masalah, tujuan dari penelitian ini adalah:

1. Mengeksplorasi karakteristik linguistik dan budaya dalam teks “*Hikayat Wayang Arjuna & Purusara*” yang relevan dengan tugas NER. Untuk mengidentifikasi dan mengklasifikasikan entitas-entitas penting berupa tokoh (*Person*), lokasi (*Location*), dan istilah budaya (*Cultural-Term*) yang terdapat dalam teks hikayat tersebut.
2. Menganalisis adaptasi model IndoBERT terhadap bahasa arkais dan istilah budaya melalui studi kasus penerapan NER.
3. Mengidentifikasi tantangan dan peluang dalam penerapan NER berbasis *transformer* pada teks sastra klasik Indonesia.
4. Menginterpretasi makna budaya dari entitas yang berhasil diekstraksi dalam konteks pelestarian warisan literasi digital.

E. Manfaat Penelitian

Penelitian mengenai penerapan Named Entity Recognition (NER) menggunakan IndoBERT pada cerita rakyat Hikayat Wayang Arjuna dan Purasara diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis bagi berbagai pihak. Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Manfaat Teoretis
 - a. Kontribusi terhadap Pengembangan Ilmu Natural Language Processing (NLP)

Penelitian ini diharapkan dapat memperkaya kajian ilmiah di bidang Natural Language Processing (NLP), khususnya pada tugas Named Entity Recognition (NER) untuk bahasa Indonesia. Hasil penelitian dapat menjadi

referensi dalam pengembangan metode ekstraksi informasi otomatis yang memanfaatkan model berbasis Transformer seperti IndoBERT.

b. Menambah Literatur Penelitian NER Bahasa Indonesia

Penelitian mengenai NER pada teks cerita rakyat dan sastra klasik Indonesia masih relatif terbatas. Oleh karena itu, penelitian ini diharapkan dapat menjadi sumber referensi akademik yang memperluas literatur mengenai penerapan NER pada domain budaya dan sastra Indonesia.

c. Memberikan Pemahaman tentang Implementasi IndoBERT

Penelitian ini dapat memberikan gambaran mengenai proses implementasi, pelatihan (fine-tuning), dan evaluasi model IndoBERT pada tugas klasifikasi entitas. Hasil penelitian juga dapat digunakan sebagai bahan pembelajaran bagi peneliti maupun mahasiswa yang ingin mengembangkan penelitian serupa.

d. Menjadi Dasar Penelitian Lanjutan

Hasil penelitian dapat digunakan sebagai landasan bagi penelitian berikutnya yang berkaitan dengan ekstraksi informasi, analisis teks, question answering system, knowledge graph, maupun pengembangan model NLP lainnya yang berbasis data budaya dan sastra Indonesia.

2. Manfaat Praktis

a. Mendukung Digitalisasi Warisan Budaya Indonesia

Penelitian ini dapat membantu proses digitalisasi cerita rakyat dan naskah budaya Indonesia melalui identifikasi otomatis terhadap tokoh, lokasi, maupun organisasi yang terdapat dalam teks. Dengan demikian, informasi yang terkandung dalam naskah budaya dapat lebih mudah diakses, dipelajari, dan dilestarikan.

b. Mempermudah Pengolahan Informasi pada Cerita Rakyat

Hasil penerapan NER dapat membantu proses pengelompokan dan pencarian informasi penting dalam cerita rakyat secara otomatis. Pengguna tidak perlu lagi melakukan identifikasi tokoh atau lokasi secara manual karena sistem dapat mengenali entitas tersebut secara otomatis.

c. Membantu Peneliti dan Akademisi

Penelitian ini dapat dimanfaatkan oleh akademisi, peneliti, maupun mahasiswa yang melakukan penelitian di bidang kecerdasan buatan, data mining, linguistik komputasional, dan pengolahan bahasa alami. Dataset, model, maupun hasil penelitian dapat menjadi bahan pendukung untuk penelitian selanjutnya.

d. Mendukung Pengembangan Aplikasi Edukasi Budaya

Model NER yang dihasilkan dapat diintegrasikan ke dalam berbagai aplikasi pendidikan berbasis digital, seperti ensiklopedia wayang, perpustakaan digital, museum virtual, maupun media pembelajaran interaktif yang mengenalkan budaya Indonesia kepada masyarakat luas.

e. Meningkatkan Efisiensi Proses Analisis Teks

Penerapan model IndoBERT memungkinkan proses identifikasi entitas dilakukan secara otomatis dan lebih cepat dibandingkan metode manual. Hal ini dapat meningkatkan efisiensi dalam pengolahan dokumen teks berukuran besar, terutama dokumen yang mengandung banyak informasi mengenai tokoh, tempat, dan organisasi.

f. Menjadi Dasar Pengembangan Sistem Cerdas Berbasis Budaya

Model yang dihasilkan dalam penelitian ini dapat digunakan sebagai komponen awal dalam pengembangan sistem yang lebih kompleks, seperti sistem tanya jawab (Question Answering System), sistem rekomendasi informasi budaya, sistem pencarian dokumen cerdas, maupun knowledge graph yang memetakan hubungan antar tokoh dan lokasi dalam cerita rakyat Indonesia.

g. Mendukung Pelestarian Sastra dan Cerita Rakyat Nusantara

Melalui pemanfaatan teknologi kecerdasan buatan, penelitian ini diharapkan dapat meningkatkan minat masyarakat terhadap cerita rakyat Indonesia. Selain itu, hasil penelitian dapat membantu menjaga keberlangsungan warisan budaya bangsa dengan cara mendokumentasikan dan mengelola informasi budaya secara lebih sistematis dan modern.

Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang Natural Language Processing dan Named Entity Recognition, tetapi juga memiliki nilai strategis dalam mendukung digitalisasi,

pelestarian, serta pemanfaatan warisan budaya Indonesia melalui teknologi kecerdasan buatan berbasis IndoBERT.

Penelitian ini diharapkan dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya pada bidang Natural Language Processing (NLP) dan kecerdasan buatan yang berfokus pada pengolahan bahasa Indonesia. Hasil penelitian ini dapat menjadi referensi dalam penerapan metode Named Entity Recognition (NER) menggunakan model berbasis Transformer, khususnya IndoBERT, pada teks berbahasa Indonesia yang memiliki karakteristik khusus seperti cerita rakyat dan sastra klasik.

Bagi kalangan akademisi dan peneliti, penelitian ini dapat menjadi sumber informasi dan bahan rujukan untuk penelitian selanjutnya yang berkaitan dengan pengenalan entitas bernama pada teks budaya Indonesia. Selain itu, penelitian ini dapat memberikan gambaran mengenai proses pengembangan dataset, pelabelan entitas menggunakan skema BIO, serta implementasi model IndoBERT dalam menyelesaikan tugas klasifikasi token pada teks berbahasa Indonesia.

Bagi pengembang sistem dan praktisi teknologi informasi, penelitian ini dapat menjadi dasar dalam pengembangan aplikasi berbasis kecerdasan buatan yang memanfaatkan teknologi ekstraksi informasi otomatis. Hasil penelitian dapat diterapkan pada berbagai sistem, seperti mesin pencari dokumen budaya, perpustakaan digital, sistem pengarsipan naskah kuno, maupun aplikasi edukasi yang memerlukan identifikasi tokoh, lokasi, dan organisasi secara otomatis dari sebuah teks.

Dari sisi pelestarian budaya, penelitian ini diharapkan dapat mendukung upaya digitalisasi dan dokumentasi warisan budaya Indonesia. Melalui penerapan teknologi Named Entity Recognition, informasi penting yang terkandung dalam cerita rakyat dapat diubah menjadi data terstruktur sehingga lebih mudah disimpan, dianalisis, dan dimanfaatkan untuk berbagai kebutuhan penelitian maupun pendidikan. Dengan demikian, pemanfaatan teknologi dapat membantu menjaga keberlangsungan dan aksesibilitas karya sastra klasik Indonesia di era digital.

DAFTAR PUSTAKA

- Adelani, David Ifeoluwa, Jade Abbott, Graham Neubig, D. Daniel, Julia Kreutzer, Constantine Lignos, Chester Palen-michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, dan Israel Abebe Azime. 2021. "MasakhaNER : Named Entity Recognition for African Languages." 9:1116–31.
- Aliff, Sulthan, Secca Ramadhani, Alifio Rahmanqa, Metode Named, dan Entity Recognition. 2021. "Jurnal Sosial dan Teknologi (SOSTECH) Deteksi Lokasi Siswa SMP di Instagram dengan Metode e-ISSN 2774-5155 Named Entity Recognition Sulthan Aliff Secca Ramadhani , Alifio Rahmanqa dan Nur Aini Rakhmawati 691." 1(7):690–96.
- Ashok, Dhananjay. 2022. "PromptNER : Prompting For Named Entity Recognition."
- Brandsen, Alex. 2025. "Can BERT Dig It ? Named Entity Recognition for Information Retrieval in the Archaeology Domain Can BERT Dig It ? Named Entity Recognition for Information." 15(3). doi:10.1145/3497842.
- Chen, Xiang, Ningyu Zhang, Lei Li, Yunzhi Yao, Shumin Deng, Chuanqi Tan, Fei Huang, Luo Si, dan Huajun Chen. 2022. "Good Visual Guidance Makes A Better Extractor : Hierarchical Visual Prefix for Multimodal Entity and Relation Extraction." 1607–18.
- Cotterell, Ryan, dan Kevin Duh. 2017. "Low-Resource Named Entity Recognition with Cross-Lingual , Character-Level Neural Conditional Random Fields." 91–96.
- Cui, Leyang, Yu Wu, Jian Liu, Sen Yang, dan Yue Zhang. 2021. "Template-Based Named Entity Recognition Using BART." 1835–45.
- Daniati, Erna, Aji Prasetya, Wahyu Sakti, Gunawan Irianto, dan Anusua Ghosh. 2024. "Analyzing event relationships in Andersen ' s Fairy Tales with BERT and Graph Convolutional Network (GCN)." 5(1):40–59.
- Daniati, Erna, Aji Prasetya, Wahyu Sakti, Gunawan Irianto, dan Leonel Hernandez. 2025. "Few-Shot-BERT-RNN Narrative Structure Analysis for Andersen ' s Stories." 9(July):1773–82.
- Daniati, Erna, Aji Prasetya, Wahyu Sakti, Gunawan Irianto, dan Andrew Nafalski. 2025. "Attention-Enhanced Convolutional Neural Network for Context Extraction in Andersen ' s Fairy Tales." 9(November):2359–65.
- Daniati, Erna, Aidina Ristyawan, Akmal Hisyam Pradhana, Dea Yuliana, Ayu Ningrum, Rina Firliana, dan Arie Nugroho. 2025. "Pengabdian Masyarakat di Unit Pelaksanaan

- Teknis Publikasi Ilmiah Universitas Negeri Malang : Analisis dan Rekomendasi Pengelolaan Jurnal Elektronik.” 6(1):86–106.
- Daniati, Erna, Hastari Utama, Sistem Informasi, Universitas Nusantara, Pgri Kediri, Teknik Informatika, Universitas Amikom, dan Kata Kunci. 2023. “ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER Abstraksi Keywords : Pendahuluan Tinjauan Pustaka.” 4(2).
- Ding, Ning, Guangwei Xu, Yulin Chen, Xiaobin Wang, Xu Han, Pengjun Xie, Hai-tao Zheng, dan Zhiyuan Liu. 2021. “F EW -NERD : A Few-shot Named Entity Recognition Dataset.” 3198–3213.
- Ferdiana, Selvia, Prasetyo Wibowo, Abid Famasya, dan Setio Basuki. 2024. “Named Entity Recognition in Medical Domain : A Systematic Literature Review.” 8(December).
- Firliana, Rina, Rini Indriati, Erna Daniati, Arie Nugroho, Ratih Kumalasari, Theo Yan, A. Pendahuluan, Implementasi Metode, Sistem Pengambilan, dan Keputusan Pemilihan. 2025. “Pemilihan menu di kantin sesuai minat mahasiswa dengan menggunakan metode profile matching.” 8(2).
- Fu, Jinlan, Xuanjing Huang, dan Pengfei Liu. 2020. : “: Named Entity Re -/Recognition as Span Prediction.”
- Huang, Yucheng, Kai He, Yige Wang, Xianli Zhang, Tieliang Gong, Rui Mao, dan Chen Li. 2022. “COPNER : Contrastive Learning with Prompt Guiding for Few-shot Named Entity Recognition.” 2515–27.
- Jarrar, Mustafa, Mohammed Khalilia, dan Sana Ghanem. 2022. “Wojood : Nested Arabic Named Entity Corpus and Recognition using.” (June):3626–36.
- Kuang, Zhanghui, Kai Chen, Wayne Zhang, dan Dahua Lin. 2021. *MMOCR : A Comprehensive Toolbox for Text Detection , Recognition and Understanding*. Vol. 1. Association for Computing Machinery.
- Le, Nicolas, dan Guillaume Wilfried. 2022. “TaxoNERD : Deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature.” 2022(April 2021):625–41. doi:10.1111/2041-210X.13778.
- Luqyana, Wanda Athira, Imam Cholissodin, dan Rizal Setya Perdana. 2018. “Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine.” 2(11):4704–13.
- Pradhana, Akmal Hisyam, Erna Daniati, dan Muhammad Najibulloh. 2025. “Penerapan Bi-LSTM Untuk Named Entity Recognition Pada Teks Bahasa Indonesia.” 4(2):96–106.

- Susanty, Meredita. 2021. "Perbandingan Pre-trained Word Embedding dan Embedding Layer untuk Named-Entity Recognition Bahasa Indonesia." 14(2):247–57.
- Tang, Ruixiang, Xiaotian Han, Xiaoqian Jiang, dan Xia Hu. n.d. "Does Synthetic Data Generation of LLMs Help Clinical Text Mining? 1."
- Tholib, Abu, M. Noer Fadli Hidayat, Resty Wulanningrum, dan Erna Daniati. 2023. "Comparison of C4 . 5 and Naive Bayes for Predicting Student Graduation Using Machine Learning Algorithms." 2(2):71–78. doi:10.30812/IJECSA.v2i2.3364.
- Ushio, Asahi, dan Jose Camacho-collados. 2021. "T-NER : An All-Round Python Library for Transformer-based Named Entity Recognition." 53–62.
- Wang, Y. U. 2025. "Nested Named Entity Recognition : A Survey Nested Named Entity Recognition : A Survey." 16(6). doi:10.1145/3522593.
- Zaratiana, Urchade, Nadi Tomeh, Pierre Holat, dan Thierry Charnois. 2023. "Bidirectional Transformer."
- Zhu, Enwei, Jinpeng Li, Hwamei Hospital, dan Health Industry. 2019. "Boundary Smoothing for Named Entity Recognition."