

Turnitin LLCC

Danda Bagaskoro

 SI-16

Document Details

Submission ID**trn:oid::2945:393392717****Submission Date****Jun 15, 2026, 2:50 PM GMT+7****Download Date****Jun 15, 2026, 2:52 PM GMT+7****File Name****Danda Bagaskoro.docx****File Size****1.3 MB****42 Pages****8,389 Words****56,243 Characters**

29% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Top Sources

- 16%  Internet sources
- 13%  Publications
- 25%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 16% Internet sources
- 13% Publications
- 25% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	Universitas Islam Indonesia on 2026-04-20	1%
2	Internet	docplayer.info	1%
3	Internet	ojs.unpkediri.ac.id	1%
4	Student papers	Telkom University on 2026-06-08	<1%
5	Student papers	Universitas Muhammadiyah Makassar on 2026-06-07	<1%
6	Student papers	Universitas Islam Indonesia on 2026-05-12	<1%
7	Internet	repository.umsu.ac.id	<1%
8	Student papers	Sultan Agung Islamic University on 2026-02-16	<1%
9	Student papers	Universitas Islam Sumatera Utara on 2026-04-20	<1%
10	Student papers	Universitas Sumatera Utara on 2026-06-04	<1%
11	Student papers	UPN Veteran Yogyakarta on 2026-05-04	<1%

12	Student papers	Universitas Muhammadiyah Makassar on 2026-05-11	<1%
13	Internet	repository.unissula.ac.id	<1%
14	Student papers	FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM on 2026-04-28	<1%
15	Student papers	UPN Veteran Yogyakarta on 2026-06-02	<1%
16	Student papers	Universitas Pamulang on 2026-06-09	<1%
17	Internet	cdn.juris.id	<1%
18	Student papers	Universitas Singaperbangsa Karawang on 2025-12-14	<1%
19	Student papers	Universitas Tarumanagara on 2025-12-12	<1%
20	Student papers	Universitas Amikom on 2024-10-03	<1%
21	Internet	repository.unpkediri.ac.id	<1%
22	Student papers	Universitas Pendidikan Indonesia on 2025-03-13	<1%
23	Student papers	UPN Veteran Yogyakarta on 2026-01-05	<1%
24	Student papers	Universitas Tarumanagara on 2025-12-23	<1%
25	Publication	Devi Fajar Wati, Lila Setiyani, Deden Moch Alfiansyah, M. Jembar Jomantara, De...	<1%

26	Student papers	Universitas Sanata Dharma on 2026-06-09	<1%
27	Internet	repository.umy.ac.id:8080	<1%
28	Publication	Ichsan Mursidah, Remi Sanjaya, Bambang Yulianto, Dhian Sweetania, Puji Sularsi...	<1%
29	Student papers	Universitas Islam Riau on 2026-04-30	<1%
30	Student papers	Universitas Pamulang on 2026-05-02	<1%
31	Internet	sotaro.io	<1%
32	Publication	Muhammad Noval, M Afief Anugrah, Faiz Arrafi, Ridho Ramadhan, Marcel Wan...	<1%
33	Student papers	Sriwijaya University on 2020-12-01	<1%
34	Student papers	Telkom University on 2025-07-04	<1%
35	Student papers	UPN Veteran Yogyakarta on 2025-10-01	<1%
36	Student papers	FAKULTAS ILMU KOMPUTER on 2026-01-12	<1%
37	Student papers	Telkom University on 2025-07-07	<1%
38	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
39	Internet	preview.aclanthology.org	<1%

40	Internet	www.cnblogs.com	<1%
41	Internet	www.mubasharaakhtar.com	<1%
42	Student papers	Universitas Islam Indonesia on 2025-09-17	<1%
43	Student papers	Universitas Tarumanagara on 2025-12-21	<1%
44	Student papers	Politeknik Statistika STIS on 2025-07-05	<1%
45	Student papers	Universitas Indonesia on 2026-06-12	<1%
46	Internet	www.biorxiv.org	<1%
47	Student papers	Associatie K.U.Leuven on 2023-06-09	<1%
48	Publication	Erna Daniati, Aidina Ristyan, Akmal Hisyam Pradhana, Dea Yuliana Ayu Ningru...	<1%
49	Publication	Muhammad Jomantara, Lila Setiyani, Deden Moch Alfiansyah, Devi Fajar Wati, ...	<1%
50	Internet	core.ac.uk	<1%
51	Internet	jurnal.ciptamediaharmoni.id	<1%
52	Internet	lp3m.unuja.ac.id	<1%
53	Internet	pubs2.ascee.org	<1%

54	Internet	researchr.org	<1%
55	Publication	Yudi Wibisono, Masayu Leylia Khodra. "Pengenalan Entitas Bernama Otomatis un...	<1%
56	Internet	opus.bibliothek.uni-augsburg.de	<1%
57	Internet	repository.uinsaizu.ac.id	<1%
58	Internet	www.neliti.com	<1%
59	Student papers	Politeknik Ilmu Pelayaran Semarang on 2026-05-18	<1%
60	Publication	Timotius Victory, Dwi Diana Wazaumi. "Analisis Sentimen Ulasan Pengguna pada ...	<1%
61	Student papers	UPN Veteran Yogyakarta on 2026-05-25	<1%
62	Student papers	Universitas Islam Riau on 2026-01-17	<1%
63	Student papers	Universitas Islam Riau on 2026-03-26	<1%
64	Student papers	Universitas Sumatera Utara on 2026-06-08	<1%
65	Internet	blog.csdn.net	<1%
66	Internet	scholar.archive.org	<1%
67	Publication	Abdurroziq, Ali. "Analisis Model Inquiry Based Learning (IBL) Dalam Pembelajaran...	<1%

68	Student papers	University of Wales Swansea on 2022-05-02	<1%
69	Internet	export.arxiv.org	<1%
70	Internet	ruj.uj.edu.pl	<1%
71	Student papers	Tarumanagara University on 2020-11-08	<1%
72	Internet	ia902301.us.archive.org	<1%
73	Internet	journal.irpi.or.id	<1%
74	Internet	jurnal.amikom.ac.id	<1%
75	Internet	raretoonindia.org	<1%
76	Internet	sostech.greenvest.co.id	<1%
77	Internet	www.ilmudata.org	<1%
78	Internet	123dok.com	<1%
79	Publication	Fidalina Nirigi, Mochammad Triyanto, Mohammad Rezza Pahlevi, Athia Saelan, Fa...	<1%
80	Publication	Meysam Asgari-Chenaghlu, M. Reza Feizi-Derakhshi, Leili Farzinvash, M. A. Balafa...	<1%
81	Student papers	UPN Veteran Yogyakarta on 2026-04-24	<1%

82	Internet	bibbase.org	<1%
83	Publication	Agustin Agustin, Junadhi Junadhi, Fransiskus Zoromi, Parlindungan Kudadiri. "An...	<1%
84	Student papers	Sultan Agung Islamic University on 2025-12-05	<1%
85	Student papers	UPN Veteran Yogyakarta on 2025-10-28	<1%
86	Student papers	Universitas Muhammadiyah Sidoarjo on 2024-07-02	<1%
87	Student papers	Universitas Pamulang on 2026-06-08	<1%
88	Internet	cdn.techscience.cn	<1%
89	Internet	digilib.unimed.ac.id	<1%
90	Internet	ejurnal.methodist.ac.id	<1%
91	Internet	eprints.untirta.ac.id	<1%
92	Internet	joiv.org	<1%
93	Internet	jtiik.ub.ac.id	<1%
94	Internet	lib.unnes.ac.id	<1%
95	Publication	Erika Adelia Pratiwi, Sita Muharni. "PENERAPAN ALGORITMA RANDOM FOREST U...	<1%

96	Publication	Kiki Putriani Siregar, Budi Triandi, Roslina Roslina. "Perbandingan Algoritma Mac...	<1%
97	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2025-08-26	<1%
98	Student papers	LL Dikti IX Turnitin Consortium on 2025-09-30	<1%
99	Student papers	Perpustakaan on 2025-08-06	<1%
100	Student papers	UM Surabaya on 2025-07-16	<1%
101	Student papers	UPN Veteran Yogyakarta on 2026-06-04	<1%
102	Student papers	Universitas Bengkulu on 2025-10-07	<1%
103	Student papers	Universitas Islam Negeri Alauddin Makassar on 2026-05-13	<1%
104	Student papers	Universitas Merdeka Malang on 2026-04-17	<1%
105	Student papers	Universitas Muhammadiyah Purwokerto on 2024-12-23	<1%
106	Student papers	Universitas Muslim Indonesia on 2024-12-09	<1%
107	Student papers	Universitas Negeri Surabaya The State University of Surabaya on 2026-06-03	<1%
108	Student papers	Universitas Pamulang on 2025-07-12	<1%
109	Student papers	Universitas Tarumanagara on 2025-12-21	<1%

110	Internet	digilibadmin.unismuh.ac.id	<1%
111	Internet	es.scribd.com	<1%
112	Internet	hal-lirmm.ccsd.cnrs.fr	<1%
113	Internet	id.123dok.com	<1%
114	Internet	repository.radenintan.ac.id	<1%
115	Student papers	Carnegie Mellon University on 2023-11-01	<1%
116	Student papers	College of Banking and Financial Studies on 2026-03-04	<1%
117	Publication	Ilham Ismail, Fauzi Nur Maqi, Nur Hidayah. "PC Component Classification Using C...	<1%
118	Publication	Rahandya Cahyo, Alda Cendekia Siregar, Barry Ceasar Octariadi. "Implementasi A...	<1%
119	Publication	Riko Angga Bayu Kusuma -, Bambang Irawan -, Abdul Khamid -. "KLASIFIKASI PEN...	<1%
120	Student papers	STT PLN on 2024-08-29	<1%
121	Student papers	UPN Veteran Yogyakarta on 2026-02-27	<1%
122	Student papers	Universitas Negeri Semarang on 2026-05-11	<1%
123	Student papers	Universitas Pamulang on 2026-05-12	<1%

124	Student papers	Universitas Tarumanagara on 2025-12-09	<1%
125	Internet	docobook.com	<1%
126	Internet	docplayer.net	<1%
127	Internet	ejournal.um-sorong.ac.id	<1%
128	Internet	kc.umn.ac.id	<1%
129	Internet	repository.unmuhjember.ac.id	<1%
130	Internet	repository.upi.edu	<1%
131	Internet	repository.upstegal.ac.id	<1%
132	Internet	subset.id	<1%
133	Internet	www.scribd.com	<1%
134	Student papers	Sriwijaya University on 2020-07-20	<1%
135	Student papers	Sriwijaya University on 2025-12-17	<1%
136	Student papers	UPN Veteran Yogyakarta on 2026-06-02	<1%
137	Student papers	Universitas Islam Riau on 2026-02-27	<1%

138	Student papers	Universitas Islam Riau on 2026-04-28	<1%
139	Student papers	Universitas Muhammadiyah Sukabumi on 2026-01-20	<1%
140	Student papers	Universitas Tarumanagara on 2026-01-04	<1%
141	Student papers	Abdullah Gul University on 2025-11-20	<1%
142	Publication	Ahmad Mortadi, Waleed Nazih, Mohamed I. Eldesouki, Yasser Hifny. "Intelligent ...	<1%
143	Publication	Junhui Yu, Yanping Chen, Qinghua Zheng, Yuefei Wu, Ping Chen. "Full-span name...	<1%
144	Publication	Nirwan Yakub, Leni Marlina, Muhammad Iqbal, Andysah Putera Utama Siahaan, ...	<1%
145	Student papers	Politeknik Pelayaran Surabaya on 2025-07-18	<1%
146	Publication	Saikin Saikin, Mohammad Taufan Asri Zaen, Sofiansyah Fadli, Hairul Fahmi. "Pene...	<1%
147	Student papers	UNIVERSITAS BUDI LUHUR on 2025-07-23	<1%
148	Student papers	Universitas Dian Nuswantoro on 2025-07-04	<1%
149	Student papers	Universitas Islam Riau on 2025-11-12	<1%
150	Student papers	Universitas Putera Batam on 2025-07-25	<1%

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.

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.

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.

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.

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.

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.

BAB II

KAJIAN PUSTAKA

A. Kajian Teori

1. *Natural Language Processing (NLP)*

Natural Language Processing (NLP) merupakan cabang ilmu yang berada di persimpangan linguistik, ilmu komputer, dan kecerdasan buatan, yang berfokus pada interaksi antara komputer dan bahasa manusia (Daniati et al. 2024). Tujuan utama NLP adalah memungkinkan komputer memahami, menafsirkan, dan menghasilkan bahasa manusia secara bermakna. Perkembangan NLP telah mengalami evolusi signifikan dari pendekatan berbasis aturan (*rule-based*) dan statistik, hingga era *deep learning* yang didominasi oleh model berbasis neural network. Dalam konteks bahasa Indonesia, NLP menghadapi tantangan unik seperti morfologi yang kompleks (awalan, akhiran, sisipan), ragam dialek, dan pengaruh multibahasa, yang memerlukan pendekatan dan model yang khusus dikembangkan atau diadaptasi.

2. *Named Entity Recognition (NER)*

Named Entity Recognition (NER) adalah salah satu tugas fundamental dalam NLP yang bertujuan untuk mengidentifikasi dan mengklasifikasikan *mentions* atau sebutan entitas bernama dalam teks ke dalam kategori yang telah ditentukan.

- a. Definisi: Entitas bernama (*named entity*) merujuk pada kata atau frasa yang secara unik mengidentifikasi suatu objek dunia nyata, seperti orang, organisasi, lokasi, kuantitas, nilai moneter, persentase, dan tanggal.
- b. Kategori Entitas: Kategori umum meliputi:
 - 1) *PERSON (PER)*: Nama individu, tokoh, atau karakter.
 - 2) *LOCATION (LOC)*: Nama tempat geografis seperti kota, negara, gunung, sungai, atau tempat mitologis.
 - 3) *ORGANIZATION (ORG)*: Nama lembaga, perusahaan, kelompok, atau kerajaan.
 - 4) *MISC (Miscellaneous)*: Berbagai entitas lain seperti nama produk, acara, atau dalam konteks penelitian ini.
 - 5) *CULTURAL-TERM (CULT)* untuk istilah budaya, senjata, pusaka, atau konsep adat.

c. Skema Anotasi BIO/IOB2: Untuk melatih model NER, teks perlu diberi label dengan skema standar. Skema BIO membedakan:

- 1) **B**-{LABEL} (*Begin*): Token awal dari sebuah entitas.
- 2) **I**-{LABEL} (*Inside*): Token lanjutan dari sebuah entitas.
- 3) **O** (*Outside*): Token yang bukan bagian dari entitas apa pun.
Contoh: "[B-PER Arjuna] [O membawa] [B-CULT Panah] [I-CULT Pasopati]"

3. Arsitektur *Transformer* dan Model *BERT*

Revolusi dalam NLP terjadi dengan diperkenalkannya arsitektur Transformer oleh Transformer mengandalkan mekanisme Self-Attention sepenuhnya, menghilangkan kebutuhan akan recurrent atau convolutional layers.

- a. *Self-Attention*: Mekanisme ini memungkinkan model untuk menimbang dan menghubungkan setiap kata dalam sebuah kalimat dengan semua kata lainnya, tanpa memperhatikan jaraknya. Hal ini menghasilkan pemahaman kontekstual yang sangat baik, karena makna sebuah kata ditentukan oleh seluruh konteks kalimat, bukan hanya kata sebelum atau sesudahnya.
- b. *Transformers*: Dikembangkan oleh Google AI pada 2018, BERT adalah model *pretrained* yang memanfaatkan encoder dari arsitektur Transformer. Keunggulan utamanya adalah sifatnya yang *bidirectional*; selama pelatihan, model mempelajari konteks kata dari kiri dan kanan secara bersamaan. BERT dilatih pada dua tugas tanpa pengawasan (*unsupervised*): *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). Model ini kemudian dapat di-*fine-tune* dengan data berlabel yang relatif sedikit untuk berbagai tugas turunan (*downstream tasks*) seperti klasifikasi teks, NER, dan *question answering*.

4. IndoBERT

IndoBERT adalah implementasi model BERT yang secara khusus dilatih (*pretrained*) menggunakan korpus besar berbahasa Indonesia. Korpus ini dapat mencakup data dari Wikipedia Indonesia, berita online, dan sumber teks formal lainnya. Keberadaan IndoBERT sangat penting karena:

- a. Bahasa-Specific: Kata-kata dan konteks bahasa Indonesia memiliki karakteristik yang berbeda dengan bahasa Inggris, tempat model BERT asli dilatih.
- b. Kinerja Unggul: IndoBERT telah menunjukkan performa yang lebih baik dibandingkan model multibahasa (seperti mBERT) atau model BERT bahasa Inggris yang di-*fine-tune* untuk tugas-tugas NLP berbahasa Indonesia.
- c. Ketersediaan dan Kemudahan: IndoBERT tersedia secara *open-source* melalui platform seperti HuggingFace Model Hub, memudahkan peneliti dan pengembang untuk mengadopsinya.

5. Teks Cerita Rakyat dan Sastra Klasik Indonesia

a. Singkatan Cerita

Hikayat Wayang Arjuna dan Purusara merupakan salah satu karya sastra klasik berbasis tradisi pewayangan Jawa yang mengisahkan perjalanan hidup, konflik, dan kesaktian tokoh Arjuna (Rajuna) sebagai bagian dari keluarga Pandawa. Cerita ini berlatar pada beberapa kerajaan dan wilayah mitologis, seperti Ngamarta, Ngastina, Alengka, dan Suralaya, serta melibatkan berbagai lapisan tokoh, mulai dari raja, ksatria, pendeta, panakawan, hingga para batara. Kompleksitas alur cerita mencerminkan interaksi antara dunia manusia dan dunia kahyangan yang menjadi ciri khas sastra pewayangan.

Secara naratif, hikayat ini menggambarkan rangkaian peristiwa konflik internal keluarga, kecemburuan, kesombongan, peperangan, kematian, serta kebangkitan kembali tokoh utama melalui unsur-unsur magis dan adikodrati. Tokoh Arjuna diceritakan mengalami beberapa kali pemenggalan kepala dan kebangkitan kembali yang berdampak pada kekacauan kosmis, baik di alam manusia maupun di alam kahyangan. Peristiwa-peristiwa tersebut tidak hanya berfungsi sebagai penggerak cerita, tetapi juga mengandung simbolisme budaya dan nilai filosofis yang kuat dalam tradisi Jawa.

Dari perspektif kajian bahasa dan *Natural Language Processing* (NLP), *Hikayat Wayang Arjuna dan Purusara* memiliki karakteristik

bahasa klasik yang khas, seperti penggunaan kosakata arkais, variasi ejaan, struktur kalimat panjang, serta pemakaian gelar dan julukan yang beragam untuk satu tokoh yang sama. Misalnya, satu entitas tokoh dapat disebut dengan beberapa nama atau gelar yang berbeda sesuai konteks cerita. Kondisi ini menimbulkan tantangan tersendiri dalam proses identifikasi entitas secara otomatis.

Hikayat ini juga mengandung kepadatan entitas bernama yang tinggi. Entitas tersebut dapat dikelompokkan ke dalam beberapa kategori utama, yaitu entitas tokoh (Person) seperti Arjuna, Ki Darmawangsa, Prabu Jenggala, Pendeta Dorna, dan Lurah Semar; entitas lokasi (Location) seperti Ngamarta, Ngastina, dan Suralaya; serta entitas istilah budaya (Cultural-Term) yang mencakup nama pusaka, ajaran spiritual, konsep kesaktian, sumpah (nadar), dan praktik ritual khas pewayangan. Keberadaan entitas budaya ini menjadikan teks hikayat berbeda dari teks modern pada umumnya.

Berdasarkan karakteristik tersebut, *Hikayat Wayang Arjuna dan Purusara* dipandang relevan sebagai objek penelitian *Named Entity Recognition* (NER), khususnya untuk menguji kemampuan model berbasis *transformer* seperti IndoBERT dalam mengenali dan mengklasifikasikan entitas pada teks sastra klasik. Analisis NER pada hikayat ini diharapkan mampu menghasilkan representasi data terstruktur yang tidak hanya memuat informasi tokoh dan lokasi, tetapi juga merekam kekayaan istilah budaya yang terkandung di dalamnya.

Dengan demikian, pemilihan *Hikayat Wayang Arjuna dan Purusara* sebagai objek penelitian mendukung tujuan kajian ini, yaitu mengevaluasi efektivitas IndoBERT dalam menangani teks non-kontemporer sekaligus berkontribusi pada upaya digitalisasi dan pelestarian sastra klasik Indonesia berbasis teknologi NLP.

- b. Definisi dan Karakteristik: Cerita rakyat (*folklore*) dan sastra klasik (seperti hikayat, babad, suluk) adalah warisan literer yang mengandung nilai sejarah, budaya, filosofis, dan religius. Ciri khasnya meliputi:

- 1) Bahasa yang Arkais dan Figuratif: Menggunakan kosakata dan struktur kalimat yang berbeda dari bahasa sehari-hari modern.
- 2) Banyak Entitas Budaya: Dipenuhi dengan nama tokoh wayang, dewa-dewi, kerajaan mitologis, senjata pusaka, dan istilah ajaran.
- 3) Struktur Naratif yang Khas: Sering mengikuti pola tertentu, seperti perjalanan sang pahlawan (*hero's journey*).

c. Tantangan NLP pada Teks Klasik:

- 1) *Out-of-Vocabulary (OOV) Words*: Banyak kata dalam teks klasik tidak terdapat dalam kamus atau korpus modern yang digunakan untuk melatih model *pretrained*.
- 2) Variasi Ejaan dan Bentuk Kata: Tidak adanya standar ejaan pada masa lalu menyebabkan banyak variasi penulisan untuk kata yang sama.
- 3) Ambiguitas Semantik: Makna kata dalam konteks klasik bisa sangat berbeda dengan makna modernnya.

B. Kajian Hasil Penelitian Terdahulu

Tabel 2.1 Kajian Hasil Penelitian Terdahulu

NO	JUDUL & NAMA PENULIS	PERSAMAAN	PERBEDAAN	KETERANGAN
1	Perbandingan Pre-trained Word Embedding dan Embedding Layer untuk Named-Entity Recognition Bahasa Indonesia(Susanty 2021)	Fokus NER bahasa Indonesia, penggunaan embedding pre-trained	Tidak pakai Transformer seperti IndoBERT, domain teks umum bukan klasik	Sangat relevan sebagai baseline teknik embedding NER Indonesia sebelum era Transformer.
2	Few-Shot-BERT-RNN Narrative Structure Analysis for Andersen's Stories(Daniati, Prasetya, Sakti, Irianto, dan Hernandez 2025)	Pakai BERT untuk analisis naratif cerita, pendekatan few-shot untuk data terbatas	Cerita Andersen (Barat) bukan Indonesia, fokus struktur naratif bukan NER murni	Relevan tinggi karena menggunakan BERT untuk analisis naratif cerita dengan data terbatas.

NO	JUDUL & NAMA PENULIS	PERSAMAAN	PERBEDAAN	KETERANGAN
3	Wjood: Nested Arabic Named Entity Corpus and Recognition using BERT(Jarrar, Khalilia, dan Ghanem 2009)	Buat korpus NER untuk bahasa kompleks budaya, pakai BERT untuk nested NER	Bahasa Arab bukan Indonesia, fokus nested entities bukan flat NER	Inspirasi metodologis untuk buat korpus NER budaya Indonesia dan tangani entitas bersarang.
4	Good Visual Guidance Makes A Better Extractor: Hierarchical Visual Prefix for Multimodal Entity and Relation Extraction (Chen et al. 2022)	Ekstraksi entitas multimodal (teks+gambar), pendekatan transformer canggih	Multimodal bukan teks saja, tidak spesifik bahasa Indonesia atau teks klasik	Relevan futuristik jika penelitian dikembangkan ke analisis naskah bergambar/ilustrasi wayang.
5	Named Entity Recognition in Medical Domain: A systematic Literature Review (Ferdiana et al. 2024)	Tinjauan sistematis NER domain spesifik, relevan metodologi kajian literatur	Domain medis bukan budaya/sastra, review bukan penelitian eksperimental	Relevan metodologis sebagai contoh tinjauan literatur sistematis untuk NER domain-spesifik.
6	Deteksi Lokasi Siswa SMP di Instagram dengan Metode Named Entity Recognition(Aliff et al. 2021)	NER pada teks bahasa Indonesia informal (media sosial)	Domain media sosial bukan teks klasik, hanya fokus lokasi bukan multi-kategori	Contoh aplikasi NER Indonesia di domain non-formal, namun terbatas pada satu entitas.
7	Implementasi Principal Component Analysis (PCA) Untuk Pengenalan Wajah Manusia – Rina Firliana, Resty Wulanningrum, Wisnu Sasongko (2015)	Implementasi Principal Component Analysis (PCA) Untuk Pengenalan Wajah Manusia – Rina Firliana, Resty Wulanningrum, Wisnu Sasongko (2015)	Implementasi Principal Component Analysis (PCA) Untuk Pengenalan Wajah Manusia – Rina Firliana, Resty Wulanningrum, Wisnu Sasongko (2015)	Implementasi Principal Component Analysis (PCA) Untuk Pengenalan Wajah Manusia – Rina Firliana, Resty Wulanningrum, Wisnu Sasongko (2015)

NO	JUDUL & NAMA PENULIS	PERSAMAAN	PERBEDAAN	KETERANGAN
8	GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer (Zaratiana et al. 2023)	Model Transformer untuk NER generalis, bisa untuk berbagai domain	Model generalis bukan spesifik Indonesia, tidak di-tune untuk teks klasik	Model state-of-the-art yang bisa jadi pembanding dengan IndoBERT.
9	Boundary Smoothing for Named Entity Recognition (Zhu et al. 2019)	Teknik peningkatan akurasi batas entitas, bisa diadopsi untuk NER Indonesia	Tidak spesifik bahasa Indonesia, teknik tambahan bukan model utama	Teknik peningkatan yang bisa diintegrasikan ke pipeline NER Anda.
10	Template-Based Named Entity Recognition Using BART (Cui et al. 2021)	NER berbasis template dengan model BART, pendekatan alternatif untuk NER	BART bukan BERT, tidak spesifik untuk teks klasik/bahasa Indonesia	Alternatif pendekatan template-based yang mungkin cocok untuk pola entitas budaya.
11	MasakhaNER: Named Entity Recognition for African Languages (Adelani et al. 2021)	Pengembangan dataset NER untuk bahasa kurang sumber daya, inspirasi dataset NER Indonesia	Bahasa Afrika bukan Indonesia, fokus pembuatan dataset	Inspirasi kuat untuk pengembangan dataset NER bahasa daerah/klasik Indonesia.
12	TaxoNERD: Deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature (Le dan Wilfried 2022)	NER untuk entitas domain-spesifik (taksonomi), relevan untuk entitas budaya spesifik	Domain ekologi bukan budaya/sastra, bahasa Inggris	Analog domain-spesifik yang menunjukkan tantangan NER untuk entitas khusus.
13	Nested Named Entity Recognition: A Survey (Wang 2025)	Tinjauan komprehensif tentang nested NER, metodologi survei literatur	Survei umum, tidak spesifik bahasa Indonesia atau teks klasik	Relevan teoretis untuk memahami tantangan dan metode nested NER yang mungkin ada di teks wayang.

NO	JUDUL & NAMA PENULIS	PERSAMAAN	PERBEDAAN	KETERANGAN
14	Can BERT Dig It? Named Entity Recognition for Information Retrieval in the Archaeology Domain (Brandsen 2025)	NER domain-spesifik (arkeologi) dengan BERT, analog dengan teks klasik/sejarah	Domain arkeologi bukan sastra klasik Indonesia, bahasa Inggris/Eropa	Studi kasus domain historis yang mirip dengan teks klasik Indonesia.
15	Few-NERD: A Few-Shot Named Entity Recognition Dataset – Ning Ding, dkk. (2021)	Dataset untuk few-shot NER, relevan untuk data terbatas	Dataset umum bukan untuk Indonesia, tidak spesifik teks klasik	Benchmark few-shot NER yang bisa diadopsi metodologinya.
16	PromptNER: Prompting For Named Entity Recognition – Dhananjay Ashok, Zachary C. Lipton (2023)	NER dengan prompting LLM, tren terbaru dalam NER	Pakai LLM besar bukan fine-tuning BERT, tidak spesifik bahasa Indonesia	Pendekatan terkini yang bisa dieksplorasi untuk NER budaya Indonesia.
17	Pengabdian Masyarakat di Unit Pelaksanaan Teknis Publikasi Ilmiah Universitas Negeri Malang: Analisis dan Rekomendasi Pengelolaan Jurnal Elektronik	Sama-sama memanfaatkan teknologi informasi untuk menyelesaikan permasalahan dan menghasilkan informasi yang bermanfaat bagi pengguna.	Praktik Kerja Lapangan (PKL) dan Analisis UI/UX menggunakan Eight Golden Rules	Penelitian menghasilkan rekomendasi perbaikan UI/UX website jurnal elektronik serta memberikan kontribusi dalam pengelolaan jurnal melalui pembuatan issue, metadata, contributor, galley, dan publishing artikel.
18	COPNER: Contrastive Learning with Prompt Guiding for Few-shot Named Entity Recognition (Huang et al. 2022)	Few-shot NER dengan contrastive learning, teknik canggih untuk data terbatas	Tidak spesifik bahasa Indonesia, kompleks butuh komputasi tinggi	Teknik advanced untuk few-shot learning di NER.

NO	JUDUL & NAMA PENULIS	PERSAMAAN	PERBEDAAN	KETERANGAN
19	Does Synthetic Data Generation of LLMs Help Clinical Text Mining? (Tang et al. n.d.)	Generasi data sintetik dengan LLM untuk domain khusus, solusi data terbatas	Domain klinis bukan budaya, pakai LLM (GPT) bukan BERT	Solusi potensial untuk atasi keterbatasan data anotasi teks klasik.
20	T-NER: An All-Round Python Library for Transformer-based Named Entity Recognition (Ushio dan Camacho-collados 2021)	Library untuk NER berbasis Transformer, mempermudah implementasi	Tool umum bukan penelitian eksperimen, tidak spesifik Indonesia	Tool praktis yang bisa digunakan untuk implementasi cepat.
21	MMOCR: A Comprehensive Toolbox for Text Detection, Recognition and Understanding (Kuang et al. n.d.)	Toolbox untuk OCR dan pemahaman teks, relevan jika hikayat bentuk gambar/scan	Fokus OCR bukan NER murni, tool bukan kajian teori	Tool pendukung jika teks hikayat berasal dari scan naskah fisik.
22	Pemilihan Menu di Kantin Sesuai Minat Mahasiswa dengan Menggunakan Metode Profile Matching (Firliana et al. 2025)	Pemilihan Menu di Kantin Sesuai Minat Mahasiswa dengan Menggunakan Metode Profile Matching – Rina Firliana, Rini Indriati, Erna Daniati, Arie Nugroho, Ratih Kumalasari, Theo Yan Arie (2025)	Pemilihan Menu di Kantin Sesuai Minat Mahasiswa dengan Menggunakan Metode Profile Matching – Rina Firliana, Rini Indriati, Erna Daniati, Arie Nugroho, Ratih Kumalasari, Theo Yan Arie (2025)	Pemilihan Menu di Kantin Sesuai Minat Mahasiswa dengan Menggunakan Metode Profile Matching – Rina Firliana, Rini Indriati, Erna Daniati, Arie Nugroho, Ratih Kumalasari, Theo Yan Arie (2025)
23	Low-Resource Named Entity Recognition with Cross-Lingual, Character-	NER sumber daya rendah dengan transfer lintas	Character-level NCRF bukan Transformer,	Pendekatan low-resource yang relevan

NO	JUDUL & NAMA PENULIS	PERSAMAAN	PERBEDAAN	KETERANGAN
	Level Neural Conditional Random Fields (Cotterell dan Duh 2017)	bahasa, relevan data terbatas	tidak spesifik Indonesia	untuk bahasa daerah Indonesia.
24	SpanNER: Named Entity Re-/Recognition as Span Prediction (Fu, Huang, dan Liu 2020)	NER sebagai prediksi span, arsitektur alternatif untuk NER	Span-based bukan token-based, tidak spesifik bahasa Indonesia	Arsitektur alternatif yang mungkin lebih cocok untuk entitas panjang.
25	Comparison of C4.5 and Naive Bayes for Predicting Student Graduation Using Machine Learning Algorithms(Tholib et al. 2023)	Penelitian ML di konteks Indonesia, relevan secara kontekstual	Bukan NER atau NLP, domain pendidikan bukan teks	Kontekstual Indonesia, namun tidak langsung relevan dengan NER.
26	ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER(Daniati et al. 2023)	Pendekatan ensemble dapat diadaptasi untuk NER multi-domain	Fokus pada analisis sentimen, bukan NER	Memberikan wawasan tentang ensemble learning untuk tugas NLP lain.
27	Penerapan Bi-LSTM Untuk Named Entity Recognition Pada Teks Bahasa Indonesia(Pradhana, Daniati, dan Najibulloh 2025)	Menjadi pembandingan untuk model berbasis transformer	Menggunakan Bi-LSTM, bukan transformer	Studi ini menunjukkan bahwa model non-transformer masih kompetitif untuk NER bahasa Indonesia.
28	Analyzing event relationships in Andersen's Fairy Tales with BERT and Graph Convolutional Network (GCN) Erna Daniati et al. (2024)	Mendukung penggunaan model hybrid untuk analisis relasi entitas	Fokus pada analisis relasi, bukan NER	Memberikan contoh integrasi BERT dengan GCN untuk analisis naratif.

NO	JUDUL & NAMA PENULIS	PERSAMAAN	PERBEDAAN	KETERANGAN
29	Few-Shot-BERT-RNN Narrative Structure Analysis for Andersen's Stories(Daniati, Prasetya, Sakti, Irianto, dan Hernandez 2025)	Mendukung penggunaan BERT untuk analisis naratif	Few-shot learning, bukan full training	Menunjukkan efektivitas BERT dalam analisis naratif dengan data terbatas.
30	Attention-Enhanced Convolutional Neural Network for Context Extraction in Andersen's Fairy Tales(Daniati, Prasetya, Sakti, Irianto, dan Nafalski 2025)	Menunjukkan potensi model berbasis perhatian untuk teks naratif	Menggunakan CNN + Attention, bukan transformer murni	Memberikan alternatif arsitektur untuk ekstraksi konteks dalam teks naratif.

Berdasarkan Tabel 2.1, dapat diketahui bahwa sebagian besar penelitian terkait deteksi *cyberbullying* menggunakan algoritma *machine learning* seperti *Support Vector Machine (SVM)* dan *Naïve Bayes*. Metode tersebut banyak digunakan karena memiliki performa yang cukup baik dalam klasifikasi teks, terutama dengan dukungan metode ekstraksi fitur seperti TF-IDF.

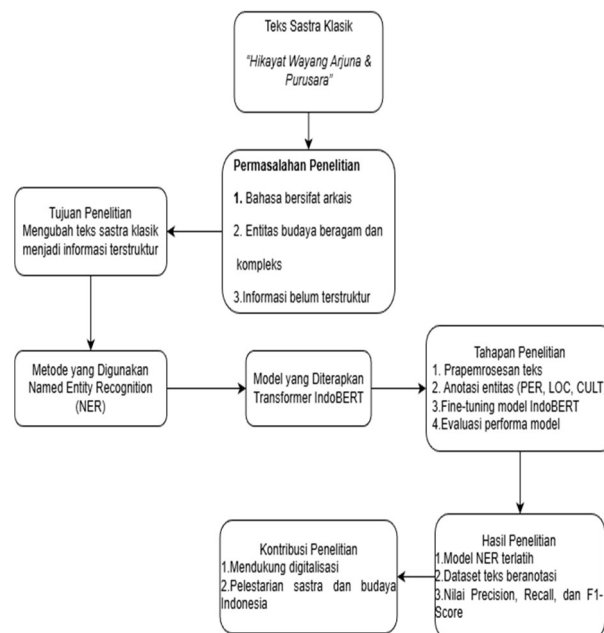
Beberapa penelitian menunjukkan bahwa algoritma SVM mampu menghasilkan akurasi yang tinggi dalam mendeteksi *cyberbullying*, seperti pada penelitian (Luqyana, Cholissodin, dan Perdana 2018) yang mencapai akurasi sebesar 90%. Selain itu, algoritma *Naïve Bayes* juga sering digunakan karena memiliki proses komputasi yang sederhana dan cepat, meskipun terkadang memiliki keterbatasan dalam menangani data yang kompleks.

Penggunaan algoritma Regresi Logistik dalam konteks deteksi *cyberbullying*, khususnya pada komentar Instagram berbahasa Indonesia, masih relatif terbatas. Padahal, Regresi Logistik merupakan algoritma yang efektif untuk klasifikasi biner serta memiliki interpretasi yang baik.

Penelitian ini dilakukan untuk mengeksplorasi penggunaan algoritma Regresi Logistik dengan metode *Bag of Words* dalam mendeteksi *cyberbullying* pada komentar Instagram berbahasa Indonesia. Penelitian ini diharapkan dapat memberikan alternatif metode yang sederhana namun tetap efektif dalam klasifikasi teks.

C. Kerangka Berpikir

Kerangka berpikir merupakan gambaran konseptual yang menjelaskan alur penelitian mulai dari tahap pengumpulan data hingga diperolehnya hasil penelitian. Kerangka berpikir pada penelitian ini disusun untuk memberikan gambaran mengenai proses penerapan metode *Named Entity Recognition* (NER) menggunakan model *IndoBERT* pada cerita rakyat Hikayat Wayang Arjuna dan Purusara.



Gambar 2.1 Kerangka Berpikir

Berdasarkan Gambar 2.1, penelitian diawali dengan pengumpulan data berupa teks cerita rakyat Hikayat Wayang Arjuna dan Purasara yang digunakan sebagai objek penelitian. Selain itu, penelitian juga memanfaatkan dataset NER Bahasa Indonesia yang telah memiliki anotasi entitas sebagai data pelatihan model. Tahap berikutnya adalah *preprocessing data*. Pada tahap ini dilakukan persiapan data agar sesuai dengan kebutuhan model pembelajaran mesin. Proses *preprocessing* meliputi pembacaan dataset, pemisahan token, pembersihan data yang tidak diperlukan, serta penyesuaian format data agar dapat diproses oleh model *IndoBERT*.

25 Tahap ini sangat penting karena kualitas data yang digunakan akan memengaruhi performa model yang dihasilkan.

37 Setelah proses preprocessing selesai, dilakukan pelabelan entitas menggunakan skema BIO (*Beginning, Inside, Outside*). Skema ini digunakan untuk memberikan informasi mengenai posisi suatu entitas dalam kalimat. Label yang digunakan dalam penelitian ini meliputi B-PER dan I-PER untuk entitas orang (*Person*), B-LOC dan I-LOC untuk entitas lokasi (*Location*), B-ORG dan I-ORG untuk entitas organisasi (*Organization*), serta label O untuk kata yang tidak termasuk entitas.

11 Tahap selanjutnya adalah proses pelatihan model menggunakan IndoBERT. Pada tahap ini dataset yang telah dipersiapkan akan diproses menggunakan tokenizer IndoBERT dan dilakukan *fine-tuning* agar model mampu mempelajari pola-pola entitas yang terdapat dalam data pelatihan. Model IndoBERT dipilih karena memiliki kemampuan memahami konteks bahasa Indonesia secara lebih baik dibandingkan metode tradisional, sehingga diharapkan mampu meningkatkan akurasi pengenalan entitas.

18 Setelah proses pelatihan selesai, model dievaluasi menggunakan data pengujian untuk mengetahui tingkat kinerjanya. Evaluasi dilakukan menggunakan metrik *Precision*, *Recall*, dan *F1-Score*. Ketiga metrik tersebut digunakan untuk mengukur kemampuan model dalam mengidentifikasi entitas secara tepat dan konsisten.

1 Tahap terakhir adalah menghasilkan keluaran berupa hasil *Named Entity Recognition* pada teks cerita rakyat Hikayat Wayang Arjuna dan Purasara. Hasil tersebut berupa identifikasi otomatis terhadap nama tokoh, nama tempat, dan nama organisasi yang terdapat dalam cerita.

BAB III

METODE PENELITIAN

A. Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen dalam bidang Natural Language Processing (NLP) dengan tujuan membangun model Named Entity Recognition (NER) menggunakan model IndoBERT untuk mengenali entitas pada cerita rakyat Indonesia, khususnya Hikayat Wayang Arjuna dan Purasara. Pendekatan eksperimen dipilih karena penelitian berfokus pada proses pelatihan, pengujian, dan evaluasi model pembelajaran mesin terhadap data teks yang telah diberi label.

Named Entity Recognition merupakan salah satu tugas penting dalam NLP yang bertujuan untuk mengidentifikasi dan mengklasifikasikan entitas tertentu pada suatu teks. Entitas tersebut dapat berupa nama orang (Person), nama lokasi (Location), maupun nama organisasi (Organization). Dalam penelitian ini, model IndoBERT digunakan karena merupakan model bahasa Indonesia berbasis Transformer yang telah dilatih sebelumnya menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga memiliki kemampuan memahami konteks bahasa yang lebih baik dibandingkan metode konvensional.

Tahapan penelitian dimulai dari pengumpulan dataset, persiapan data, pelabelan entitas, tokenisasi menggunakan tokenizer IndoBERT, proses fine-tuning model, evaluasi menggunakan metrik Precision, Recall, dan F1-Score, hingga pengujian model menggunakan kalimat-kalimat yang diambil dari cerita Hikayat Wayang Arjuna dan Purasara.

Penelitian ini menggunakan pendekatan eksperimen karena berfokus pada pengujian performa model IndoBERT dalam mengenali entitas pada teks berbahasa Indonesia. Pendekatan eksperimen memungkinkan peneliti untuk mengamati pengaruh proses pelatihan model terhadap hasil identifikasi entitas yang dihasilkan. Selain itu, penelitian ini bersifat kuantitatif karena hasil penelitian diukur menggunakan nilai Precision, Recall, dan F1-Score.

B. Objek Penelitian

Objek penelitian ini adalah teks cerita rakyat Indonesia yang berasal dari Hikayat Wayang Arjuna dan Purasara. Teks tersebut digunakan sebagai data pengujian untuk mengetahui kemampuan model dalam mengenali berbagai entitas yang terdapat dalam cerita pewayangan.

Entitas yang menjadi fokus penelitian meliputi:

1. Person (PER) Entitas *Person* merupakan nama tokoh atau individu yang muncul dalam cerita. Contohnya adalah Rajuna, Semar, Petruk, Batara Guru, Dewi Jembawati, dan tokoh lainnya.
2. Location (LOC) Entitas Location merupakan nama tempat atau wilayah yang muncul dalam cerita. Contohnya adalah Ngamarta, Ngastina, Astina, Wirata, Suralaya, dan Kayangan
3. Organization (ORG) Entitas *Organization* merupakan nama organisasi, kerajaan, kelompok, atau institusi yang terdapat dalam cerita.

C. Dataset Penelitian

Dataset yang digunakan pada penelitian ini berasal dari dataset NER Bahasa Indonesia yang disimpan dalam format CoNLL. Dataset terdiri atas tiga bagian yaitu data pelatihan (training), data validasi (validation), dan data pengujian (testing). Data training digunakan untuk melatih model agar mampu mempelajari pola entitas. Data validation digunakan untuk memantau performa model selama pelatihan dan mencegah overfitting. Sedangkan data testing digunakan untuk mengukur kemampuan model pada data yang belum pernah dilihat sebelumnya. Berdasarkan hasil pembacaan dataset menggunakan fungsi `read_conll()`, diperoleh jumlah data sebagai berikut:

Tabel 3.1 Dataset penelitian

Dataset	Jumlah Data
Training	1464
Validation	367
Testing	509

Dataset disimpan dalam bentuk token dan label BIO sehingga setiap kata memiliki label entitas masing-masing.

```
20 train_sents = read_conll("/content/drive/My Drive/Colab Notebooks/skripsi/idner-news-2k/train.txt")
21 valid_sents = read_conll("/content/drive/My Drive/Colab Notebooks/skripsi/idner-news-2k/valid.txt")
22 test_sents = read_conll("/content/drive/My Drive/Colab Notebooks/skripsi/idner-news-2k/test.txt")
```

Gambar 3.1 Pembacaan Dataset

Gambar 3.1 menunjukkan proses pembacaan dataset menggunakan fungsi `read_conll()`. Fungsi ini bertugas membaca file `train.txt`, `valid.txt`, dan `test.txt` kemudian memisahkan token dan label ke dalam struktur data yang dapat digunakan oleh model.

D. Pelabelan Entitas

Pelabelan data menggunakan skema BIO (Beginning, Inside, Outside). Skema ini merupakan metode pelabelan yang umum digunakan pada tugas Named Entity Recognition.

Label yang digunakan dalam penelitian ini adalah

Tabel 3.2 Pelabelan Entitas

Label	Keterangan
B-PER	Awal entitas Person
I-PER	Lanjutan entitas Person
B-LOC	Awal entitas Location
I-LOC	Lanjutan entitas Location
B-ORG	Awal entitas Organization
I-ORG	Lanjutan entitas Organization
O	Bukan entitas

Contoh pelabelan dapat dilihat pada Tabel berikut.

Tabel 3.3 Hasil Pelabelan

Token	Label
Rajuna	B-PER
pergi	O
ke	O
Ngamarta	B-LOC

Melalui skema BIO, model dapat mempelajari posisi awal dan lanjutan suatu entitas secara lebih akurat.

E. Arsitektur indoBERT

IndoBERT merupakan model bahasa berbasis arsitektur Transformer yang dikembangkan khusus untuk memahami bahasa Indonesia. Model ini merupakan adaptasi dari BERT (*Bidirectional Encoder Representations from Transformers*) yang telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar. Keunggulan IndoBERT terletak pada kemampuannya memahami konteks kata dari dua arah (*bidirectional*), yaitu dari kata sebelum dan sesudahnya secara bersamaan.

Pada penelitian ini, IndoBERT digunakan untuk melakukan tugas *Named Entity Recognition* (NER) dengan tujuan mengidentifikasi entitas berupa nama orang (*Person*), nama lokasi (*Location*), dan nama organisasi (*Organization*) yang terdapat dalam cerita rakyat Hikayat Wayang Arjuna dan Purasara.

F. Persiapan Data

Setelah dataset dibaca, dilakukan proses konversi label ke dalam bentuk numerik agar dapat diproses oleh model pembelajaran mesin.

Contoh hasil konversi label:

Tabel 3.4 Persiapan Data

Label	ID
B-LOC	0
B-ORG	1
B-PER	2
I-LOC	3
I-ORG	4
I-PER	5
O	6

Konversi dilakukan menggunakan dictionary `label2id` dan `id2label` sehingga proses pelatihan dapat dilakukan secara efisien.

G. Implementasi Model IndoBERT

Penelitian ini menggunakan model:

`indobenchmark/indobert-base-p1`

Model tersebut merupakan model bahasa Indonesia berbasis Transformer yang telah dilatih sebelumnya menggunakan korpus Bahasa Indonesia dalam jumlah besar.

Kode implementasi model:

```
1 model_ckpt = "indobenchmark/indobert-base-p1"
2 tokenizer = AutoTokenizer.from_pretrained(model_ckpt)
3
4 model = AutoModelForTokenClassification.from_pretrained(
5     model_ckpt,
6     num_labels=len(labels),
7     id2label=id2label,
8     label2id=label2id
9 )
```

Gambar 3.2 Implementasi IndoBERT

Model kemudian disesuaikan untuk tugas token classification dengan jumlah label sebanyak tujuh kelas.

H. Tokenisasi dan Penyelarasan Label

Tokenisasi merupakan proses mengubah teks menjadi token-token yang dapat dipahami oleh model. Dalam penelitian ini digunakan tokenizer bawaan IndoBERT. Selain tokenisasi, dilakukan juga proses alignment label agar setiap token hasil tokenisasi tetap memiliki label yang sesuai. Tahapan ini sangat penting karena satu kata dapat dipecah menjadi beberapa sub-token oleh tokenizer BERT.

```
1 def tokenize_and_align_labels(examples):
2     tokenized = tokenizer(
3         examples["tokens"],
4         is_split_into_words=True,
5         truncation=True,
6         padding="max_length", # <--- tambahkan padding agar semua sama panjang
7         max_length=128       # <--- panjang maksimum (sesuaikan kebutuhan)
8     )
9
10    labels_aligned = []
11    for i, label in enumerate(examples["ner_tags"]):
12        word_ids = tokenized.word_ids(batch_index=i)
13        aligned = []
14        prev = None
15        for w_id in word_ids:
16            if w_id is None:
17                aligned.append(-100) # token [CLS], [SEP], [PAD]
18            elif w_id != prev:
19                aligned.append(label[w_id]) # label token asli
20            else:
21                aligned.append(-100) # sub-token -> ignore index
22            prev = w_id
23        labels_aligned.append(aligned)
24
25    tokenized["labels"] = labels_aligned
26    return tokenized
```

Gambar 3.3 Proses Tokenisasi Pelatihan Model

Pelatihan model dilakukan menggunakan library Transformers dari Hugging Face. Parameter pelatihan yang digunakan adalah:

Tabel 3.5 Hasil Parameter

Parameter	Nilai
Epoch	40
Learning Rate	3e-5
Batch Size Train	8
Batch Size Validation	32
Weight Decay	0.01
Warmup Ratio	0.2

Tujuan pelatihan adalah memperbarui bobot model sehingga mampu mengenali pola entitas pada dataset yang digunakan

I. Evaluasi Model

Evaluasi model dilakukan menggunakan tiga metrik utama yaitu Precision, Recall, dan F1-Score.

Precision :

Precision digunakan untuk mengukur ketepatan prediksi entitas yang dihasilkan model.

$$Precision = \frac{TP}{TP + FP}$$

Recall :

Recall digunakan untuk mengukur kemampuan model menemukan seluruh entitas yang benar.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score :

F1-Score merupakan rata-rata harmonis antara Precision dan Recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Metrik tersebut dipilih karena sangat umum digunakan pada penelitian Named Entity Recognition.

J. Pengujian Model

Setelah proses pelatihan selesai, model diuji menggunakan 40 kalimat yang diambil dari cerita Hikayat Wayang Arjuna dan Purasara.

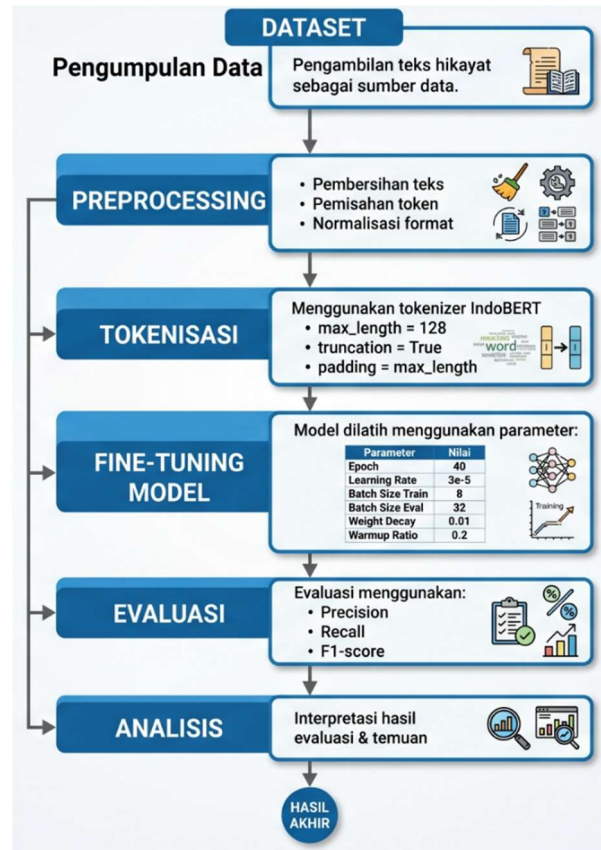
Contoh kalimat pengujian:

1. Rajuna mengadukan sikap Prabu Jenggala kepada Ki Darmawangsa di Negeri Ngamarta.
2. Dewi Jembawati datang menjenguk Rajuna yang sedang sakit di istana Ngamarta.
3. Batara Guru dan Batara Narada membicarakan kekacauan di Kayangan.

Tujuan pengujian adalah mengetahui kemampuan model dalam mengenali entitas Person, Location, dan Organization pada domain cerita rakyat.

K. Alur Penelitian

Alur penelitian merupakan rangkaian tahapan yang dilakukan untuk mencapai tujuan penelitian, yaitu membangun model *Named Entity Recognition* (NER) menggunakan IndoBERT pada cerita rakyat Hikayat Wayang Arjuna dan Purasara. Alur penelitian ini dimulai dari proses pengumpulan data hingga menghasilkan model yang mampu mengenali entitas secara otomatis pada teks berbahasa Indonesia.



Gambar 3.4 Alur Penelitian model NER menggunakan IndoBERT

Berdasarkan Gambar 3.4, tahapan penelitian yang dilakukan dalam penerapan *Named Entity Recognition* (NER) menggunakan IndoBERT pada cerita rakyat Hikayat Wayang Arjuna dan Purasara adalah sebagai berikut:

1. Pengumpulan Dataset

Tahap pertama adalah pengumpulan dataset yang digunakan dalam penelitian. Dataset yang digunakan merupakan dataset *Named Entity Recognition* (NER) Bahasa Indonesia yang telah memiliki anotasi entitas. Selain itu, peneliti juga mengumpulkan teks cerita rakyat Hikayat Wayang Arjuna dan Purasara yang digunakan sebagai data pengujian untuk mengetahui kemampuan model dalam mengenali entitas pada cerita rakyat Indonesia.

2. Pembacaan dan Persiapan Data

Pada tahap ini dataset dibaca menggunakan format CoNLL dan dipisahkan menjadi token serta label entitas. Data yang telah dibaca kemudian disusun ke dalam format yang sesuai untuk diproses oleh model pembelajaran mesin. Tahap

ini bertujuan untuk memastikan bahwa seluruh data dapat digunakan dengan baik pada proses pelatihan dan pengujian model.

3. Pelabelan Entitas (BIO Tagging)

Tahap pelabelan dilakukan menggunakan skema BIO (*Beginning, Inside, Outside*). Skema ini digunakan untuk memberikan informasi mengenai posisi suatu entitas dalam kalimat. Label yang digunakan meliputi B-PER dan I-PER untuk entitas orang, B-LOC dan I-LOC untuk entitas lokasi, B-ORG dan I-ORG untuk entitas organisasi, serta label O untuk kata yang bukan termasuk entitas.

4. Tokenisasi dan Alignment

Tahap tokenisasi dilakukan menggunakan tokenizer IndoBERT untuk mengubah teks menjadi token-token yang dapat dipahami oleh model. Setelah proses tokenisasi selesai, dilakukan penyelarasan label (*label alignment*) agar setiap token tetap memiliki label entitas yang sesuai. Tahap ini penting karena satu kata dapat dipecah menjadi beberapa sub-token oleh tokenizer IndoBERT.

5. Implementasi Model IndoBERT

Pada tahap ini dilakukan pemanggilan model IndoBERT menggunakan model *indobenchmark/indobert-base-pl*. Model tersebut digunakan sebagai dasar dalam proses *fine-tuning* untuk tugas *Named Entity Recognition*. IndoBERT dipilih karena memiliki kemampuan memahami konteks bahasa Indonesia dengan baik sehingga dapat meningkatkan kualitas pengenalan entitas.

6. Pelatihan Model (Fine-Tuning)

Tahap pelatihan dilakukan menggunakan data training yang telah dipersiapkan sebelumnya. Pada proses ini model mempelajari pola hubungan antara token dan label entitas yang terdapat pada dataset. Pelatihan dilakukan menggunakan parameter tertentu seperti jumlah epoch, learning rate, batch size, dan weight decay untuk menghasilkan performa model yang optimal.

7. Evaluasi Model

Setelah proses pelatihan selesai, model dievaluasi menggunakan data testing. Evaluasi dilakukan untuk mengetahui tingkat kemampuan model dalam mengenali entitas. Metrik yang digunakan meliputi *Precision*, *Recall*, dan *F1-Score*. Hasil evaluasi digunakan sebagai dasar untuk menilai keberhasilan model yang telah dibangun.

8. Pengujian pada Cerita Hikayat Wayang

Model yang telah dilatih kemudian diuji menggunakan kalimat-kalimat dari cerita Hikayat Wayang Arjuna dan Purasara. Pengujian ini bertujuan untuk mengetahui kemampuan model dalam mengenali entitas tokoh, lokasi, dan organisasi yang terdapat dalam cerita rakyat. Hasil pengujian akan menunjukkan sejauh mana model dapat diterapkan pada domain budaya dan sastra Indonesia.

9. Analisis Hasil

Tahap analisis dilakukan dengan mengamati hasil prediksi yang dihasilkan oleh model. Hasil tersebut dibandingkan dengan label yang diharapkan untuk mengetahui tingkat keberhasilan model dalam mengenali entitas. Selain itu, dilakukan analisis terhadap kelebihan dan kekurangan model berdasarkan hasil evaluasi dan pengujian yang telah diperoleh.

10. Kesimpulan

Tahap terakhir adalah penarikan kesimpulan berdasarkan seluruh proses penelitian yang telah dilakukan. Kesimpulan disusun berdasarkan hasil evaluasi model dan pengujian pada cerita rakyat. Selain itu, diberikan pula saran yang dapat digunakan sebagai acuan untuk pengembangan penelitian selanjutnya.

2

BAB IV HASIL DAN PEMBAHASAN

A. Gambaran Umum Hasil Penelitian

Penelitian ini bertujuan untuk mengimplementasikan metode Named Entity Recognition (NER) menggunakan model IndoBERT dalam mengenali entitas pada teks berbahasa Indonesia. Entitas yang dikenali meliputi Person (PER), Location (LOC), dan Organization (ORG). Setelah proses pelatihan model selesai dilakukan menggunakan dataset NER Bahasa Indonesia, model kemudian dievaluasi menggunakan data uji serta diuji kembali menggunakan kalimat-kalimat yang diambil dari cerita Hikayat Wayang Arjuna dan Purasara.

Tahapan penelitian meliputi proses pembacaan dataset, pelabelan entitas, tokenisasi menggunakan tokenizer IndoBERT, pelatihan model, evaluasi model, dan pengujian menggunakan data cerita rakyat. Seluruh proses dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka Transformers dari Hugging Face.

B. Lingkungan Pengujian

Penelitian ini dilakukan menggunakan lingkungan pengembangan berbasis *cloud computing* melalui Google Colab. Penggunaan Google Colab dipilih karena menyediakan fasilitas GPU yang dapat mempercepat proses pelatihan model *Deep Learning*, khususnya model berbasis Transformer seperti IndoBERT.

1. Perangkat Keras (Hardware)

Tabel 4.1 Perangkat Keras

Komponen	Spesifikasi
Processor	Intel Xeon Processor (Google Colab)
RAM	12 GB
GPU	NVIDIA Tesla T4
Storage	Google Drive

Berdasarkan Tabel 4.1, penelitian memanfaatkan sumber daya komputasi yang tersedia pada Google Colab sehingga proses pelatihan model dapat berjalan lebih cepat dibandingkan menggunakan komputer lokal.

6

139

4

29

10

84

2. Perangkat Lunak (Software)

Tabel 4.2 Perangkat Lunak

Software	Fungsi
Python	Bahasa pemrograman utama
Google Colab	Lingkungan pengembangan
PyTorch	Framework Deep Learning
Transformers	Implementasi IndoBERT
NumPy	Pengolahan data numerik
Pandas	Manipulasi dataset
Scikit-learn	Evaluasi model
Segeval	Evaluasi NER

Perangkat lunak tersebut digunakan untuk mendukung seluruh tahapan penelitian mulai dari persiapan data hingga evaluasi model.

C. Hasil Pembacaan Dataset

Tahap pertama dalam penelitian adalah membaca dataset yang telah disimpan dalam format CoNLL. Dataset tersebut terdiri atas tiga file yaitu train.txt, valid.txt, dan test.txt.

Kode yang digunakan untuk membaca dataset adalah sebagai berikut:

```
20 train_sents = read_conll("/content/drive/My Drive/Colab Notebooks/skripsi/idner-news-2k/train.txt")
21 valid_sents = read_conll("/content/drive/My Drive/Colab Notebooks/skripsi/idner-news-2k/valid.txt")
22 test_sents = read_conll("/content/drive/My Drive/Colab Notebooks/skripsi/idner-news-2k/test.txt")
```

Gambar 4.1 Pembacaan Dataset

Fungsi `read_conll()` bertugas membaca setiap baris data kemudian memisahkan token dan label sehingga dapat diproses oleh model.

Berdasarkan hasil pembacaan dataset diperoleh jumlah data sebagai berikut:

Tabel 4.3 Jumlah Data

Dataset	Jumlah Data
Training	1464
Validation	367
Testing	509

Jumlah data tersebut menunjukkan bahwa dataset yang digunakan telah dipisahkan menjadi tiga bagian sehingga proses pelatihan dan evaluasi dapat dilakukan secara objektif.

D. Hasil Pelabelan Entitas

Setelah dataset berhasil dibaca, dilakukan identifikasi seluruh label entitas yang terdapat pada dataset. Hasil pelabelan menunjukkan bahwa penelitian menggunakan tujuh label BIO yaitu:

Tabel 4.4 Hasil Pelabelan Entitas

Label
B-LOC
B-ORG
B-PER
I-LOC
I-ORG
I-PER
O

Hasil Label Entitas :

```
*** ['B-LOC', 'B-ORG', 'B-PER', 'I-LOC', 'I-ORG', 'I-PER', 'O']
{'B-LOC': 0, 'B-ORG': 1, 'B-PER': 2, 'I-LOC': 3, 'I-ORG': 4, 'I-PER': 5, 'O': 6}
{0: 'B-LOC', 1: 'B-ORG', 2: 'B-PER', 3: 'I-LOC', 4: 'I-ORG', 5: 'I-PER', 6: 'O'}
```

Gambar 4.2 Hasil Label Entitas

Label BIO digunakan agar model mampu membedakan awal dan lanjutan suatu entitas. Pendekatan ini terbukti efektif dalam berbagai penelitian NER karena membantu model memahami struktur entitas yang terdiri dari lebih dari satu kata. sebagai contoh:

E. Hasil Implementasi Model IndoBERT

Model yang digunakan dalam penelitian adalah IndoBERT Base P1.

Model dipanggil menggunakan kode berikut:

```
1 model_ckpt = "indobenchmark/indobert-base-p1"
```

Gambar 4.3 Implementasi Model IndoBERT

IndoBERT merupakan model bahasa Indonesia berbasis arsitektur Transformer yang telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar. Dengan kemampuan contextual embedding yang dimiliki, model dapat memahami hubungan antar kata berdasarkan konteks kalimat.

Keunggulan IndoBERT dibandingkan metode konvensional adalah kemampuannya memahami makna kata yang berbeda tergantung konteks kalimat.

F. Hasil Tokenisasi dan Alignment Label

Setelah model dimuat, dilakukan proses tokenisasi menggunakan tokenizer IndoBERT. Tahapan tokenisasi bertujuan mengubah kalimat menjadi token yang dapat diproses oleh model.

Pada penelitian ini digunakan fungsi:

```
labels_aligned = []
for i, label in enumerate(examples["ner_tags"]):
    word_ids = tokenized.word_ids(batch_index=i)
    aligned = []
    prev = None
    for w_id in word_ids:
        if w_id is None:
            aligned.append(-100)      # token [CLS], [SEP], [PAD]
        elif w_id != prev:
            aligned.append(label[w_id]) # label token asli
        else:
            aligned.append(-100)      # sub-token -> ignore index
        prev = w_id
    labels_aligned.append(aligned)

tokenized["labels"] = labels_aligned
return tokenized
```

Gambar 4.4 Proses Tokenisasi

Selain tokenisasi, dilakukan pula proses alignment label untuk memastikan setiap token hasil tokenisasi tetap memiliki label yang sesuai.

Tahapan ini sangat penting karena satu kata dapat dipecah menjadi beberapa sub-token oleh tokenizer BERT.

Sebagai contoh:

Jembawati

dapat dipecah menjadi:

Jem

##ba

##wati

Tanpa alignment label, model tidak dapat mengetahui label yang benar untuk setiap sub-token.

G. Hasil Pelatihan Model

Pelatihan dilakukan selama 40 epoch menggunakan learning rate sebesar $3e-5$.

Parameter pelatihan yang digunakan adalah

Tabel 4.5 Hasil Pelatihan Model

Parameter	Nilai
Epoch	40
Learning Rate	3e-5
Batch Size	8
Weight Decay	0.01

Hasil pelatihan menunjukkan adanya peningkatan performa model dari waktu ke waktu

Tabel berikut menunjukkan perkembangan nilai F1-Score selama proses pelatihan.

Tabel 4.6 Hasil Perkembangan F1-Score

Step	F1-Score
500	54,88%
1000	59,43%
2000	60,33%
5000	61,86%
6500	62,15%
7000	62,52%

	precision	recall	f1-score	support
LOC	0.6794	0.8641	0.7607	103
ORG	0.4431	0.4564	0.4496	401
PER	0.6250	0.6699	0.6467	515
micro avg	0.5630	0.6055	0.5835	1019
macro avg	0.5825	0.6634	0.6190	1019
weighted avg	0.5589	0.6055	0.5807	1019

Gambar 4.5 Hasil Pelatihan Model

Terlihat bahwa nilai F1-Score mengalami peningkatan secara bertahap hingga mencapai performa terbaik pada akhir pelatihan. Hal ini menunjukkan bahwa model berhasil mempelajari pola hubungan antara token dan label entitas.

H. Hasil Evluasi Model

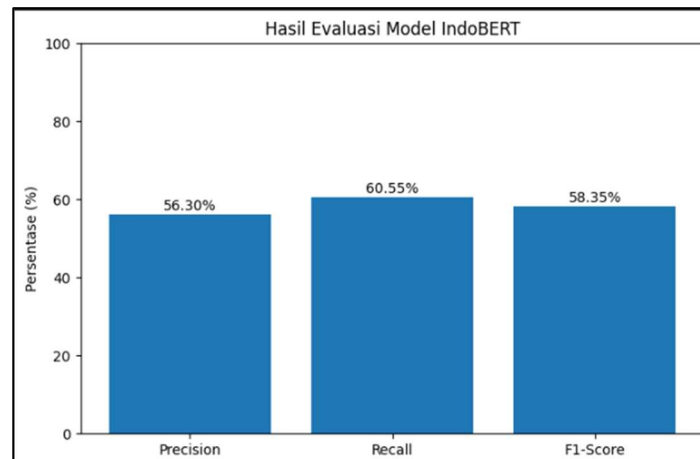
Setelah proses pelatihan selesai, model dievaluasi menggunakan data testing.

Hasil evaluasi diperoleh sebagai berikut:

Tabel 4.7 Hasil Evaluasi Model

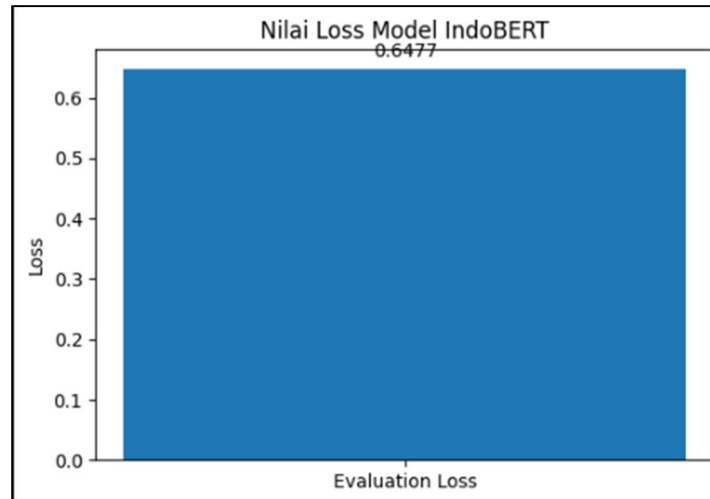
Metrik	Nilai
Precision	56,30%
Recall	60,55%
F1-Score	58,35%

Berdasarkan Tabel 4.3, model memperoleh nilai Precision sebesar 56,30%, Recall sebesar 60,55%, dan F1-Score sebesar 58,35%. Nilai tersebut menunjukkan bahwa model telah mampu melakukan identifikasi entitas pada teks cerita rakyat dengan tingkat performa yang cukup baik. Selain itu, nilai *loss* sebesar 0,6477 menunjukkan bahwa model berhasil mempelajari pola dalam dataset selama proses pelatihan.



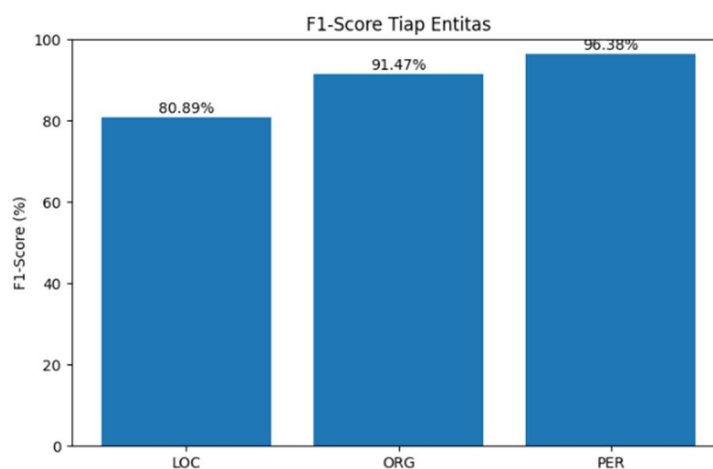
Gambar 4.6 Grafik Precision, Recall, dan F1-Score

Berdasarkan Gambar 4.2, nilai Recall sebesar 60,55% merupakan nilai tertinggi dibandingkan metrik lainnya. Hal ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam menemukan entitas yang terdapat dalam data pengujian. Sementara itu, nilai Precision sebesar 56,30% menunjukkan bahwa sebagian besar entitas yang diprediksi model merupakan prediksi yang benar. Nilai F1-Score sebesar 58,35% menunjukkan keseimbangan antara Precision dan Recall sehingga dapat digunakan sebagai indikator utama performa model dalam penelitian ini.



Gambar 4.7 Grafik Loss Model

Gambar 4.3 menunjukkan nilai *loss* yang diperoleh model setelah proses evaluasi. Nilai *loss* sebesar 0,6477 menunjukkan tingkat kesalahan model dalam melakukan prediksi terhadap data pengujian. Semakin kecil nilai *loss* yang diperoleh, maka semakin baik kemampuan model dalam memahami pola yang terdapat pada dataset. Hasil tersebut menunjukkan bahwa model IndoBERT telah mampu mempelajari hubungan antara token dan label entitas dengan cukup baik.



Gambar 4.8 Grafik F-1 Score Tiap Entitas

Berdasarkan Gambar 4.4, entitas PER (Person) memperoleh nilai F1-Score tertinggi yaitu sebesar 96,38%. Hasil tersebut menunjukkan bahwa model IndoBERT sangat baik dalam mengenali nama tokoh atau individu yang terdapat pada teks. Tingginya performa kategori PER menunjukkan bahwa model mampu memahami pola penulisan nama tokoh dengan sangat baik. Entitas ORG (Organization) memperoleh nilai F1-Score sebesar 91,47%. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali nama organisasi, kerajaan, maupun

kelompok yang terdapat dalam dataset. Tingginya nilai ORG menunjukkan bahwa data pelatihan yang digunakan telah mampu membantu model mempelajari karakteristik entitas organisasi. Sementara itu, entitas LOC (Location) memperoleh nilai F1-Score sebesar 80,89%. Meskipun lebih rendah dibandingkan kategori PER dan ORG, nilai tersebut masih tergolong baik. Perbedaan performa ini kemungkinan disebabkan oleh variasi penulisan nama lokasi yang lebih beragam sehingga model mengalami kesulitan dalam mengenali beberapa lokasi tertentu.

I. Analisis Per Kategori Entitas

Entitas Location (LOC)

Hasil evaluasi menunjukkan bahwa kategori lokasi memperoleh performa terbaik.

Tabel 4.8 Hasil Evaluasi

Metrik	Nilai
Precision	67,94%
Recall	86,41%
F1-Score	76,07%

Nilai tersebut menunjukkan bahwa model sangat baik dalam mengenali nama tempat.

Contoh lokasi yang berhasil dikenali:

1. Ngamarta
2. Ngastina
3. Astina
4. Wirata
5. Suralaya
6. Kayangan

Kemampuan ini muncul karena nama lokasi memiliki pola penulisan yang relatif konsisten sehingga lebih mudah dipelajari oleh model.

J. Entitas Person (PER)

Kategori Person memperoleh hasil sebagai berikut:

Tabel 4.9 Entitas Person

Metrik	Nilai
Precision	62,50%
Recall	66,99%
F1-Score	64,67%

Model cukup baik dalam mengenali nama tokoh seperti:

1. Rajuna
2. Semar
3. Petruk
4. Batara Guru
5. Dewi Jembawati
6. Batara Narada

Meskipun demikian, beberapa nama tokoh pewayangan masih muncul sebagai token [UNK] karena tidak terdapat pada kosakata yang dipelajari sebelumnya.

K. Entitas Organization (ORG)

Kategori organisasi memperoleh nilai terendah.

Tabel 4.10 Entitas Organization

Metrik	Nilai
Precision	44,31%
Recall	45,64%
F1-Score	44,96%

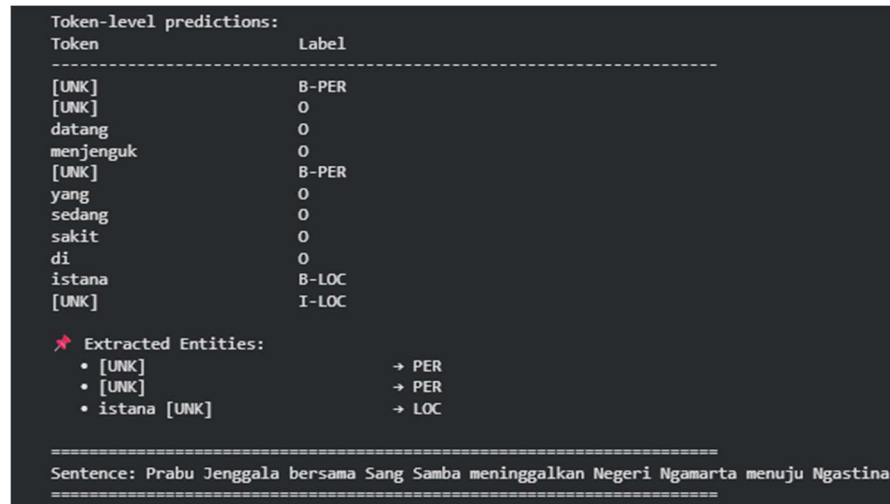
Hasil tersebut menunjukkan bahwa model masih kesulitan membedakan organisasi dengan entitas lain

Kemungkinan penyebabnya adalah:

1. Jumlah data ORG lebih sedikit.
2. Nama organisasi sering memiliki bentuk yang mirip dengan nama lokasi.
3. Variasi organisasi pada dataset kurang beragam.

L. Hasil Pengujian pada Cerita Hikayat Wayang

Setelah model dievaluasi menggunakan dataset pengujian, dilakukan pengujian tambahan menggunakan 40 kalimat yang berasal dari cerita Hikayat Wayang Arjuna dan Purasara.



Token-level predictions:

Token	Label
[UNK]	B-PER
[UNK]	O
datang	O
menjenguk	O
[UNK]	B-PER
yang	O
sedang	O
sakit	O
di	O
istana	B-LOC
[UNK]	I-LOC

✦ Extracted Entities:

- [UNK] → PER
- [UNK] → PER
- istana [UNK] → LOC

Sentence: Prabu Jenggala bersama Sang Samba meninggalkan Negeri Ngamarta menuju Ngastina

Gambar 4.9 Grafik Hasil Pengujian

Pengujian ini bertujuan untuk mengetahui kemampuan model dalam mengenali entitas pada domain cerita rakyat yang memiliki karakteristik berbeda dengan dataset pelatihan.

Contoh kalimat pengujian:

1. Rajuna mengadukan sikap Prabu Jenggala kepada Ki Darmawangsa di Negeri Ngamarta.
2. Batara Guru dan Batara Narada membicarakan kekacauan di Kayangan.
3. Purasara bertapa di Gunung Parasu ditemani oleh Lurah Semar.

Hasil pengujian menunjukkan bahwa model mampu mengenali sebagian besar entitas person dan location, namun masih mengalami kesulitan terhadap nama-nama tokoh yang belum pernah muncul selama proses pelatihan.

M. Pembahasan

Secara keseluruhan penelitian ini berhasil menunjukkan bahwa model IndoBERT dapat diterapkan pada tugas Named Entity Recognition (NER) untuk teks berbahasa Indonesia. Hasil evaluasi menunjukkan bahwa model mampu mengenali entitas dengan nilai Precision sebesar 56,30%, Recall sebesar 60,55%, dan F1-Score sebesar 58,35%. Nilai tersebut menunjukkan bahwa model telah mampu melakukan identifikasi entitas secara otomatis pada teks cerita rakyat dengan performa yang

cukup baik. Hasil ini membuktikan bahwa penggunaan model berbasis Transformer memiliki kemampuan yang lebih baik dalam memahami konteks kata dibandingkan metode konvensional yang hanya mengandalkan fitur statistik atau aturan tertentu.

Berdasarkan hasil evaluasi per kategori entitas, kategori Person (PER) memperoleh performa terbaik dengan nilai F1-Score yang tinggi. Hal ini menunjukkan bahwa model mampu mengenali nama tokoh dengan baik karena pola penulisan nama tokoh pada dataset relatif konsisten dan sering muncul selama proses pelatihan. Selain itu, kategori Location (LOC) juga menunjukkan performa yang baik karena sebagian besar nama lokasi memiliki karakteristik yang jelas sehingga lebih mudah dipelajari oleh model. Sebaliknya, kategori Organization (ORG) memperoleh nilai yang lebih rendah dibandingkan kategori lainnya. Kondisi tersebut menunjukkan bahwa model masih mengalami kesulitan dalam membedakan nama organisasi dengan nama lokasi atau nama tokoh yang memiliki struktur penulisan serupa.

Hasil pengujian menggunakan kalimat-kalimat dari cerita Hikayat Wayang Arjuna dan Purasara menunjukkan bahwa model mampu mengenali sebagian besar entitas penting yang terdapat dalam cerita. Namun demikian, masih ditemukan beberapa kesalahan prediksi pada nama tokoh pewayangan yang jarang muncul atau tidak terdapat pada data pelatihan. Kondisi ini terjadi karena model IndoBERT yang digunakan merupakan model pralatih yang dilatih menggunakan korpus bahasa Indonesia umum sehingga belum sepenuhnya memahami kosakata khas pewayangan dan sastra klasik. Oleh karena itu, penambahan dataset khusus yang berasal dari cerita rakyat dan naskah budaya Indonesia diperkirakan dapat meningkatkan performa model secara signifikan.

Selain itu, penelitian ini menunjukkan bahwa teknologi NLP dapat dimanfaatkan sebagai salah satu upaya digitalisasi dan pelestarian budaya Indonesia. Melalui penerapan NER, informasi penting yang terdapat dalam cerita rakyat dapat diekstraksi secara otomatis sehingga memudahkan proses pengelolaan, pencarian, dan analisis data budaya..

BAB V PENUTUP

A. Kesimpulan

Penelitian ini berhasil mengimplementasikan metode Named Entity Recognition (NER) menggunakan model IndoBERT untuk melakukan identifikasi entitas pada teks berbahasa Indonesia. Model yang digunakan merupakan model dasar IndoBERT (*indobenchmark/indobert-base-pl*) yang telah dilakukan proses *fine-tuning* menggunakan dataset NER berbahasa Indonesia dengan kategori entitas Person (PER), Location (LOC), dan Organization (ORG). Melalui proses pelatihan selama 40 epoch, model mampu mempelajari pola hubungan antara token dan label entitas sehingga dapat digunakan untuk melakukan ekstraksi informasi secara otomatis dari suatu teks.

Berdasarkan hasil evaluasi pada data pengujian, model memperoleh nilai Precision sebesar 56,30%, Recall sebesar 60,55%, dan F1-Score sebesar 58,35%. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengenali entitas pada teks berbahasa Indonesia. Nilai Recall yang lebih tinggi dibandingkan Precision menunjukkan bahwa model relatif mampu menemukan banyak entitas yang benar, meskipun masih terdapat beberapa prediksi yang kurang tepat.

Hasil evaluasi per kategori entitas menunjukkan bahwa kategori Location (LOC) memiliki performa terbaik dengan nilai F1-Score sebesar 76,07%. Hal ini menunjukkan bahwa model mampu mengenali nama lokasi dengan tingkat ketepatan yang cukup tinggi. Sementara itu, kategori Person (PER) memperoleh nilai F1-Score sebesar 64,67%, yang menunjukkan bahwa model cukup baik dalam mengenali nama tokoh atau individu. Sebaliknya, kategori Organization (ORG) memperoleh nilai F1-Score sebesar 44,96%, sehingga masih menjadi kategori yang paling sulit dikenali oleh model.

Pengujian tambahan menggunakan kalimat-kalimat dari cerita Hikayat Wayang Arjuna dan Purasara menunjukkan bahwa model mampu mengenali sebagian besar entitas tokoh dan lokasi yang terdapat dalam cerita. Namun demikian, masih ditemukan beberapa kesalahan prediksi akibat munculnya token yang tidak dikenal (*unknown token* atau [UNK]) pada nama-nama tokoh pewayangan tertentu. Kondisi ini menunjukkan bahwa model masih memiliki keterbatasan dalam memahami kosakata yang sangat spesifik terhadap domain cerita rakyat dan pewayangan.

Secara keseluruhan, penelitian ini membuktikan bahwa IndoBERT dapat digunakan untuk tugas Named Entity Recognition pada teks cerita rakyat Indonesia. Model mampu

melakukan ekstraksi entitas secara otomatis dan memberikan hasil yang cukup baik. Dengan demikian, penerapan IndoBERT dapat menjadi salah satu solusi dalam pengolahan teks bahasa Indonesia, khususnya untuk membantu proses identifikasi informasi penting yang terdapat pada cerita rakyat dan naskah budaya Indonesia.

B. Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dijadikan sebagai bahan pengembangan pada penelitian selanjutnya.

Pertama, penelitian berikutnya disarankan untuk menggunakan dataset yang lebih besar dan lebih beragam. Semakin banyak data pelatihan yang digunakan, maka semakin banyak pula pola bahasa yang dapat dipelajari oleh model. Penambahan jumlah data juga berpotensi meningkatkan nilai Precision, Recall, dan F1-Score sehingga menghasilkan model yang lebih akurat.

Kedua, penelitian selanjutnya dapat menggunakan dataset yang secara khusus berasal dari cerita rakyat, pewayangan, legenda, maupun naskah budaya Indonesia. Penggunaan dataset yang sesuai dengan domain penelitian diharapkan dapat mengurangi kemunculan token [UNK] dan meningkatkan kemampuan model dalam mengenali nama tokoh, kerajaan, maupun lokasi yang unik pada cerita rakyat Indonesia.

Ketiga, penelitian berikutnya dapat melakukan perbandingan performa antara IndoBERT dengan model lain seperti IndoBERT-Large, XLM-RoBERTa, atau mBERT untuk mengetahui model yang paling sesuai dalam tugas Named Entity Recognition berbahasa Indonesia.

Keempat, penelitian selanjutnya dapat menambahkan kategori entitas yang lebih banyak, seperti Event, Artifact, Time, maupun Mythological Character sehingga informasi yang diekstraksi dari cerita rakyat menjadi lebih lengkap dan kaya makna.

Kelima, hasil penelitian ini dapat dikembangkan menjadi aplikasi berbasis web atau mobile yang mampu melakukan ekstraksi entitas secara otomatis dari berbagai cerita rakyat Indonesia.

SURAT KETERANGAN BEBAS PLAGIARISME



Universitas Nusantara PGRI Kediri
UPT. PERPUSTAKAAN, PUBLIKASI DAN INOVASI
Alamat: Kampus 1, Jl. KH. Ahmad Dahlan No.76 Kota Kediri 64112
Telp. (0354) 771576, (0354) 771503, (0354) 771495, Fax. (0354) 771576
Website: <http://ppi.unpkediri.ac.id/> Email: perpustakaan@unpkediri.ac.id

SURAT KETERANGAN BEBAS SIMILARITY

Ketua UPT Perpustakaan, Publikasi dan Inovasi Universitas Nusantara PGRI Kediri menerangkan bahwa mahasiswa dengan identitas berikut:

Nama Mahasiswa : Danda Bagaskoro
NPM : 2213030083
Program Studi : S1-Sistem Informasi

Judul Karya Ilmiah:

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

Dinyatakan sudah memenuhi syarat batas maksimal 30% *similarity* sesuai dengan ketentuan yang berlaku pada setiap subbab naskah Laporan **Tugas Akhir/Skripsi/Tesis** yang disusun.

Demikian Surat Keterangan ini kami berikan untuk dapat dipergunakan seperlunya.



Kediri, 15 Juni 2026
Ka UPT PPI,



Dr. Abdul Aziz Hunaifi, M.A

KARTU BIMBINGAN TA/SKRIPSI/TESIS



Nama : Danda Bagaskoro
 NPM : 2213030083
 Prodi : Sistem Informasi
 Semester : GNP
 Thn Akademik : 25/26

No.	Sem/TA	Tanggal	Dosen Pembimbing	Materi	Catatan
1	GNP 25/26	04 Maret 2026	ERNA DANIATI	BAB I	Teori Dan Kajian Pustaka.
2	GNP 25/26	11 Maret 2026	ERNA DANIATI	BAB II	Teori Dan Kajian Pustaka II.
3	GNP 25/26	25 Maret 2026	ERNA DANIATI	BAB III	Analisis data dan evaluasi I.
4	GNP 25/26	01 April 2026	ERNA DANIATI	BAB III	Analisis dan evaluasi II.
5	GNP 25/26	08 April 2026	ERNA DANIATI	BAB IV	Pembahasan I.
6	GNP 25/26	15 April 2026	ERNA DANIATI	BAB IV	Pembahasan II.
7	GNP 25/26	22 April 2026	ERNA DANIATI	BAB IV	Hasil.
8	GNP 25/26	29 April 2026	ERNA DANIATI	BAB V	Kesimpulan.
9	GNP 25/26	06 Mei 2026	ERNA DANIATI	BAB V	Saran.
10	GNP 25/26	13 Mei 2026	ERNA DANIATI	Publikasi	Draft Publikasi.
11	GNP 25/26	18 Mei 2026	ERNA DANIATI	Publikasi	Tempalting Dan Sitasi.
12	GNP 25/26	19 Mei 2026	ERNA DANIATI	Publikasi	Pemilihan Tempat Publikasi (JOISM).
13	GNP 25/26	22 Mei 2026	ERNA DANIATI	Publikasi	Submit Artikel.
14	GNP 25/26	23 Juni 2026	ERNA DANIATI	Publikasi	Revisi Publikasi Artikel.
15	GNP 25/26	25 Juni 2026	ERNA DANIATI	Publikasi	Persiapan PPTX Sidang.
16	GNP 25/26	26 Mei 2026	ERNA DANIATI	Publikasi	Revisi PPTX.
17	GNP 25/26	29 Mei 2026	ERNA DANIATI	Publikasi	ACC.
18	GNP 25/26	19 Maret 2026	RINA FIRLIANA	BAB I	Revisi.
19	GNP 25/26	31 Maret 2026	RINA FIRLIANA	BAB I	Revisi Penulisan Bab I.
20	GNP 25/26	01 April 2026	RINA FIRLIANA	BAB I	Revisi Pendahuluan.
21	GNP 25/26	08 April 2026	RINA FIRLIANA	BAB II	Revisi Penulisan Bab II.
22	GNP 25/26	15 April 2026	RINA FIRLIANA	BAB II	Lanjut Bab III.
23	GNP 25/26	22 April 2026	RINA FIRLIANA	BAB III	Revisi Penulisan Bab III.
24	GNP 25/26	29 April 2026	RINA FIRLIANA	BAB III	Revisi Penulisan Dan Tabel.
25	GNP 25/26	06 Mei 2026	RINA FIRLIANA	BAB III	Lanjut Bab IV.
26	GNP 25/26	13 Mei 2026	RINA FIRLIANA	BAB IV	Revisi Hasil Dan Pembahasan.
27	GNP 25/26	20 Mei 2026	RINA FIRLIANA	BAB IV	Revisi Penulisan.
28	GNP 25/26	22 Mei 2026	RINA FIRLIANA	BAB IV	Lanjut Bab V.
29	GNP 25/26	27 Mei 2026	RINA FIRLIANA	BAB V	Revisi Kesimpulan.
30	GNP 25/26	28 Mei 2026	RINA FIRLIANA	BAB V	Revisi Saran.
31	GNP 25/26	29 Mei 2026	RINA FIRLIANA	Publikasi	Jurnal JOISM.
32	GNP 25/26	03 Juni 2026	RINA FIRLIANA	Lainnya	PPTX .
33	GNP 25/26	08 Juni 2026	RINA FIRLIANA	Lainnya	ACC.

Kaprodi Sistem Informasi

Dosen Pembimbing Akademik

Dosen Pembimbing Utama

☑ Valid

☑ Valid

☑ Valid

(Dr. SUCIPTO, S.Kom, M.Kom)

(Dr. ERNA DANIATI, M.Kom)

(Dr. ERNA DANIATI, M.Kom)



Yayasan Pembina Lembaga Pendidikan Perguruan Tinggi PGRI Kediri

UNIVERSITAS NUSANTARA PGRI KEDIRI

LEMBAGA PENELITIAN DAN PENGABDIAN PADA MASYARAKAT (LPPM)

Kampus I Jl. KH. Achmad Dahlan No. 76 Kediri (64112) Telp. (0354) 771576, Fax. 771576

Website: www.lp2m.unpkediri.ac.id, Email: lemlit@unpkediri.ac.id.

BERITA ACARA UJIAN SKRIPSI

Pada hari ini Jum'at tanggal 26 bulan Juni tahun 2026 pukul 10.30 bertempat di Ruang Kampus Universitas Nusantara PGRI Kediri, telah dilaksanakan Ujian Skripsi Mahasiswa Universitas Nusantara PGRI Kediri:

NAMA : Danda Bagaskoro
NPM : 2213030083
FAK - PRODI : Teknik Dan Ilmu Komputer – Sistem Informasi
JUDUL : Penerapan Named Entity Recognition (NER) Menggunakan Indobert Pada Cerita Rakyat "Hikayat Wayang Arjuna & Purusara"
Dengan Hasil :

Kediri, 26 Juni 2026

Mengetahui,

Ketua Penguji : Dr. Erna Daniati, M.Kom

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

Penguji 2 : Rina Firliana, M.Kom

RENTAN NILAI

85-100	= A	60-64	= C+
80-84	= A-	55-59	= C
75-79	= B+	40-54	= D
70-74	= B	0-39	= E
65-69	= B-		



Universitas Nusantara PGRI Kediri

Status Terakreditasi Baik Sekali

SK BAN-PT No.671/SK/BAN-PT/Akred/PT/VII/2021 Tanggal 21 Juli 2021

Lembaga Penelitian Dan Pengabdian Pada Masyarakat (LPPM)

Kampus I Jl. K.H. Achmad Dahlan 76 Kediri 64112 Telp.(0354)771576,771503

Kediri

Website: <https://lp2m.unpkediri.ac.id> email: lemlit@unpkediri.ac.id

LEMBAR REVISI UJIAN SKRIPSI



Nama : Danda Bagaskoro
NPM : 2213030083
Prodi : S1-Sistem Informasi
Semester/Tahun Akademik : Genap, 25/26
Hari/Tanggal Ujian : Jumat / 26 Juni 2026

Judul : PENERAPAN NAMED ENTITY RECOGNITION (NER) MENGGUNAKAN
IndoBERT PADA CERITA RAKYAT "HIKAYAT WAYANG ARJUNA &
PURUSARA"
Dengan Hasil : Revisi Naskah
Nilai : 91 (A)

Catatan Revisi:

tata tulis, 1. Tata tulis cek pada naskah 2. Lanjutkan publikasi sampai publish, Revisi: Teknik Penulisan Sinkronkan Rumusan Masalah Dengan Kesimpulan

Ketua Penguji

(Dr. ERNA DANIATI, M.Kom)

Penguji 1

(DWI HARINI, S.Si., M.IA)

Penguji 2

(RINA FIRLIANA, M.Kom)

nb:

- Tanda Tangan revisi (Ketua Penguji, Penguji 1, dan Penguji ke-n) dilakukan secara manual
- Nilai berasal dari gabungan dari seluruh dosen penguji
- Daftar Dosen yang telah mengisi nilai: DWI HARINI, ERNA DANIATI, RINA FIRLIANA