

**MENGGALI ENTITAS PADA CERITA KLASIK SASTRA JAWA
MENGUNAKAN NAMED ENTITY RECOGNITION : PELESTARIAN
BUDAYA MELALUI PENDEKATAN MACHINE LEARNING**

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

AKMAL HISYAM PRADHANA

NPM: 2113030019

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

2025

Skripsi oleh :

AKMAL HISYAM PRADHANA

NPM : 2113030019

Judul

**MENGKALI ENTITAS PADA CERITA KLASIK SASTRA JAWA
MENGUNAKAN NAMED ENTITY RECOGNITION : PELESTARIAN
BUDAYA MELALUI PENDEKATAN MACHINE LEARNING**

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

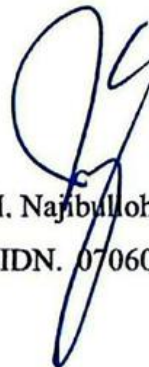
Tanggal: 17 Juni 2025

Pembimbing I



Erna Daniati, M.Kom.
NIDN. 0723058501

Pembimbing II



M. Najibulloh Muzaki, S.Kom., M.Cs
NIDN. 0706098902

Skripsi oleh :

AKMAL HISYAM PRADHANA

NPM : 2113030019

Judul

**MENGGALI ENTITAS PADA CERITA KLASIK SASTRA JAWA
MENGUNAKAN NAMED ENTITY RECOGNITION : PELESTARIAN
BUDAYA MELALUI PENDEKATAN MACHINE LEARNING**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 8 Juli 2025

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

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



2. Penguji I : Arie Nugroho, S.Kom., M.M., M.Kom.



3. Penguji II : M. Najibulloh Muzaki, S.Kom., M.Cs



Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer



Sulistiono, M.Si.

NIDN. 0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama	: Akmal Hisyam Pradhana
Jenis Kelamin	: Laki-Laki
Tempat/tgl. lahir	: Trenggalek / 22 Juli 2002
NPM	: 2113030019
Fak/Jur./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, 17 Juni 2025

Yang Menyatakan



AKMAL HISYAM PRADAHANA
NPM: 2113030019

MOTTO & PERSEMBAHAN

MOTTO

“Menuntut ilmu adalah takwa. Menyampaikan ilmu adalah ibadah. Mengulang-ulang ilmu adalah zikir. Mencari ilmu adalah jihad”

“Ilmu tanpa amal adalah kegilaan, amal tanpa ilmu adalah kesia-siaan.”

(Abu Hamid Muhammad bin Muhammad al-Ghazali ath-Thusi asy-Syafi'i)

PERSEMBAHAN

Puji Segala puji dan syukur saya haturkan ke hadirat Tuhan Yang Maha Esa atas anugerah dan petunjuk-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan baik. Dengan penuh rasa hormat dan ketulusan, saya persembahkan karya ini kepada: Kedua orang tua tercinta, yang senantiasa hadir dalam setiap langkah kehidupan saya melalui doa, kasih yang tak terbatas, serta pengorbanan yang tiada henti. Kehadiran dan cinta kalian menjadi sumber kekuatan dan motivasi terbesar bagi saya. Para dosen yang telah membagikan ilmu serta memberikan arahan selama masa perkuliahan, khususnya dosen pembimbing saya yang dengan kesabaran, ketulusan, dan komitmen tinggi telah membimbing saya hingga skripsi ini rampung. Sahabat-sahabat dan rekan seperjuangan, atas segala kebersamaan, dukungan, dan semangat yang senantiasa menguatkan selama proses ini berlangsung. Semoga karya sederhana ini dapat memberikan manfaat serta menjadi langkah awal menuju masa depan yang lebih baik dan penuh harapan.

ABSTRAK

Akmal Hisyam Pradhana : Menggali Entitas Pada Cerita Klasik Sastra Jawa Menggunakan Named Entity Recognition : Pelestarian Budaya Melalui Pendekatan Machine Learning Kediri, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2025.

Kata Kunci: Bi-LSTM, Cerita Rakyat, Named Entity Recognition, NLP, spaCy.

Cerita rakyat merupakan bagian penting dari warisan budaya Indonesia yang diwariskan secara turun-temurun dan mencerminkan nilai-nilai moral, sejarah, serta identitas lokal. Salah satu bentuk cerita rakyat yang menonjol adalah Cerita Klasik Sastra Jawa, yang seringkali dituturkan secara lisan dan memuat unsur mitos, legenda, serta simbol budaya. Di era digital saat ini, kebutuhan untuk mengolah cerita-cerita tersebut secara otomatis semakin mendesak, khususnya dalam konteks pelestarian dan pemanfaatan informasi budaya. Namun, proses identifikasi entitas penting seperti tokoh, lokasi, dan organisasi dalam cerita-cerita tersebut masih banyak dilakukan secara manual, yang memerlukan waktu lama dan berisiko tidak konsisten.

Penelitian ini bertujuan untuk mengembangkan sistem otomatis dalam identifikasi entitas menggunakan metode Named Entity Recognition (NER) pada teks Cerita Klasik Sastra Jawa. Penelitian menggunakan pendekatan kuantitatif berbasis Natural Language Processing (NLP) dan mengimplementasikan dua model, yaitu NER dengan library spaCy dan arsitektur Bi-LSTM. Dataset yang digunakan berupa cerita Panji dalam format JSON. Proses meliputi ekstraksi entitas menggunakan model `xx_ent_wiki_sm` dari spaCy dan evaluasi kinerja model melalui metrik akurasi, precision, recall, dan F1-score.

Hasil penelitian menunjukkan bahwa model NER berbasis spaCy berhasil mengidentifikasi total 289 entitas, dengan dominasi entitas Person (73%). Akurasi rata-rata model spaCy mencapai 0,96 dengan F1-score tertinggi pada entitas Person dan Organization (0,99). Model Bi-LSTM menunjukkan akurasi lebih tinggi, yaitu 99%, meskipun performa terbaik hanya terlihat pada entitas dengan jumlah data pelatihan yang besar. Visualisasi entitas dalam bentuk diagram pie juga memperkuat temuan distribusi entitas dalam teks.

Kesimpulan dari penelitian ini menegaskan bahwa NER mampu secara efektif mengidentifikasi entitas dalam cerita klasik sastra Jawa. Namun, terdapat keterbatasan pada variasi data dan kompleksitas bahasa naratif lokal. Penelitian selanjutnya disarankan untuk mengeksplorasi model berbasis Transformer seperti IndoBERT dan memperluas cakupan dataset agar sistem dapat lebih adaptif terhadap variasi budaya dan bahasa yang lebih luas di Indonesia.

KATA PENGANTAR

Puji Syukur Kami panjat kan kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan- Nya tugas penyusunan proposal ini dapat diselesaikan. Penyusunan skripsi ini merupakan bagian dari rencana penelitian guna penyusunan skripsi sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi. Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus-tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri.
3. Sucipto, M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri.
4. Erna Daniati, M.Kom. selaku Dosen pembimbing Pertama yang telah sukarela memberikan waktu, tenaga, dan pemikiran dalam mendampingi penulis selama proses penyusunan skripsi ini.
5. Muhammad Najibulloh Muzaki, M.Cs selaku Dosen pembimbing Kedua yang telah memberikan waktu, tenaga, dan pemikiran dalam mendampingi penulis selama proses penyusunan skripsi ini.
6. Orang tua yang selalu mendoakan dan memberikan semangat agar tetap berusaha dan tidak mudah menyerah. Terima kasih atas semua dukungan baik spiritual, moral, maupun material hingga terselesaikannya laporan ini.

Demikian kata pengantar ini penulis sampaikan. Penulis berharap skripsi ini bisa bermanfaat serta berkah bagi penulis, pembaca, dan pihak-pihak terkait. Akhir kata, penulis mengucapkan terima kasih atas semua bimbingan dan dukungan yang diberikan oleh semua pihak.

Kediri, 17 Juni 2025



AKMAL HISYAM PRADHANA
NPM : 2113030019

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN MOTTO	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	
A. Latar Belakang Masalah	1
B. Identifikasi Masalah	4
C. Batasan Masalah	4
D. Rumusan Masalah	4
E. Tujuan Penelitian	5
F. Manfaat Penelitian	5
BAB II KAJIAN PUSTAKA	
A. Kajian Teori	6
1. <i>Natural Language Processing</i> (NLP)	6
2. Cerita Rakyat	6
3. <i>Named Entity Recognition</i> (NER)	7
4. <i>SpaCy</i>	7
5. Visualisasi Data	8
6. Evaluasi Skor	9
B. Kajian Hasil Penelitian Terdahulu	10
C. Kerangka Berpikir	30
D. Hipotesis	31
BAB III METODE PENELITIAN	
A. Desain Penelitian	32
B. Alat, Bahan, dan atau Instrumen Penelitian	34

C. Populasi dan Sampel/ Objek Penelitian/ Subyek Penelitian.....	34
D. Prosedur Penelitian	35
E. Tempat dan Waktu Penelitian.....	36
F. Teknik Analisis Data	37
 BAB IV HASIL DAN PEMBAHASAN	
A. Gambaran Model	38
B. Proses Penerapan <i>Named Entity Recognition</i>	38
C. Hasil Jumlah Entitas	46
D. Hasil <i>Pie Chart</i>	47
E. Hasil <i>Classification Report</i>	48
 BAB V PENUTUP	
A. Kesimpulan.....	50
B. Saran	51
DAFTAR PUSTAKA	52
LAMPIRAN	57

DAFTAR TABEL

Tabel 2.1 Kajian Hasil Penelitian Terdahulu	10
Tabel 3.1 Waktu Penelitian Dalam Format Gantt Chart	36
Tabel 4.1 Cerita Panji Yang Diteliti.....	41
Tabel 4.2 Kategori Entitas <i>Library Spacy</i>	42
Tabel 4.3 Kategori Entitas Model <i>Bi-LSTM</i>	43
Tabel 4.4 Jumlah Entitas Dari 3 Cerita Panji	46
Tabel 4.5 <i>Classification Report Library Spacy</i>	48

DAFTAR GAMBAR

Gambar 2.1 Entitas <i>Named Entity Recognition</i>	7
Gambar 2.2 Tampilan Website <i>spaCy</i>	8
Gambar 2.3 Diagram Pie Chart.....	9
Gambar 2.4 <i>Confusion Matrix</i>	10
Gambar 2.5 Kerangka Berpikir.....	30
Gambar 3.1 Alur Penelitian.....	32
Gambar 4.1 Alur <i>Named Entity Recognition</i>	38
Gambar 4.2 Source Code Memproses File <i>JSON</i>	39
Gambar 4.3 Mapping Kata Menjadi Indeks.....	40
Gambar 4.4 Source Code Perhitungan Jumlah Entitas Pada <i>Spacy</i>	41
Gambar 4.5 Alur Kerja <i>Named Entity Recognition</i>	43
Gambar 4.6 Source Code <i>Pie Chart</i> Pada <i>Spacy</i>	44
Gambar 4.7 Hasil <i>Pie Chart</i>	45
Gambar 4.8 <i>Classification Report</i> Model <i>Bi-LSTM</i>	45

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Skripsi	57
Lampiran 2. Surat Keterangan Bebas Similarity Dari PPI.....	58
Lampiran 3. Bukti Skor Similarity.....	59
Lampiran 4. Berita Acara Ujian Skripsi.....	60
Lampiran 5. Lembar Revisi Ujian SIAKAD.....	61
Lampiran 6. Bukti Submit & Publish Artikel	62

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Cerita rakyat merupakan salah satu warisan budaya yang dilestarikan turun temurun di berbagai daerah khususnya negara Indonesia. Direktorat Jenderal Kebudayaan Indonesia mencatat bahwa, di Indonesia terdapat 945 cerita rakyat yang tersebar di berbagai daerah. Cerita tersebut tidak hanya mengandung nilai-nilai moral dan budaya, namun juga berfungsi sebagai sumber informasi tentang sejarah, geografis, dan karakter tokoh yang pernah hidup di masa lampau (Yulianti et al., 2019). Cerita rakyat yang berkembang di Indonesia misalnya Cerita Klasik Sastra Jawa. Sebuah cerita lokal yang berkembang dan dipelihara oleh warga di Indonesia, sering kali dipercaya sebagai peristiwa nyata meski juga dianggap hanya sebagai mitos dan dongeng (Indriana et al., 2021). Seiring perkembangan teknologi, kebutuhan akan pengolahan informasi digital dari Cerita Klasik Sastra Jawa semakin meningkat, terutama dalam upaya melestarikan, mengklasifikasikan, memanfaatkan, dan pencarian informasi tersebut untuk kepentingan akademik, budaya, hingga pariwisata (Santyaningtyas & Noor, 2016).

Salah satu masalah saat pengolahan informasi tersebut adalah bagaimana mengidentifikasi dan mengelompokkan entitas penting, seperti Tokoh, Lokasi, dan Organisasi, yang terdapat dalam Cerita Klasik Sastra Jawa tersebut secara otomatis (Sukardi et al., 2020a). Identifikasi entitas untuk pencarian informasi dalam cerita klasik sastra jawa sering kali dilakukan secara manual oleh peneliti dan sejarawan. Metode ini tidak hanya memakan waktu yang lama dan melelahkan, tetapi juga rentan terhadap kesalahan, terutama saat mengelola data yang besar (Kim et al., 2020). Oleh karena itu, diperlukan solusi berbasis teknologi yang mampu melakukan proses pengenalan entitas secara otomatis dengan hasil yang akurat dan efisien. Penerapan *Named Entity Recognition* (NER) menjadi salah satu solusi yang dapat diandalkan dalam identifikasi entitas secara otomatis pada Cerita Klasik Sastra Jawa.

Pengambilan informasi menggunakan *Named Entity Recognition* (NER) merupakan komponen utama dalam pemrosesan bahasa alami (NLP), yang berperan penting dalam proses identifikasi dan klasifikasi entitas secara otomatis (Daniati et al., 2025a). Dengan *Named Entity Recognition* (NER), proses ekstraksi entitas seperti Nama Tokoh, Lokasi, dan Organisasi dapat dilakukan secara otomatis, sehingga mempercepat proses identifikasi entitas dan mengurangi ketergantungan identifikasi entitas secara manual. Sehingga, *Named Entity Recognition* (NER) dapat menyelesaikan berbagai permasalahan identifikasi entitas secara manual pada Cerita Klasik Sastra Jawa seperti banyak memakan waktu lama, bersifat subjektif yaitu hasil yang diperoleh dapat berbeda tergantung individu yang melakukan identifikasi, tidak efisien dan rentan tidak konsisten saat identifikasi entitas secara manual.

Saat pembentukan model *Named Entity Recognition* (NER), peneliti menggunakan *library spaCy* karena merupakan salah satu library yang simpel dan efisien. Selain itu masih relatif jarang penelitian yang menggunakan *library spaCy* secara eksplisit untuk *Named Entity Recognition* (NER). Pada penelitian ini, Peneliti tertarik mengidentifikasi dan mengklasifikasi entitas pada teks Cerita Klasik Sastra Jawa karena penerapan *Named Entity Recognition* (NER) pada teks Cerita Klasik Sastra Jawa masih sangat jarang dan terbatas. Cerita Klasik Sastra Jawa memiliki karakteristik khas seperti unsur tokoh mitos, tempat legendaris, simbol budaya (Fatubun, 2021). Peneliti berupaya membuat penelitian baru dengan menerapkan *Named Entity Recognition* (NER) pada Cerita Klasik Sastra Jawa. Dataset yang digunakan untuk menguji *Named Entity Recognition* (NER) yaitu Cerita Panji yang merupakan salah satu Cerita Klasik Sastra Jawa.

Penelitian ini memberikan kontribusi dalam bidang pemrosesan bahasa alami (*Natural Language Processing*), khususnya dalam penerapan *Named Entity Recognition* (NER) pada bahasa Indonesia, melalui eksplorasi data uji berupa cerita klasik sastra Jawa. Salah satu kontribusi utama dari penelitian ini terletak pada pemanfaatan cerita klasik sastra Jawa sebagai data testing, yang hingga saat ini belum banyak dijadikan objek pengujian dalam studi-studi sejenis. Sebagian besar penelitian NER sebelumnya masih berfokus pada jenis teks formal seperti berita, dokumen administratif, atau artikel ilmiah. Dengan menggunakan cerita rakyat

sebagai data testing, penelitian ini berhasil menunjukkan kemampuan model NER dalam mengenali entitas yang muncul dalam konteks budaya lokal dengan gaya bahasa naratif yang khas, termasuk penggunaan nama tokoh legenda, tempat tradisional, serta struktur kalimat yang lebih bebas dan penuh nuansa lokal.

Kontribusi menjadi penting karena memperluas cakupan dan validitas sistem NER terhadap jenis teks yang lebih kompleks dan kaya secara kultural. Selain memberikan tantangan bahasa yang unik, pendekatan ini juga membuka peluang penelitian lanjutan dalam pengembangan model NER yang lebih adaptif terhadap domain dan konteks budaya tertentu. Tidak hanya itu, penelitian ini juga berkontribusi terhadap pelestarian budaya lokal, karena hasil ekstraksi entitas dari cerita klasik sastra Jawa dapat dimanfaatkan untuk mendokumentasikan informasi penting seperti tokoh, tempat, serta nilai-nilai budaya yang terkandung di dalamnya. Dengan demikian, penelitian ini tidak hanya memiliki nilai teknis dalam pengembangan sistem NER, tetapi juga nilai strategis dalam upaya pelestarian dan digitalisasi warisan budaya Nusantara. Penerapan *Named Entity Recognition* (NER) pada Cerita Klasik Sastra Jawa juga menghadapi tantangan, terutama karena karakteristik bahasa yang digunakan dalam Cerita Klasik Sastra Jawa bersifat figuratif, dan tidak selalu mengikuti kaidah bahasa modern. *Accuracy*, *precision*, *recall*, dan *F1-score* menjadi metrik penting dalam mengevaluasi seberapa baik sistem *Named Entity Recognition* (NER) dapat mengidentifikasi entitas dalam Cerita Klasik Sastra Jawa tersebut.

Dengan demikian, penelitian ini bertujuan untuk menerapkan *Named Entity Recognition* (NER) pada teks Cerita Klasik Sastra Jawa, mengetahui jumlah dan presentase entitas serta mengukur kinerjanya berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan sistem otomatis dalam pengelolaan cerita rakyat dan pelestarian budaya, serta menjadi dasar bagi penelitian di bidang *Natural Language Processing* (NLP) dalam konteks humaniora budaya dan sastra (Daniati et al., 2025b).

B. Identifikasi Masalah

1. Identifikasi entitas dalam cerita klasik sastra Jawa sering kali dilakukan secara manual oleh peneliti dan sejarawan.
2. Banyak memakan waktu lama saat identifikasi entitas secara manual pada cerita klasik sastra Jawa.
3. Identifikasi entitas dengan cara manual bersifat subjektif. Hasil yang diperoleh dapat berbeda tergantung individu yang melakukan pencarian.
4. Ekstraksi Informasi dengan cara lama sering tidak efisien dan rentan tidak konsisten. Sehingga menghasilkan entitas yang kurang optimal.

C. Batasan Masalah

Peneliti membatasi masalah sebagai berikut :

1. Terdapat 4 kategori entitas yaitu *Person*, *Organization*, *Location*, dan *Miscellaneous*.
2. Dataset untuk Testing berupa teks Cerita Panji Klasik Sastra Jawa, data source : doi.org/10.17632/fry8x5rkz7.2 yang berbahasa Indonesia. Sedangkan data training dalam bentuk anotasi Named Entity Recognition NER dengan sumber data github.com/yudanta/indonesia-ner-spacy/blob/main/ner_spacy_fmt_datasets.pickle.
3. *Software library* yang digunakan untuk *Named Entity Recognition* adalah *Spacy*.
4. Data keluaran yang dihasilkan berupa *classification report* dan jumlah label kategori entitas.

D. Rumusan Masalah

1. Bagaimana proses klasifikasi suatu entitas menggunakan *library Spacy* ?
2. Bagaimana hasil klasifikasi entitas dari cerita klasik sastra Jawa berbahasa Indonesia ?

3. Bagaimana hasil evaluasi *Named Entity Recognition* (NER) Menggunakan *Spacy* Untuk klasifikasi suatu entitas berdasarkan *classification report* ?

E. Tujuan Penelitian

1. Membuat *Named Entity Recognition* (NER) untuk klasifikasi suatu entitas menggunakan *software library Spacy*.
2. Mengetahui hasil, jumlah dan presentase entitas pada teks Cerita Klasik Sastra Jawa berbahasa Indonesia.
3. Mengetahui hasil evaluasi *Named Entity Recognition* (NER) Menggunakan *Spacy* Untuk Klasifikasi Suatu Entitas berdasarkan *Classification Report*.

F. Manfaat Penelitian

Mengacu pada tujuan penelitian yang hendak dicapai, maka penelitian ini diharapkan mempunyai manfaat pada bidang ilmu pengetahuan dan teknologi. Salah satu manfaat spesifiknya adalah peneliti dapat memahami konsep perancangan dan pengembangan *Named Entity Recognition* (NER) menggunakan *software library Spacy*. Selain itu memberikan kemudahan bagi pengguna untuk memperoleh entitas informasi berupa nama Orang, Organisasi, Lokasi, Dan sebagainya pada teks berita secara cepat dan konsisten. Membantu pengguna mempersingkat waktu untuk mendapatkan informasi penting karena menggunakan sebuah sistem *Named Entity Recognition* (NER) yang berjalan secara otomatis. Menjadi sebuah solusi dari ekstraksi pencarian informasi secara manual yang memiliki sifat subjektif dan tidak konsisten. Adapun hasil evaluasi berguna untuk mengukur seberapa baik model mengklasifikasi entitas sebagai penelitian lebih lanjut

DAFTAR PUSTAKA

- Akhtar, Z., Arshad, A., Zubair, M., Qasim, A., Ullah, F., Zamir, M. T., Ahmad, M., Sidorov, G., & Gelbukh, A. (2024). NAMED ENTITY RECOGNITION TOOLS AND TECHNIQUES IN CUSTOM NATURAL LANGUAGE PROCESSING MODELS. *Journal of Population Therapeutics and Clinical Pharmacology*. <https://doi.org/10.53555/e8b2km73>
- Akhundova, N. (2021). Named Entity Recognition for the Azerbaijani Language. *2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT)*, 1–7. <https://doi.org/10.1109/AICT52784.2021.9620336>
- Azalia, F. Y., Bijaksana, M. A., & Huda, A. F. (2019). Name Indexing in Indonesian Translation of Hadith using Named Entity Recognition with Naïve Bayes Classifier. *Procedia Computer Science*, 157, 142–149. <https://doi.org/10.1016/j.procs.2019.08.151>
- Azarine, I. S., Arif Bijaksana, M., & Asror, I. (2019). Named Entity Recognition on Indonesian Tweets using Hidden Markov Model. *2019 7th International Conference on Information and Communication Technology (ICoICT)*, 1–5. <https://doi.org/10.1109/ICoICT.2019.8835277>
- bin Ali, N., & Tanveer, B. (2022). A Comparison of Citation Sources for Reference and Citation-Based Search in Systematic Literature Reviews. *E-Informatica Software Engineering Journal*, 16(1), 220106. <https://doi.org/10.37190/e-Inf220106>
- Daniati, E., Firliana, R., Wardani, A. S., & Zarkasi, A. C. (2021). Evaluation Framework for Decision Making Based On Sentiment Analysis in Social Media. *2021 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, 47–51. <https://doi.org/10.1109/ICAMIMIA54022.2021.9807790>
- Daniati, E., & Utama, H. (2023). ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER. *Journal of Information System Management (JOISM)*, 4(2), 125–131. <https://doi.org/10.24076/joism.2023v4i2.973>
- Daniati, E., Wibawa, A. P., & Irianto, W. S. G. (2025a). Event Extraction in Narrative Texts: A Zero-Shot Approach Using Bert and Bi-LSTM on Andersen’s Fairy Tales. *2025 17th International Conference on Knowledge and Smart Technology (KST)*, 208–213. <https://doi.org/10.1109/KST65016.2025.11003342>
- Daniati, E., Wibawa, A. P., & Irianto, W. S. G. (2025b). *Extracting Narrative Events in Andersen’s Fairy Tales Using a Hybrid BERT-LSTM Model* (pp. 355–372). https://doi.org/10.1007/978-981-96-4613-5_26
- Daniati, E., Wibawa, A. P., & Irianto, W. S. G. (2025c). *The Relevance of Andersen’s Children’s Stories to the Growth and Development of Indonesian Children* (pp. 108–124). https://doi.org/10.2991/978-2-38476-352-8_10
- Daniati, E., Wibawa, A. P., Irianto, W. S. G., Ghosh, A., & Hernandez, L. (2024). Analyzing event relationships in Andersen’s Fairy Tales with BERT and Graph

- Convolutional Network (GCN). *Science in Information Technology Letters*, 5(1), 40–60. <https://doi.org/10.31763/sitech.v5i1.1810>
- Fatubun, R. R. (2021). Archetypes in Biak folktales: characters, symbols, and concepts. *Linguistics and Culture Review*, 5(1), 230–246. <https://doi.org/10.21744/lingcure.v5n1.1483>
- Goyal, A., Gupta, V., & Kumar, M. (2018). Recent Named Entity Recognition and Classification techniques: A systematic review. *Computer Science Review*, 29, 21–43. <https://doi.org/10.1016/j.cosrev.2018.06.001>
- Gunawan, W., Suhartono, D., Purnomo, F., & Ongko, A. (2018). Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs. *Procedia Computer Science*, 135, 425–432. <https://doi.org/10.1016/j.procs.2018.08.193>
- HARIYANI, Y. S., HADIYOSO, S., & SIADARI, T. S. (2020). Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 8(2), 443. <https://doi.org/10.26760/elkomika.v8i2.443>
- Hidayatullah, A. F., Apong, R. A., Lai, D. T. C., & Qazi, A. (2022). Extracting Tourist Attraction Entities from Text using Conditional Random Fields. *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, 1–6. <https://doi.org/10.1109/ICITDA55840.2022.9971310>
- Huang, W., & Yu, X. (2025). Named Entity Recognition Method Based on Multi-Feature Fusion. *Applied Sciences*, 15(1), 388. <https://doi.org/10.3390/app15010388>
- Indriana, F. N. L., Pangestu, Y. A., Amanda, B., Ranti, A. D., & Amaruli, R. J. (2021). A Preliminary Study on the Javanese Folklore as a Disaster Mitigation Strategy. *E3S Web of Conferences*, 317, 01021. <https://doi.org/10.1051/e3sconf/202131701021>
- Junaidi, S., Devegi, M., & Kurniawan, H. (2023). Pelatihan Pengolahan dan Visualisasi Data Penduduk menggunakan Python. *ADMA: Jurnal Pengabdian Dan Pemberdayaan Masyarakat*, 4(1), 151–162. <https://doi.org/10.30812/adma.v4i1.2963>
- Kim, J., Ko, Y., & Seo, J. (2020). Construction of Machine-Labeled Data for Improving Named Entity Recognition by Transfer Learning. *IEEE Access*, 8, 59684–59693. <https://doi.org/10.1109/ACCESS.2020.2981361>
- Kusumawardani, R. P., & Kusumawati, K. N. (2024). Named entity recognition in the medical domain for Indonesian language health consultation services using bidirectional-lstm-crf algorithm. *Procedia Computer Science*, 245, 1146–1156. <https://doi.org/10.1016/j.procs.2024.10.344>
- Liu, H., Jun, G., & Zheng, Y. (2021). Chinese named entity recognition model based on BERT. *MATEC Web of Conferences*, 336, 06021. <https://doi.org/10.1051/matecconf/202133606021>
- McNutt, A. M. (2022). No Grammar to Rule Them All: A Survey of JSON-style DSLs for Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 1–11. <https://doi.org/10.1109/TVCG.2022.3209460>
- Munarko, Y., Sutrisno, M. S., Mahardika, W. A. I., Nuryasin, I., & Azhar, Y. (2018). Named entity recognition model for Indonesian tweet using CRF classifier. *IOP*

- Conference Series: Materials Science and Engineering*, 403, 012067. <https://doi.org/10.1088/1757-899X/403/1/012067>
- Nurchim, N., Nurmalitasari, N., & Long, Z. A. (2023). Indonesian news classification application with named entity recognition approach. *JURNAL INFOTEL*, 15(2), 130–134. <https://doi.org/10.20895/infotel.v15i2.909>
- Putra, A. A., & Kurniawan, R. (2021). BIDIRECTIONAL LSTM-CNNs UNTUK EKSTRAKSI ENTITY LOKASI KEBAKARAN PADA BERITA ONLINE BERBAHASA INDONESIA. *Seminar Nasional Official Statistics*, 2020(1), 319–327. <https://doi.org/10.34123/semnasoffstat.v2020i1.601>
- Ramadhanty, E., Effendi, D., & Hetilaniar, H. (2022). Antropologi Sastra Dalam Kumpulan Cerita Rakyat Ogan Komerling Ilir. *Jurnal Pembahsi (Pembelajaran Bahasa Dan Sastra Indonesia)*, 12(1), 26–38. <https://doi.org/10.31851/pembahsi.v12i1.6142>
- Rifani, R., Bijaksana, M. A., & Asror, I. (2019). Named Entity Recognition for an Indonesian Based Language Tweet using Multinomial Naive Bayes Classifier. *Indonesia Journal on Computing (Indo-JC)*, 4(2), 119–126.
- Riyanto, S., Sitanggang, I. S., Djatna, T., & Atikah, T. D. (2023). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. *International Journal of Advanced Computer Science and Applications*, 14(6). <https://doi.org/10.14569/IJACSA.2023.01406116>
- Salleh, M. S., Asmai, S. A., Basiron, H., & Ahmad, S. (2017). A Malay named entity recognition using conditional random fields. *2017 5th International Conference on Information and Communication Technology (ICoICT7)*, 1–6. <https://doi.org/10.1109/ICoICT.2017.8074647>
- Sanjaya, H., Kusri, K., Yuana, K. A., & Salio, J. R. M. (2024). Multilingual Named Entity Recognition Model for Location and Time Extraction of Forest Fire. *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, 611–615. <https://doi.org/10.1109/ICSINTESA62455.2024.10747844>
- Santyaningtyas, A. C., & Noor, M. Z. M. (2016). Preserving of Traditional Culture Expression in Indonesia. *Asian Social Science*, 12(7), 59. <https://doi.org/10.5539/ass.v12n7p59>
- Sedik, R. R., & Romadhony, A. (2023). Information Extraction from Indonesian Crime News with Named Entity Recognition. *2023 15th International Conference on Knowledge and Smart Technology (KST)*, 1–5. <https://doi.org/10.1109/KST57286.2023.10086789>
- Shelar, H., Kaur, G., Heda, N., & Agrawal, P. (2020). Named Entity Recognition Approaches and Their Comparison for Custom NER Model. *Science & Technology Libraries*, 39(3), 324–337. <https://doi.org/10.1080/0194262X.2020.1759479>
- Sukardi, S., Susanty, M., Irawan, A., & Putra, R. F. (2020a). Low Complexity Named-Entity Recognition for Indonesian Language using BiLSTM-CNNs. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, 137–142. <https://doi.org/10.1109/ICOIACT50329.2020.9331989>

- Sukardi, S., Susanty, M., Irawan, A., & Putra, R. F. (2020b). Low Complexity Named-Entity Recognition for Indonesian Language using BiLSTM-CNNs. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, 137–142. <https://doi.org/10.1109/ICOIACT50329.2020.9331989>
- Susanto, H. D., & Yuniarto, B. (2023). Analisis Pola Kecelakaan Lalu Lintas Menggunakan Algoritma Decision Tree Berdasarkan Ekstraksi Informasi dari Berita Online Menggunakan Named Entity Recognition (NER). *Seminar Nasional Official Statistics*, 2023(1), 213–224. <https://doi.org/10.34123/semnasoffstat.v2023i1.1751>
- Suwarningsih, W., Supriana, I., & Purwarianti, A. (2014). ImNER Indonesian medical named entity recognition. *2014 2nd International Conference on Technology, Informatics, Management, Engineering & Environment*, 184–188. <https://doi.org/10.1109/TIME-E.2014.7011615>
- Syahrani, G., Sevira, S., & Yunizar Pratama Yusuf, A. (2024). RANCANGAN CHATBOT REKOMENDASI COFFEE SHOP JABODETABEK DENGAN MENGGUNAKAN DIALOGFLOW NATURAL LANGUAGE PROCESSING. *SKANIKA: Sistem Komputer Dan Teknik Informatika*, 7(1), 74–84. <https://doi.org/10.36080/skanika.v7i1.3139>
- Widiyanti, N. F., Sukmana, H. T., Hulliyah, K., Khairani, D., & Oh, L. K. (2023). Improving Indonesian Named Entity Recognition for Domain Zakat Using Conditional Random Fields. *Jurnal Online Informatika*, 8(2), 131–138. <https://doi.org/10.15575/join.v8i2.898>
- Wildannissa Pinasti, & Lya Hulliyyatus Suadaa. (2024). Named Entity Recognition in Statistical Dataset Search Queries. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 13(3), 171–177. <https://doi.org/10.22146/jnteti.v13i3.11580>
- Wintaka, D. C., Bijaksana, M. A., & Asror, I. (2019). Named-Entity Recognition on Indonesian Tweets using Bidirectional LSTM-CRF. *Procedia Computer Science*, 157, 221–228. <https://doi.org/10.1016/j.procs.2019.08.161>
- Yanti, R. M., Santoso, I., & Suadaa, L. H. (2021a). Application of Named Entity Recognition via Twitter on *SpaCy* in Indonesian (Case Study : Power Failure in the Special Region of Yogyakarta). *Indonesian Journal of Information Systems*, 76–86. <https://doi.org/10.24002/ijis.v4i1.4677>
- Yanti, R. M., Santoso, I., & Suadaa, L. H. (2021b). Application of Named Entity Recognition via Twitter on *SpaCy* in Indonesian (Case Study : Power Failure in the Special Region of Yogyakarta). *Indonesian Journal of Information Systems*, 76–86. <https://doi.org/10.24002/ijis.v4i1.4677>
- Yulianti, A., Andika, B. P., & Labellapansa, A. (2019). Application of Batu Belah Batu Bertangkup Folklore In Riau Province With Augmented Reality. *2019 23rd International Computer Science and Engineering Conference (ICSEC)*, 60–64. <https://doi.org/10.1109/ICSEC47112.2019.8974761>
- Yunefri, Y., Fadrial, Y. E., & Sutejo, S. (2021). Chatbot Pada Smart Cooperative Oriented Problem Menggunakan Natural Language Processing dan Naive Bayes Classifier. *INTECOMS: Journal of Information Technology and Computer Science*, 4(2), 131–140. <https://doi.org/10.31539/intecom.v4i2.2704>

- Zahra, S. N., & Utomo, P. E. P. (2023). Visualisasi Data Penjualan Barang Retail di Seluruh Dunia Menggunakan Tableau. *Jurnal Nasional Ilmu Komputer*, 4(3), 12–21. <https://doi.org/10.47747/jurnalnik.v4i3.1217>
- Zyam, N. S. S., & Umam, N. K. (2022). ANALISIS KETERAMPILAN MEMIRSA PADA VIDEO PEMBELAJARAN CERITA RAKYAT MELALUI WHATSAPP. *COLLASE (Creative of Learning Students Elementary Education)*, 5(4), 645–652. <https://doi.org/10.22460/collase.v5i4.11334>