

## PROTOTYPE SISTEM PENDUKUNG KEPUTUSAN TERINTEGRASI MODEL NER UNTUK VALIDASI DAN PENETAPAN PEMUKTAHIRAN DATA ASN

Nur Muhamad Holik<sup>\*1</sup>, Surya Sumpeno<sup>2</sup>, Reza Fuad Rachmadi<sup>3</sup>

<sup>1,2,3</sup>Institut Teknologi Sepuluh Nopember, Surabaya  
Email: <sup>1</sup>nurmuhamadh@gmail.com, <sup>2</sup>surya@its.ac.id, <sup>3</sup>fuad@its.ac.id  
<sup>\*</sup>Penulis Korespondensi

(Naskah masuk: 04 September 2024, diterima untuk diterbitkan: 19 Juni 2025)

### Abstrak

Untuk mendukung dan memperlancar penyelenggaraan manajemen aparatur sipil negara serta pengambilan keputusan yang efisien, efektif, dan akurat, diperlukan data pegawai ASN. Data tersebut harus dimutakhirkan dan divalidasi secara berkala sebelum disebarluaskan dan diakses oleh instansi pemerintah sesuai kewenangannya masing-masing serta dapat diakses oleh masyarakat melalui portal data sesuai dengan ketentuan peraturan perundang-undangan. Pada paper ini ditampilkan prototipe berbasis web untuk menunjukkan bahwa model NER yang dikembangkan dapat diintegrasikan sebagai subsistem dari sistem pendukung keputusan dalam melakukan validasi dan penetapan persetujuan pemutakhiran data ASN. Prototipe menunjukkan tingkat kemiripan hasil prediksi model dengan data yang diusulkan, tertinggi sebesar 100% dan terendah sebesar 41,34%. Pengukuran kinerja model menggunakan *spacy* menunjukkan bahwa model terbaik memperoleh nilai F1-score rata-rata sebesar 99,01 menggunakan dataset training, 98,20 menggunakan dataset testing, dan 94,26 menggunakan dataset *other*.

**Kata kunci:** Aparatur Sipil Negara; Named Entity Recognition; Sistem Pendukung Keputusan; Optical Character Recognition.

## PROTOTYPE INTEGRATED DECISION SUPPORT SYSTEM WITH NER MODEL FOR VALIDATION AND DETERMINATION OF ASN DATA UPDATES

### Abstract

To support and facilitate the implementation of state civil apparatus management and efficient, effective, and accurate decision-making, ASN employee data are required. The data must be updated and validated periodically before being disseminated and accessed by government agencies according to their respective authorities and can be accessed by the public through a data portal in accordance with the provisions of laws and regulations. This paper presents a web-based prototype to demonstrate that the proposed NER model can be integrated as a subsystem of a decision support system to validate and determine approval for ASN data updates. The prototype shows the level of similarity between the model's prediction results and the proposed data, with the highest value being 100% and the lowest being 41.34%. The measurement of model performance using *spacy* demonstrated that the best model obtained an average F1-score of 99.01 using the training dataset, 98.20 using the testing dataset, and 94.26 using the other dataset.

**Keywords:** State Civil Apparatus; Named Entity Recognition; Decision Support System; Optical Character Recognition

### 1. PENDAHULUAN

Badan Kepegawaian Negara (BKN) bertugas membina dan mengoordinasikan pengelolaan aparatur sipil negara (ASN) nasional. Pengelolaan ASN menghasilkan pegawai ASN yang kompeten, memiliki nilai-nilai dan berperilaku etis, serta terhindar dari pengaruh politik dan perilaku koruptif, termasuk kolusi dan nepotisme.

Untuk mendukung dan memperlancar pelaksanaan pengelolaan ASN serta pengambilan

keputusan yang efisien, efektif, dan akurat, diperlukan data pegawai ASN yang wajib dimutakhirkan dan divalidasi secara berkala sebelum disebarluaskan dan diakses oleh instansi pemerintah sesuai kewenangannya serta dapat diakses oleh masyarakat dengan memanfaatkan portal data sesuai ketentuan peraturan perundang-undangan (BKN, 2023).

Validasi dan persetujuan penetapan pemutakhiran data dilakukan oleh pegawai BKN Pusat dan pegawai BKN pada Kantor Regional di

seluruh Indonesia sesuai wilayah kerjanya dengan cara memeriksa dokumen yang telah diterbitkan oleh pejabat yang berwenang dan kebenaran usulan pemutakhiran data yang telah dilakukan. Salah satu data pegawai ASN yang wajib divalidasi adalah dokumen dan data surat keputusan tentang kenaikan pangkat.

Validasi dan persetujuan penetapan pemutakhiran data dilakukan dengan mencermati dan memperhatikan informasi entitas penting yang terdapat dalam dokumen surat keputusan kenaikan pangkat seperti NIP ASN, nama ASN, nomor persetujuan BKN, tanggal persetujuan BKN, nomor surat keputusan kenaikan pangkat, dan tanggal surat keputusan kenaikan pangkat. Entitas penting pada dokumen tersebut akan dicocokkan dengan kebenaran usulan pemutakhiran data yang telah diajukan oleh masing-masing ASN.

Berdasarkan laporan terakhir yang dibuat pada tanggal 25 November 2022, dapat disebutkan bahwa data pegawai ASN yang wajib divalidasi dan disetujui yang telah menjadi beban kerja Kantor Wilayah XI Badan Kepegawaian Negara sebanyak 260.269 data, dan data tersebut tersebar pada tiga provinsi yaitu Gorontalo, Maluku Utara, dan Sulawesi Utara. Data pegawai ASN tersebut telah divalidasi dan disetujui oleh seluruh pegawai pada Kantor Regional XI BKN sesuai dengan beban kerja yang diberikan oleh Kepala Kantor. Validasi data pegawai ASN dimulai pada tanggal 12 September 2022 dan dilakukan selama 12 (dua belas) jam per hari (Kanreg XI BKN, 2022).

Dalam *Natural Language Processing* (NLP), mengidentifikasi dan mengkategorikan entitas seperti orang, tempat, organisasi, dan tanggal disebut pengenalan *Named Entity Recognition* (NER). Entitas menyimpan data penting dalam dokumen yang berisi informasi pribadi dan bisnis (Zhang, 2013). Tugas NLP, termasuk ekstraksi informasi, sangat bergantung pada NER (Royesh, 2024), pengambilan informasi penting (Kanhaiya et al., 2023), peringkasan teks (Thottempudi, 2021), dan tanya jawab (Mishra et al., 2019).

Penelitian sebelumnya yang telah difokuskan membahas tentang NER pada beberapa aplikasi, seperti spesimen herbarium, untuk secara otomatis memperoleh data entitas dari gambar spesimen herbarium, sehingga membantu entri data ke dalam database spesimen (Takano et al., 2024), menggunakan pendekatan berbasis aturan untuk mendapatkan nama-nama nabi, individu, ciptaan, lokasi, dan berbagai gelar Tuhan sebagai komponen penting dalam domain teks Al-Qur'an (Tarmizi et al., 2022) dan mengidentifikasi makanan halal berdasarkan bahan-bahan seperti Halal, Haram, dan Syubhat (ragu-ragu), yang dapat membantu pelanggan muslim dalam mengidentifikasi produk makanan halal yang dapat dikonsumsi (Khairani et al., 2022).

Dalam penelitian ini, model NER akan dimanfaatkan pada dokumen keputusan kenaikan pangkat ASN yang diterbitkan oleh pejabat yang berwenang, seperti Presiden, Menteri, Kepala Badan atau Lembaga, Gubernur, Walikota, dan Bupati. Untuk dapat mengidentifikasi dan mengenali entitas penting seperti nomor surat keputusan (NOMOR\_SK), nomor persetujuan BKN (NOMOR\_LG\_BKN), tanggal persetujuan BKN (TANGGAL\_LG\_BKN), nama ASN (NAMA\_PNS), nomor induk pegawai ASN (NIP\_PNS), tanggal mulai berlakunya pangkat baru (TMT\_KP), dan nama pangkat baru (PANGKAT\_BARU), sehingga model NER dapat diintegrasikan sebagai salah satu bagian pada sistem pendukung keputusan dalam mendukung proses validasi dan penetapan pemutakhiran data yang telah diajukan oleh masing-masing ASN.

Sistem pendukung keputusan (SPK) adalah sistem komputer yang menawarkan informasi kepada pengguna dan menganalisisnya berdasarkan kebutuhan pengguna dalam pengambilan keputusan. Sistem pendukung keputusan menyediakan data, informasi, dan analisis penting bagi pengguna untuk memfasilitasi pengambilan keputusan yang efektif. Dalam skenario yang tidak terstruktur atau tidak dapat diprediksi, sistem pendukung keputusan membantu pengguna mengelola informasi dan membuat keputusan berdasarkan data (Khadir, 2023).

Sistem pendukung keputusan memiliki tiga komponen atau subsistem utama yang menentukan kemampuan teknis sistem pendukung keputusan, termasuk subsistem data, yang merupakan bagian dari sistem pendukung keputusan sebagai penyedia informasi; subsistem model, yang menampilkan hasil prediksi; dan subsistem dialog, yang memungkinkan pengguna berinteraksi dengan sistem melalui antarmuka dialog.

Dengan menggunakan model NER, entitas dapat diidentifikasi secara akurat, sehingga dapat meningkatkan kinerja validasi dan penetapan pemutakhiran data data ASN.

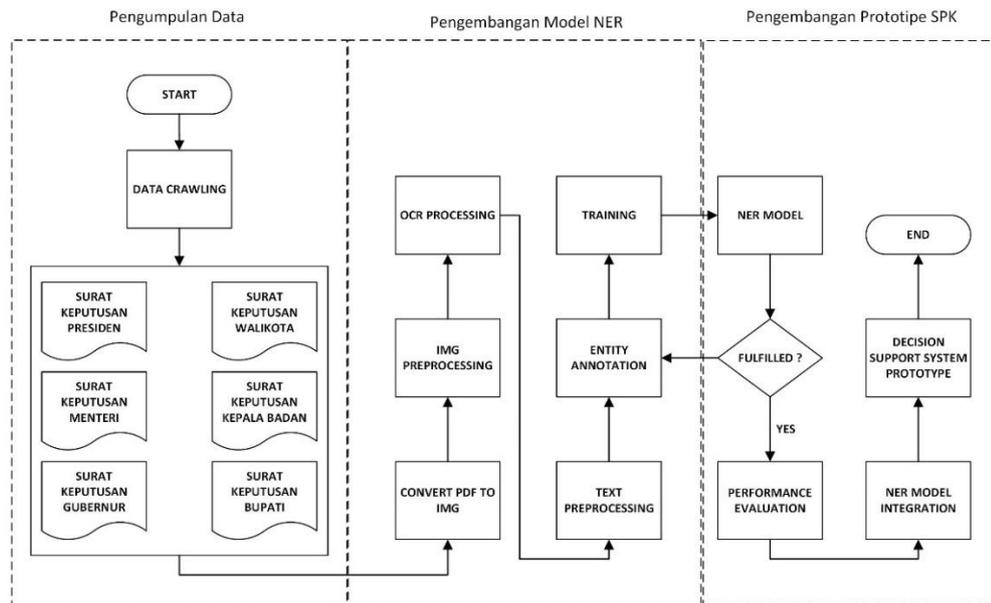
## 2. METODE PENELITIAN

Alur kerja penelitian dirangkum dalam blok diagram. Diagram ini akan membantu peneliti pada setiap tahap penelitian. Gambar 1 mengilustrasikan alur penelitian yang terdiri dari tiga tahapan utama yaitu pengumpulan data, pengembangan model NER dan pembuatan prototipe sistem pendukung keputusan.

### 2.1 Data Crawling

Data dapat dikumpulkan dan diindeks secara otomatis dari berbagai sumber, termasuk situs web, basis data, dan dokumen. Program komputer yang disebut sebagai "*web crawler*" atau "*spider*" bertanggung jawab untuk menelusuri halaman web, mengumpulkan data, dan mengaturnya secara sistematis. Fungsi utama perayapan data adalah untuk memfasilitasi akses ke informasi oleh mesin pencari

dan aplikasi melalui pengembangan, analisis, dan pengindeksan (Ani et al., 2023).



Gambar 1. Alur Penelitian

Data yang dikumpulkan melalui data crawling adalah dokumen terkait keputusan kenaikan pangkat ASN yang bersumber dari Sistem Informasi ASN (SIASN). SIASN merupakan serangkaian dokumen dan data terkait ASN yang disusun secara sistematis, komprehensif, dan terintegrasi dengan teknologi. Dokumen tersebut adalah surat yang telah diterbitkan oleh pejabat berwenang, seperti Presiden, Menteri, Kepala Badan/Lembaga, Gubernur, Wali Kota, dan Bupati, dokumen dikumpulkan dari beberapa instansi pemerintah yang telah diterbitkan oleh pejabat yang berwenang mulai tahun 2023 sampai dengan tahun 2024 berdasarkan data ASN yang bekerja pada wilayah kerja Kantor Regional XI BKN. Dokumen yang diperoleh kemudian disimpan secara offline atau lokal dalam bentuk file PDF dengan format nama file SK\_KP\_TAHUN\_NIPASN.

### 2.2 Convert PDF To Image

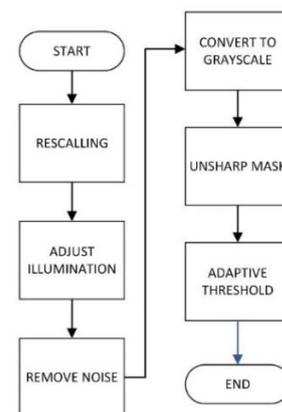
Pada tahap ini, setiap halaman dokumen surat keputusan kenaikan pangkat diubah menjadi file citra menggunakan pustaka "pdf2image". Hal ini diperlukan untuk memudahkan manipulasi citra menggunakan berbagai teknik pemrosesan citra daripada menggunakan berkas PDF. citra dapat dengan mudah diubah ukurannya, dipotong, atau disesuaikan kecerahannya dan kontrasnya, yang sering kali diperlukan sebelum diproses lebih lanjut.

### 2.3 Image Preprocessing

Membaca teks pada dokumen hasil pemindaian terkadang menimbulkan masalah seperti dokumen buram, teks memudar, terdapat tanda air, stempel, tanda tangan, dan bekas lipatan pada dokumen. Masalah seperti ini dapat menyebabkan penurunan kualitas *optical character recognition* (OCR) dan

masalah OCR lainnya. Untuk meningkatkan kualitas dokumen hasil pemindaian dan meningkatkan hasil OCR, diperlukan pra-pemrosesan citra (Ramdhani et al., 2021).

Pada penelitian ini dilakukan prapemrosesan citra menggunakan pustaka *OpenCV* dengan enam langkah algoritma yaitu *resizing* dengan faktor skala 4 dan metode *INTER\_LINEAR*, *adjust illumination*, *remove noise*, *unsharp mask*, dan *adaptive threshold*, prapemrosesan citra dilakukan sebelum proses OCR untuk meningkatkan kualitas pengenalan karakter. Alur kerja prapemrosesan citra diilustrasikan pada Gambar 3.

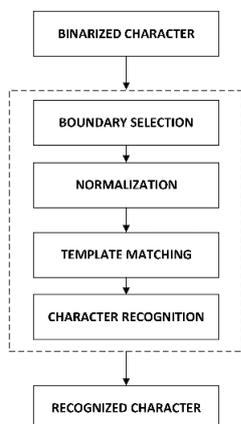


Gambar 2. Alur prapemrosesan citra

### 2.4 OCR Processing

Metode yang dikenal sebagai OCR mengubah gambar dalam bentuk teks menjadi teks terstruktur yang dapat dikenali oleh mesin (Wahyuddin et al., 2023). OCR banyak digunakan untuk membaca teks dari gambar yang dipindai dan juga digunakan untuk membaca teks dari dokumen identitas seperti kartu identitas atau paspor.

Proses klasifikasi karakter dilakukan secara berurutan dengan OCR (Mollah et al., 2011), alur kerja *OCR Processing* ditampilkan pada Gambar 4.



Gambar 3. Alur OCR Processing

*Binarized character* adalah berkas citra digital yang diubah menjadi citra biner. Citra biner hanya memiliki dua tingkat keabuan, hitam dan putih. Piksel objek bernilai 1, dan piksel latar belakang bernilai 0. Saat citra ditampilkan, 1 mewakili hitam dan 0 mewakili putih. Oleh karena itu, dalam citra biner, latar belakang berwarna putih, dan objek berwarna hitam.

*Boundary selection* bertujuan untuk mengelompokkan piksel objek ke dalam wilayah yang mewakili suatu objek. Batas antara objek dan latar belakang terlihat jelas dalam citra biner. Piksel objek dan latar belakang berwarna hitam, dan piksel latar belakang berwarna putih. Perpotongan antara piksel hitam dan putih dimodelkan sebagai segmen garis.

*Normalization* merupakan tahap dalam pra-proses citra yang dilakukan sebelum pengenalan. Proses ini menyesuaikan data citra masukan dengan data citra dalam basis data.

*Template Matching* merupakan teknik pemrosesan citra digital yang mencocokkan setiap bagian citra dengan citra referensi. Citra masukan dibandingkan dengan citra template dalam basis data, dan kesamaan ditentukan menggunakan aturan tertentu. Pencocokan citra yang menghasilkan kesamaan tinggi menunjukkan bahwa citra dikenali sebagai citra template.

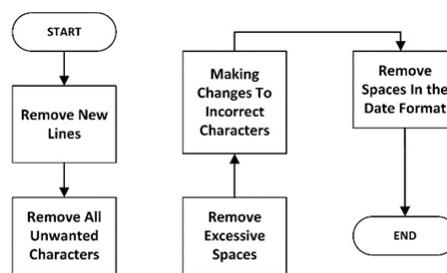
*Character recognition* adalah proses mengenali karakter yang diamati dengan membandingkan ciri-ciri karakter yang diperoleh dengan ciri-ciri karakter dalam basis data.

*OCR Processing* pada penelitian ini menggunakan *Tesseract-OCR*, sebuah platform *OCR* sumber terbuka untuk melakukan ekstraksi teks dari gambar dan dokumen. Teknologi ini dapat mengidentifikasi dan membaca teks dari berbagai gambar, seperti dokumen dan foto. *Tesseract* besutan *Google*, yang awalnya dibuat oleh *Hewlett-Packard*, beroperasi dalam berbagai bahasa dan sistem penulisan, termasuk *Unicode (UTF-8)*.

## 2.5 Text Preprocessing

Berbagai teknik *text-preprocessing* berkontribusi pada keberhasilan *NER*. Penulis menyatakan bahwa penggunaan satu atau kombinasi prosedur *text-preprocessing*, seperti kontraksi ekspansi, konversi angka ke kata, dan pemisahan koma, tanda hubung, dapat meningkatkan kinerja model. Dapat dilihat bahwa model *conditional random fields (CRF)* tanpa *text-preprocessing* dan model *CRF* dengan *text-preprocessing* lebih baik saat menjalankan tugas *NER* (Situmeang, 2021).

*Natural Language Processing (NLP)* sangat bergantung pada tahap *text-preprocessing*. Gambar 5 menampilkan alur *text-preprocessing*



Gambar 4. Alur text-preprocessing

## 2.6 Entity Annotation

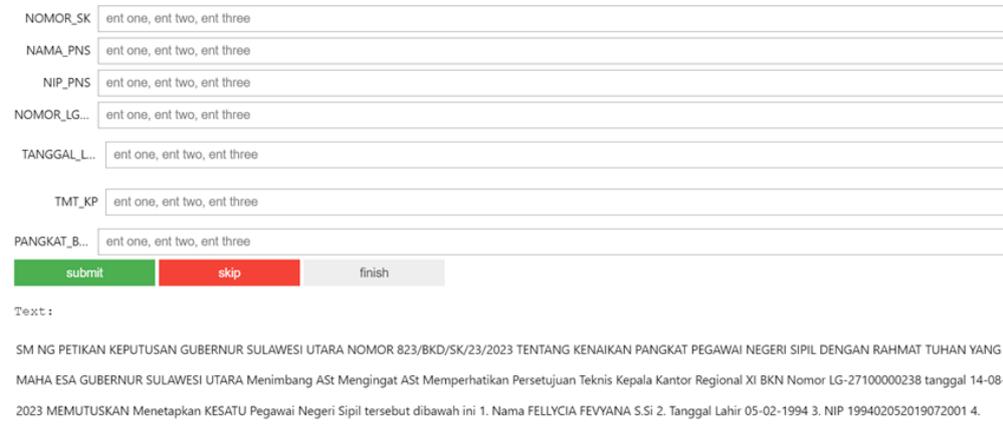
Anotasi entitas, yang juga disebut sebagai pelabelan data, adalah teknik pelabelan data yang dapat diterapkan pada teks, foto, atau video. Pelabelan data adalah proses pemberian satu atau beberapa label deskriptif pada data yang belum diproses (gambar, teks, video, dll.) sehingga model *machine learning* dapat menggunakannya untuk belajar. Hal ini dilakukan untuk memudahkan model memahami sumber data tertentu dan mengidentifikasi objek, format, data, atau pola tertentu di masa mendatang berdasarkan pelabelan data (Kumar et al., 2022).

Berbagai metode anotasi telah disarankan dalam mengidentifikasi entitas. Skema anotasi dan pengkodean diterapkan untuk menetapkan identitas dan lokasi entitas dalam kalimat tertentu. Dalam *NER*, tujuannya adalah untuk menetapkan label sekuens yang sesuai ke sekuens teks tertentu. Dalam skema anotasi entitas bernama, token pertama ditandai(B), token terakhir dari entitas bernama ditandai(E), sedangkan token perantara ditandai(I). Token non-entitas dilambangkan sebagai(O). Skema tag *NER* dapat terdiri dari empat tag sebagai berikut: *BIO*, *IO*, *IOE*, *IOBES*, *IE*, dan *BIES* (Keraghel et al., 2024). Pada *NER Arab* (Alshammari et al., 2021) penulis menunjukkan bahwa skema tag *IO*, merupakan skema paling sederhana saat dipelajari, mencapai pengukuran *F1-score* tertinggi.

Pada penelitian ini, anotasi entitas menggunakan tools "*spacy-annotator*". *spacy-annotator* digunakan untuk membuat data beranotasi yang berisi teks dan label entitas. *spacy-annotator*

dapat menyederhanakan proses anotasi data secara signifikan karena hanya memerlukan input teks yang

berisi entitas dalam kotak teks yang disediakan sesuai dengan jenis entitas.



Gambar 5. Tools spacy-annotator

Hal ini memudahkan dan lebih efisien untuk mengembangkan model NER untuk NLP dalam domain machine learning. Data beranotasi kemudian akan disimpan dalam format "spacy" yang akan digunakan untuk pelatihan, pengujian, dan evaluasi model. Gambar 6 menampilkan tools spacy-annotator.

## 2.7 Training

Pengetahuan yang sesuai untuk model *machine learning* ditentukan selama training. Model *machine learning* belajar untuk melakukan generalisasi melalui proses training ini. Metode *machine learning* bervariasi dalam prosedur training-nya. Model pembelajaran mesin dapat dilatih menggunakan *reinforcement*, *unsupervised*, dan *supervised learning* (Elwirehardja et al., 2023).

Pada penelitian (Kotevska, 2021), *semi-supervised learning* digunakan untuk melatih model NER, yaitu *joint training*, dimana beberapa label data dilatih. Kemudian, secara periodik, data yang telah dilatih sebelumnya dapat memberikan label secara instan pada data yang belum diberi label, sehingga membuat model lebih reliabel (menghasilkan hasil yang sama pada setiap percobaan yang berhasil).

## 2.8 NER Model

Model NER yang telah menjalani proses training sehingga dapat digunakan untuk mengidentifikasi bagian teks yang dianggap entitas dan mengklasifikasikan entitas ke dalam tipe entitas yang ditentukan sebelumnya dalam teks tidak terstruktur. Model NER inilah yang dipergunakan juga dalam membantu saat proses pelabelan.

## 2.9 Performance Evaluation

Dalam konteks *machine learning*, metrik evaluasi digunakan untuk mengukur sejauh mana model dapat mempelajari pola yang ada dalam data pelatihan dan menerapkannya pada data yang sebelumnya tidak terlihat (data validasi dan pengujian). Beberapa pengukuran NER yang umum

digunakan melibatkan penggunaan metrik evaluasi seperti *recall*, *F1-score*, dan *precision* (Keraghel et al., 2024).

Pada penelitian ini, evaluasi model dilakukan melalui *pipeline spacy* dengan menggunakan perintah "*spacy benchmark*" pada command line.

## 2.10 NER Model Integration

Sistem terintegrasi merupakan gabungan dari beberapa sistem komputerisasi dan perangkat lunak. Sistem terintegrasi menyatukan fungsi komponen-komponen subsistem menjadi satu kesatuan. Integrasi data dari sistem sangatlah penting dikarenakan beberapa hal berikut (Yuliana, 2022):

1. Kolaborasi antar pemangku kepentingan.
2. Pemrosesan data dalam sistem terpisah memerlukan pertukaran data untuk memperoleh informasi yang komprehensif.
3. Menawarkan akses instan dan *realtime*.

Pada penelitian ini Model NER akan diintegrasikan sebagai subsistem manajemen model sehingga dapat memberikan prediksi dalam melakukan validasi dan penetapan persetujuan data pegawai ASN berdasarkan surat keputusan kenaikan pangkat yang telah diterbitkan oleh pejabat berwenang.

## 2.11 Prototipe Decision Support System

Sistem Pendukung Keputusan (SPK) biasanya dikembangkan menggunakan pendekatan pembuatan prototipe (Turban et al., 2005). Dalam pengembangan perangkat lunak, prototipe dapat digunakan dalam berbagai cara (Sommerville, 2011):

1. Prototipe berperan penting dalam proses rekayasa persyaratan untuk memperoleh dan memvalidasi persyaratan sistem.
2. Dalam proses desain, prototipe digunakan untuk menyelidiki solusi perangkat lunak dan berkontribusi pada antarmuka pengguna.

- Selama pengujian pengguna, prototipe berfungsi sebagai pengganti sistem yang sebenarnya.

SM NG PETIKAN KEPUTUSAN GUBERNUR SULAWESI UTARA NOMOR 823/BKD/SK/23/2023 TENTANG KENAIKAN PANGKAT PEGAWAI NEGERI SIPIL DENGAN RAHMAT TUHAN YANG MAHA ESA GUBERNUR SULAWESI UTARA Menimbang AST Mengingat AST Memperhatikan Persetujuan Teknis Kepala Kantor Regional XI BKN Nomor LG-27100000159 tanggal 04-08-2023 MEMUTUSKAN Menetapkan KESATU Pegawai Negeri Sipil tersebut dibawah ini 1. Nama WIRDAWATI PAPUTUNGAN A.Md. 2. Tanggal Lahir 24-09-1971 3. NIP 197109242015042001 4. Pendidikan D-III MANAJEMEN KEUANGAN DAN PERBANKAN 5. Pangkat/Gol.ruang lama/TMT Pengatur Tingkat I / II/d / 01-04-2019 6. Jabatan PRANATA KEARSIPAN / . Unit Kerja DINAS PENDIDIKAN DAERAH 8. Instansi Induk PEMERINTAH PROVINSI SULAWESI UTARA Terhitung mulai tanggal 01-10-2023 dinaikkan pangkatnya menjadi Penata Muda golongan ruang III/a dengan masa kerja golongan 16 Tahun 9 Bulan dan diberikan gaji pokok sebesar Rp. 3.305.700 ditambah dengan penghasilan lain berdasarkan ketentuan peraturan perundang-undangan yang berlaku. KEDUA Apabila dikemudian hari ternyata terdapat kekeliruan dalam keputusan ini akan diadakan perbaikan dan penghitungan kembali sebagaimana mestinya. PETIKAN Keputusan ini diberikan kepada yang bersangkutan untuk diketahui dan dipergunakan sebagaimana mestinya. Tembusan Keputusan ini disampaikan kepada 1. Kepala Kantor Regional XL BKN di Manado 2. Kepala BKAD Provinsi Sulawesi Utara 3. Pertinggal. Ditetapkan di Manado Pada tanggal 18 September 2023 MEA Rena N Aa AA GUBERNUR SULAWESI UTARA AAN Ma TTD Sa Ma OLLY DONDOKAMBEY Untuk petikan yang sah KEPALA BADAN KEPEGAWAIAN DAERAH PROVINSI SULAWESI UTARA Dr. JEMMY S. KUMENDONG M.Si PEMBINA UTAMA MADYA NIP. 197006121990101001

Gambar 6. Hasil OCR

E	#	LOSS	TOK2VEC	LOSS	NER	ENTS_F	ENTS_P	ENTS_R	DEP_LAS	TAG_ACC	TOKEN_ACC	SPEED	SCORE
0	0	0.00	257.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	2249.47	0.00
0	200	4195.68	8875.55	82.39	76.69	89.01	0.00	0.00	100.00	2484.93	0.00	0.33	
0	400	423.33	659.46	96.39	97.18	95.61	0.00	0.00	100.00	2477.55	0.00	0.39	
0	600	33.36	158.81	95.53	96.42	94.66	0.00	0.00	100.00	2239.78	0.00	0.38	
0	800	78.47	189.47	97.10	96.99	97.20	0.00	0.00	100.00	2596.42	0.00	0.39	
0	1000	79.35	163.07	97.80	97.86	97.75	0.00	0.00	100.00	2324.95	0.00	0.39	
0	1200	503.86	430.36	97.64	97.38	97.91	0.00	0.00	100.00	2424.25	0.00	0.39	
0	1400	108.39	182.37	95.57	93.97	97.22	0.00	0.00	100.00	2337.17	0.00	0.38	
1	1600	61.20	123.31	97.60	97.88	97.32	0.00	0.00	100.00	2596.09	0.00	0.39	
1	1800	198.93	155.84	97.27	97.38	97.15	0.00	0.00	100.00	2481.18	0.00	0.39	
1	2000	115.24	98.43	97.61	97.44	97.77	0.00	0.00	100.00	2320.35	0.00	0.39	
1	2200	86.81	112.25	97.88	97.66	98.10	0.00	0.00	100.00	2553.66	0.00	0.39	
1	2400	186.52	139.22	97.12	97.42	96.82	0.00	0.00	100.00	3911.29	0.00	0.39	
1	2600	430.83	125.85	97.51	97.55	97.46	0.00	0.00	100.00	2259.56	0.00	0.39	
1	2800	88.61	71.83	97.06	95.50	98.67	0.00	0.00	100.00	2466.79	0.00	0.39	
2	3000	138.07	91.84	97.80	97.14	98.46	0.00	0.00	100.00	2352.61	0.00	0.39	
2	3200	549.62	103.37	97.95	97.13	98.79	0.00	0.00	100.00	2441.07	0.00	0.39	
2	3400	142.93	104.48	97.76	96.94	98.60	0.00	0.00	100.00	2446.75	0.00	0.39	
2	3600	101.20	92.71	97.11	98.30	95.94	0.00	0.00	100.00	2317.62	0.00	0.39	
2	3800	264.09	113.95	97.80	97.45	98.15	0.00	0.00	100.00	2458.63	0.00	0.39	
2	4000	261.08	83.16	97.77	97.10	98.46	0.00	0.00	100.00	2342.40	0.00	0.39	
2	4200	197.70	86.94	98.09	97.81	98.39	0.00	0.00	100.00	2440.47	0.00	0.39	
3	4400	81.68	68.40	97.99	98.42	97.56	0.00	0.00	100.00	2404.81	0.00	0.39	
3	4600	127.94	69.97	98.20	98.59	97.82	0.00	0.00	100.00	2405.91	0.00	0.39	
3	4800	156.00	60.78	98.12	97.58	98.67	0.00	0.00	100.00	2379.05	0.00	0.39	
3	5000	153.74	69.65	98.11	97.87	98.34	0.00	0.00	100.00	2167.83	0.00	0.39	
3	5200	471.01	91.93	97.97	97.87	98.08	0.00	0.00	100.00	2490.37	0.00	0.39	
3	5400	381.40	114.16	97.96	97.94	97.98	0.00	0.00	100.00	2403.40	0.00	0.39	
3	5600	157.86	97.19	97.95	97.98	97.91	0.00	0.00	100.00	4147.58	0.00	0.39	
4	5800	397.07	74.21	97.74	97.25	98.24	0.00	0.00	100.00	5056.80	0.00	0.39	
4	6000	238.06	82.09	98.09	97.99	98.20	0.00	0.00	100.00	4781.44	0.00	0.39	
4	6200	804.04	78.46	97.22	96.98	97.46	0.00	0.00	100.00	4723.73	0.00	0.39	

Gambar 7. Hasil Training NER

### 3. HASIL DAN PEMBAHASAN

Berdasarkan diagram alur penelitian (Gambar 1), hasil dan pembahasan secara sistematis disusun ke dalam tahapan-tahapan berikut:

Tabel 1. Dataset surat keputusan

Diterbitkan oleh	Tahun		Jumlah
	2023	2024	
Presiden	145	142	287
Menteri/Kepala Badan	384	134	518
Gubernur	316	151	467
Walikota	168	145	313
upati	297	159	456
Total			2041

#### 3.2. OCR Processing

Data File PDF surat keputusan kenaikan pangkat ASN yang terkumpul kemudian diubah menjadi file citra menggunakan pustaka "pdf2image". Selanjutnya, untuk meningkatkan kualitas citra, dilakukan preprocessing citra menggunakan pustaka "opencv" sebelum proses OCR dilakukan oleh Tesseract-OCR, kemudian hasil yang diperoleh setelah OCR disimpan secara lokal sebagai file TXT.

Gambar 9 menyajikan hasil proses OCR yang dilakukan oleh Tesseract-OCR menggunakan contoh dokumen surat keputusan kenaikan pangkat yang telah diterbitkan oleh Gubernur Pemerintah Daerah Provinsi Sulawesi Utara.

Tabel 2. Dataset Training

Diterbitkan oleh	Tahun		Jumlah
	2023	2024	
Presiden	94	105	199
Menteri/Kepala Badan	261	95	356
Gubernur	226	101	327
Walikota	116	103	219
Bupati	222	105	327
Total			1428

#### 3.3 Entity Annotation

Teks hasil OCR kemudian dibersihkan melalui tahapan text-preprocessing, selanjutnya dibagi menjadi dua bagian secara acak, 70% berfungsi sebagai dataset training, seperti ditunjukkan pada Tabel 2, sedangkan 30% data berfungsi sebagai dataset testing, seperti ditunjukkan pada Tabel 3.

Selanjutnya, proses pelabelan dilakukan menggunakan tools *spacy-annotator*. Pelabelan merupakan proses yang memakan waktu dalam pembuatan dataset untuk training model, oleh karena itu, proses pelabelan atau anotasi dalam penelitian ini menggunakan anotasi *semi-supervised*, yaitu dengan membangun model dari sebagian dataset yang telah diberi label. Kemudian, dataset yang belum diberi label akan diberi label atau anotasi oleh model yang telah terbentuk.

Tabel 3. Dataset Testing

Diterbitkn oleh	Tahun		amount
	2023	2024	
Presiden	51	37	88
Menteri/Kepala Badan	123	39	162
Gubernur	90	50	140
Walikota	52	42	94
Bupati	74	54	128
Total			613

### 3.4 Training

Training model dilakukan menggunakan notebook *Lenovo Yoga Slim 7 Pro 14ACH5* dengan Prosesor *AMD Ryzen7 5800H* dengan unit pemrosesan grafis *Radeon*, RAM terpasang 16 GB, dan Penyimpanan *SSD 1 TB*. Hasil pelatihan NER dengan *spacy* ditunjukkan pada Gambar 7. Nilai *F1-score* pada training terakhir adalah 97,22, *Recall* 97,46, *Precision* 96,98, *Speed* 4723,73 kata per detik, *Score* 0,39, dan *Loss* 78,46.

Tabel 4 menjelaskan deskripsi output setelah melakukan training model NER menggunakan *spacy*.

Tabel 4. Deskripsi training output

Nama	Deskripsi
F-Score (F)	Rata-rata harmonik presisi dan recall. Biasanya meningkat.
Precision (P)	Persentase anotasi yang diprediksi benar. Biasanya meningkat.
Recall (R)	Persentase anotasi referensi yang dipulihkan. Biasanya meningkat.
UAS / LAS	Skor lampiran berlabel dan tidak berlabel untuk pengurai ketergantungan, yaitu persentase busur yang benar. Biasanya meningkat.
Speed	Kecepatan prediksi dalam kata per detik (WPS). Biasanya tetap stabil.
Loss	Loss pelatihan yang mewakili jumlah pekerjaan yang tersisa untuk pengoptimal. Harus berkurang, tetapi biasanya tidak sampai 0

Setelah proses pelatihan, model disimpan dalam dua folder: *model-last*, yang berisi versi model pada iterasi pelatihan terakhir, yang dapat digunakan untuk melanjutkan proses pelatihan di masa mendatang, dan *model-best*, yang berisi versi model yang memiliki performa terbaik dalam proses pelatihan. Secara umum, model ini dapat digunakan untuk membuat prediksi pada data yang tidak berlabel.

Tabel 5 menyajikan dataset yang dipergunakan oleh model *machine learning* pada penelitian ini dalam mempelajari pola menggunakan dataset training dan testing, menampilkan tipe entitas, dan jumlah entitas dalam dataset. Pada dataset training jumlah token sebanyak 539.628, dan jumlah kata sebanyak 477.730. Pada dataset testing jumlah token sebanyak 232.588 dan jumlah kata sebanyak 205.987.

Tabel 5. Dataset berdasar tipe entitas

Tipe Entitas	Dataset	
	Training	Testing
NOMOR_SK	1423	612
NOMOR_LG_BKN	1434	617
TANGGAL_LG_BKN	1431	613
NAMA_PNS	1335	560
NIP_PNS	1332	563
TMT_KP	1440	616
PANGKAT_BARU	1394	599
Total	9789	4180

### 3.5 Evaluation

Kinerja model NER dinilai menggunakan matriks *precision*, *recall*, dan *F1-score* berdasarkan klasifikasi *True Positives*, *False Positives*, dan *False Negatives*. Kinerja model terbaik dievaluasi, dan hasilnya diperoleh.

Sebanyak 500 data tambahan digunakan untuk menilai kinerja model NER sebagai dataset *other* yang diambil dari usulan peremajaan atau pemutakhiran data pangkat atau golongan yang disampaikan melalui SIASN (Sistem Informasi Aparatur Sipil Negara) dan juga dari data Keputusan Kenaikan Pangkat ASN yang diterbitkan oleh pejabat berwenang sebelum tahun 2023.

Perbandingan performa model training, testing, dan *other* disajikan dalam Tabel 6. Perbandingan performa model berdasarkan tipe entitas ditunjukkan dalam Tabel 7.

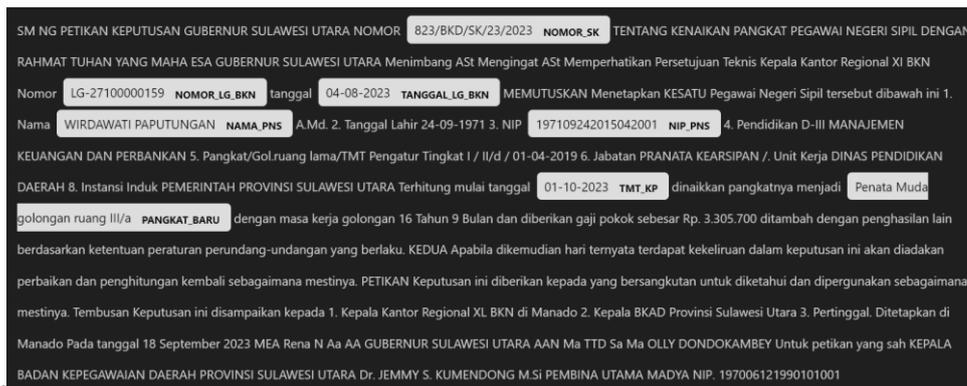
Tabel 6. Evaluasi kinerja model

Tipe Evaluasi	Training	Testing	Other
Precision	99.58	98.59	96.39
Recall	98.45	97.82	92.22
F1-Score	99.01	98.20	94.26

Gambar 8 menampilkan kinerja model NER menggunakan contoh dokumen surat keputusan kenaikan pangkat yang telah diterbitkan oleh Gubernur Pemerintah Daerah Provinsi Sulawesi Utara, dalam melakukan prediksi untuk mengenali entitas nomor surat keputusan (NOMOR\_SK), nomor persetujuan BKN (NOMOR\_LG\_BKN), tanggal persetujuan BKN (TANGGAL\_LG\_BKN), nama ASN (NAMA\_PNS), NIP ASN (NIP\_PNS), tanggal mulai berlaku pangkat baru (TMT\_KP), dan nama pangkat baru (PANGKAT\_BARU)

Tabel 7. Evaluasi kinerja berdasarkan tipe entitas

Tipe Entitas	Precision			Recall			F1-Score		
	Training	Testing	Other	Trainin g	Testing	Other	Trainin g	Testing	Other
NOMOR_SK	99.72	99.02	96.67	97.86	98.86	88.44	98.78	98.94	92.37
NOMOR_LG_BKN	99.58	99.68	98.18	100	100	99.39	99.79	99.84	98.78
TANGGAL_LG_BK N	99.09	98.53	92.10	96.73	96.95	94.06	97.89	97.73	93.07
NIP_PNS	99.70	98.05	97.33	99.25	97.18	74.13	99.48	97.61	84.16
PANGKAT_BARU	99.93	99	100	97.48	97.37	98.77	98.69	98.18	99.38
TMT_KP	99.86	99.84	96.15	99.38	99.51	98.04	99.62	99.68	97.08
NAMA_PNS	99.18	95.71	94.52	98.59	94.53	92.64	98.88	95.12	93.57
Average	99.58	98.59	96.39	98.45	97.82	92.22	99.01	98.20	94.26



Gambar 8. Kinerja Model

PREDIKSI MODEL NER  
UNTUK SURAT KEPUTUSAN KENAIKAN PANGKAT ASN

DATA USULAN		DATA PREDIKSI MODEL	
NIP	196212052014072001	NIP	196212052014072001 1.0
NAMA	MIN JAFAR	NAMA	MIN JAFAR 1.0
NOMOR SK	800.1.3.2/KEP/BU/BB/12/2/X/2022	NOMOR SK	800.1.3.2/KEP/BU/BB/12/2/X/2022 1.0
NOMOR LG BKN	LG-27504000106	NOMOR LG BKN	LG-27504000106 1.0
TANGGAL LG BKN	09-09-2022	TANGGAL LG BKN	09-09-2022 1.0
TMT KP	01-10-2022	TMT KP	01-19-2022 0.9
PANGKAT BARU	Penata golongan ruang III/c	PANGKAT BARU	Penata golongan ruang III/c 1.0

**TINGKAT KEMIRIPAN PREDIKSI MODEL DENGAN USULAN : 98.57 %**

Gambar 9. Prototipe

### 3.6 Prototype

Pada penelitian ini digunakan prototipe untuk menguji kinerja model NER seperti ditampilkan pada Gambar 9, prototipe yang dikembangkan berbasis web yang telah mengintegrasikan model NER sebagai salah satu elemen sistem pendukung keputusan (SPK) dalam melakukan validasi dan penetapan persetujuan pemutakhiran data kenaikan pangkat ASN. Dalam prototipe hasil prediksi entitas yang dihasilkan oleh model NER dibandingkan kedekatannya masing-

## 4. KESIMPULAN

Hasil penelitian menunjukkan keberhasilan model NER dalam mengenali entitas penting pada surat keputusan kenaikan pangkat ASN. Model NER dapat mengenali entitas nomor surat keputusan (NOMOR\_SK), nomor persetujuan BKN (NOMOR\_LG\_BKN), tanggal persetujuan BKN

masing entitas dengan data usulan dan akan diberi poin 1 (satu) apabila data prediksi model mirip dengan data usulan. Prototipe diujicobakan pada 50 dataset acak. Dari 50 data yang diujicobakan menggunakan pustaka "Levenshtein Ratio", tingkat kemiripan prediksi model dengan data usulan memperoleh skor tertinggi 100%, dan skor terendah 41,34%. Skor yang rendah terkadang disebabkan oleh hasil OCR yang tidak sempurna akibat dokumen yang buram, terbalik, atau bukan dokumen surat keputusan kenaikan pangkat.

(TANGGAL\_LG\_BKN), nama ASN (NAMA\_PNS), NIP ASN (NIP\_PNS), tanggal mulai berlakunya pangkat baru (TMT\_KP), dan nama pangkat baru (PANGKAT\_BARU).

Pengukuran kinerja model NER menggunakan *spacy* menunjukkan bahwa model terbaik memperoleh nilai F1-score rata-rata sebesar 99,01 menggunakan dataset training, 98,20 menggunakan dataset testing, dan 94,26 menggunakan dataset *other*.

Dalam makalah ini, prototipe berbasis web ditampilkan untuk menunjukkan bahwa model NER yang telah dikembangkan dapat diintegrasikan sebagai subsistem dari sistem pendukung keputusan yang dapat dipergunakan dalam melakukan validasi dan menetapkan persetujuan pemutakhiran data pangkat dan golongan ruang ASN.

Prototipe menampilkan persentase tingkat kemiripan hasil prediksi model NER dengan data usulan, tertinggi sebesar 100% dan terendah sebesar 41,34%.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kementerian Komunikasi dan Informatika atas bantuan dan dukungan yang diberikan kepada penulis melalui Program Beasiswa Dalam Negeri sehingga penulis dapat melanjutkan studi di Institut Teknologi Sepuluh Nopember.

#### DAFTAR PUSTAKA

- ANI, N., SINAGA, D. Y., JUNIOR, N., & MUNGgaran, M. D., 2023. Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Fitur Pencarian Dokumen Standar Nasional Indonesia. *JSAI (Journal Scientific and Applied Informatics)*, 6(3), 517-522.
- ALSHAMMARI, N., & ALANAzi, S., 2021. The impact of using different annotation schemes on named entity recognition. *Egyptian Informatics Journal*, 22(3), 295-302.
- ELWIREHARDJA, G. N., SUPARYANTO, T., DAN PARDAMEAN, B., 2023. Machine Learning Untuk Pemula. Yogyakarta : INSTIPER PRESS.
- GAO, S., KOTEVSKA, O., SOROKINE, A., & CHRISTIAN, J. B., 2021. A pre-training and self-training approach for biomedical named entity recognition. *PloS one*, 16(2), e0246310.
- KERAGHEL, I., MORBIEU, S., & NADIF, M., 2024. A survey on recent advances in named entity recognition. *arXiv preprint arXiv:2401.10825*.
- KUMAR, D., CHOUDHARI, K., PATEL, P., PANDEY, S., HAJARE, A., & JANTE, S., 2022. STAT simple text annotation tool (STAT): Web-based tool for creating training data for spaCy models. In *ICT Analysis and Applications* (pp. 299-305). Springer Singapore.
- KHADIR, ABDUL., 2023. Sistem Pendukung Keputusan. Sumatera Utara: PT. Mifandi Mandiri Digital.
- KHAIRANI, D., BANGKIT, D. A., ROZI, N. F., MASRUROH, S. U., OKTAVIANA, S., & ROSYADI, T., 2022. Named-entity recognition and optical character recognition for detecting halal food ingredients: Indonesian case study. In *2022 10th International Conference on Cyber and IT Service Management (CITSM)* (pp. 01-05). IEEE.
- KANHAIYA, K., SHARMA, A. K., GAUTAM, K., & RATHORE, P. S., 2023. AI Enabled-Information Retrieval Engine (AI-IRE) in Legal Services: An Expert-Annotated NLP for Legal Judgements. In *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAIS)* (pp. 206-210). IEEE.
- Laporan Approval Peremajaan Data Mandiri ASN Tahun 2022. Manado :Kantor Regional XI Badan Kepegawaian Negara.
- MISHRA, A., SHEKHAR, S., SINGH, A. K., & CHAKRABORTY, A., 2019. Ocr-vqa: Visual question answering by reading text in images. In *2019 international conference on document analysis and recognition (ICDAR)* (pp. 947-952). IEEE.
- MOLLAH, A. F., MAJUMDER, N., BASU, S., & NASIPURI, M., 2011. Design of an optical character recognition system for camera-based handheld devices. *arXiv preprint arXiv:1109.3317*.
- Peraturan Badan Kepegawaian Negara Nomor 7 Tahun 2023 tentang Sistem Informasi Aparatur Sipil Negara. Jakarta : Badan Kepegawaian Negara.
- RAMDHANI, T. W., BUDI, I., & PURWANDARI, B., 2021. Optical character recognition engines performance comparison in information extraction. *International Journal of Advanced Computer Science and Applications*, 12(8).
- ROYESH, A., & OLADEJI, O., 2024. Information Extraction: An application to the domain of hyper-local financial data on developing countries. *arXiv preprint arXiv:2403.09077*.
- SITUMEANG, S., 2022. Impact of text preprocessing on named entity recognition based on conditional random field in Indonesian text. *Jurnal Mantik*, 6(1), 423-430.
- SOMMERVILLE, I., 2011. *Software Engineering*. United States of America: Pearson Education, Inc
- TAKANO, A., COLE, T. C., & KONAGAI, H., 2024. A novel automated label data extraction and data base generation system from herbarium specimen images using OCR and NER. *Scientific Reports*, 14(1), 112.
- TARMIZI, S. A., & SAAD, S., 2022. NAMED ENTITY RECOGNITION FOR QURANIC TEXT USING RULE BASED APPROACHES. *Asia-Pacific Journal of Information Technology & Multimedia*, 11(2).
- THOTTEMPUDI, S. G., 2021. A visual narrative of ramayana using extractive summarization topic modeling and named entity recognition.

- In CEUR Workshop Proc. (Vol. 2823, pp. 3-10).
- TURBAN, E. AND ARONSON, J.E. and LIANG, T.P., 2005. *Decision Support and Intelligent System*. Pearson/Prentice Hall.
- WAHYUDDIN, W., & HASIM, A., 2023. Aplikasi Ekstraksi Data Kartu Vaksin Berbasis Web Menggunakan Metode Ocr. *Jurnal Sintaks Logika*, 3(2), 53-57.
- YULIANA, R., 2022. *Integrasi Aplikasi dan Informasi*, Jawa Barat: Widina Bhakti Persada Bandung.
- ZHANG, Z., 2013. *Named entity recognition: challenges in document annotation, gazetteer construction and disambiguation* (Doctoral dissertation, University of Sheffield).