

Joseph Dedy irawan

17224

 Supporting Children with Special Needs: Early Identification and Intervention

Document Details

Submission ID

trn:oid::3117:546055638

Submission Date

Jan 14, 2026, 11:42 AM GMT+7

Download Date

Jan 14, 2026, 11:45 AM GMT+7

File Name

17224.pdf

File Size

624.2 KB

6 Pages

2,705 Words

17,395 Characters

9% Overall Similarity

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

Filtered from the Report

- Bibliography

Top Sources

- 7%  Internet sources
 - 4%  Publications
 - 7%  Submitted works (Student Papers)
-

Top Sources

- 7% Internet sources
- 4% Publications
- 7% 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 Muhammadiyah Purwokerto on 2025-07-09	3%
2	Internet	journal.sinov.id	<1%
3	Publication	Ade Eviyanti, Hindarto hindarto, Sumarno, Herlian Aliyasa Alamj Duddin. "Epileps...	<1%
4	Internet	docplayer.info	<1%
5	Student papers	Universitas Muhammadiyah Makassar on 2024-01-13	<1%
6	Student papers	University of Computer Studies on 2022-11-17	<1%
7	Internet	jurnal.wicida.ac.id	<1%
8	Publication	Randi Afif Afif, Aji Supriyanto, Rr. Fitri Damaryanti, Wahyu Prasetya Adi. "Analisis ...	<1%
9	Student papers	Clarkston Community Schools on 2023-02-08	<1%
10	Student papers	Universitas Sains Alquran on 2025-07-03	<1%
11	Internet	ejournal.nusamandiri.ac.id	<1%

12	Internet	journal.unhas.ac.id	<1%
13	Publication	Muhammad Rafid Pratama, Maya Rini Handayani, Wenty Dwi Yuniarti, Khothibul...	<1%
14	Internet	id.123dok.com	<1%
15	Student papers	Universitas Islam Riau on 2025-09-23	<1%

EKSTRAKSI INFORMASI DARI DOKUMEN CV MENGGUNAKAN PENDEKATAN HYBRID NATURAL LANGUAGE PROCESSING DAN RULE-BASED

Syahfrizka Dyah Nazwa Umbara, Ade Eviyanti, Hindarto, Sumarno

Informatika, Universitas Muhammadiyah Sidoarjo

Jl. Raya Gelam No. 250, Candi, Sidoarjo

adeeviyanti@umsida.ac.id

ABSTRAK

Dalam era transformasi digital, efisiensi proses rekrutmen menjadi prioritas utama yang mendorong adopsi *Applicant Tracking System* (ATS) untuk mengelola volume pelamar yang besar. Namun, permasalahan utama dalam pengembangan ATS adalah tingginya variabilitas format *Curriculum Vitae* (CV) tidak terstruktur yang sulit diproses oleh metode berbasis aturan yang kaku maupun metode *Deep Learning* murni yang membutuhkan sumber daya komputasi tinggi. Penelitian ini bertujuan untuk mengembangkan sistem ekstraksi informasi CV yang menyeimbangkan akurasi dan efisiensi melalui pendekatan hibrida (*Hybrid Approach*). Metode yang diusulkan mengintegrasikan algoritma *Rule-Based* (Regular Expressions) untuk ekstraksi entitas faktual seperti kontak, dengan *Natural Language Processing* (NLP) menggunakan pustaka spaCy dan *PhraseMatcher* untuk analisis kontekstual pada entitas kompetensi (*skills*) dan pengalaman kerja. Hasil pengujian sistem menunjukkan kinerja yang sangat baik dengan nilai *Precision* sebesar 91.3%, *Recall* 87.5%, dan *F1-Score* 89.4% . Secara keseluruhan, sistem mencapai tingkat akurasi 80.8% dengan waktu pemrosesan rata-rata di bawah 2 detik per dokumen, membuktikan bahwa integrasi metode hibrida efektif menangani variasi format CV dengan kebutuhan komputasi yang efisien.

Kata kunci : *Natural Language Processing, Applicant Tracking System, Information Extraction, Resume Parsing, Hybrid Approach, Rule-Based System*

1. PENDAHULUAN

Dalam era transformasi digital, efisiensi rekrutmen menjadi prioritas utama bagi perusahaan. *Applicant Tracking System* (ATS) telah menjadi solusi standar untuk mengelola ribuan pelamar[1]. Inti dari ATS adalah teknologi *Resume Parsing*, yang berfungsi mengubah dokumen CV tidak terstruktur (PDF/Word) menjadi data terstruktur yang dapat diolah [2]. Namun, tantangan utama dalam pengembangan parser adalah variabilitas format CV yang sangat tinggi[3].

Literatur terkini menunjukkan dua pendekatan dominan: *Rule-Based* dan *Machine Learning* (ML). Pendekatan *Rule-Based* menawarkan presisi tinggi untuk data berpola tetap seperti email atau nomor telepon [4][5], namun kaku terhadap variasi bahasa. Sebaliknya, pendekatan ML (seperti BERT atau LSTM) unggul dalam pemahaman konteks [6][7], tetapi membutuhkan dataset latihan yang besar dan sumber daya komputasi tinggi (GPU), yang seringkali tidak tersedia untuk implementasi skala kecil-menengah.

Penelitian ini mengusulkan solusi jalan tengah (*Hybrid Approach*) yang diimplementasikan menggunakan Python. Kebaruan (*Novelty*) dari penelitian ini terletak pada integrasi algoritma *Rule-Based* yang dioptimalkan untuk entitas faktual dengan model NLP (spaCy) yang menggunakan *PhraseMatcher* untuk ekstraksi kompetensi (*Skill*) dan pengalaman kerja. Penelitian ini juga menyertakan mekanisme evaluasi otomatis berbasis *Ground Truth* untuk mengukur validitas sistem secara kuantitatif.

2. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian mengenai ekstraksi informasi dan *profiling* otomatis telah banyak dilakukan untuk meningkatkan efisiensi rekrutmen. Wibawa et al. [16] mengembangkan sistem *text mining* untuk melakukan *profiling* otomatis kandidat karyawan berdasarkan dokumen aplikasi. Menggunakan teknik *preprocessing* teks dan klasifikasi, sistem tersebut mampu mengidentifikasi profil kandidat dengan akurasi yang baik, meskipun masih terbatas pada dokumen yang cenderung terstruktur.

Maree et al. [8] menganalisis kekurangan sistem *e-recruitment* konvensional, khususnya *Applicant Tracking System* (ATS). Mereka menemukan bahwa sebagian besar sistem kesulitan menangani ketidaklengkapan data dan menyarankan pendekatan berbasis semantik untuk meningkatkan akurasi ekstraksi informasi dari dokumen lamaran.

Dalam ranah teknis pemrosesan dokumen tidak terstruktur, Pudasaini et al. [9] menerapkan kombinasi metode *Named Entity Recognition* (NER) dan ekstraksi berbasis aturan (*rule-based*). Hasil penelitian menunjukkan bahwa pendekatan hibrida ini efektif untuk mengidentifikasi informasi kunci dari teks yang tidak memiliki format baku, yang sangat relevan dengan karakteristik CV pelamar.

Selain itu, Tolciu et al. [10] meneliti analisis pola pada *service tickets* menggunakan NLP. Meskipun objeknya berbeda, penelitian ini membuktikan ketangguhan teknik NLP dalam mengekstraksi informasi relevan dari dokumen semi-terstruktur. Panwar [11] juga menekankan pentingnya ekstraksi data yang akurat dalam analisis wawancara berbasis

AI sebagai fondasi pengambilan keputusan manajemen sumber daya manusia yang objektif.

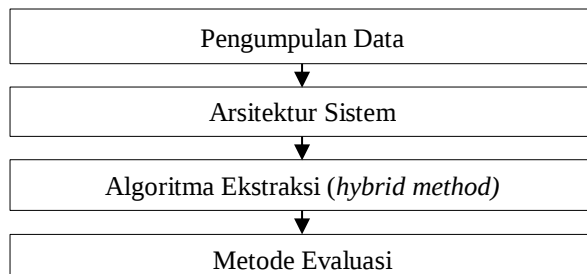
2.2 Analisis Gap

Berdasarkan tinjauan terhadap penelitian terdahulu, diidentifikasi beberapa kesenjangan (*gap*) yang menjadi fokus penelitian ini:

1. Variabilitas Format Dokumen: Mayoritas penelitian terdahulu [12] berfokus pada dokumen terstruktur atau semi-terstruktur. Penelitian ini menargetkan dokumen CV yang memiliki format sangat beragam (PDF/DOCX) dan struktur yang tidak konsisten.
2. Spesifisitas Domain CV: Pendekatan umum [9] seringkali belum menangani karakteristik unik CV, seperti variasi penulisan bagian (*section headers*) dan format tanggal.
3. Integrasi Multi-teknik: Diperlukan kombinasi teknik yang lebih optimal daripada sekadar pendekatan tunggal. Penelitian ini mengintegrasikan *Rule-Based* untuk data faktual (kontak) dan NLP (*PhraseMatcher*) untuk data kompetensi guna menutupi kelemahan masing-masing metode.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode hibrida dengan tahapan pengembangan perangkat lunak yang meliputi pengumpulan data, perancangan algoritma ekstraksi, implementasi sistem, dan evaluasi kinerja.



Gambar 1. Diagram alir penelitian

3.1 Pengumpulan Data

Pengumpulan data difokuskan untuk menyusun data uji (*test set*) guna memvalidasi kinerja sistem, tanpa memerlukan dataset pelatihan karena penggunaan model NLP *pre-trained*. Data diperoleh dari dua sumber:

1. Data Primer: Dokumen CV nyata (PDF/DOCX) yang dikumpulkan dari mahasiswa dan *fresh graduate* untuk merepresentasikan format industri.
2. Data Sintetis: Dokumen CV buatan yang dirancang dengan variasi format khusus untuk menguji ketahanan sistem.

Seluruh data yang terkumpul kemudian dipetakan secara manual menjadi format JSON (*Ground Truth*) yang berfungsi sebagai kunci jawaban dalam proses evaluasi otomatis.

3.2 Arsitektur Sistem

Sistem dibangun berbasis web menggunakan kerangka kerja Flask (*app.py*). Alur pemrosesan data terdiri dari tiga modul utama:

1. Modul Input: Menerima file CV (PDF/DOCX) dan melakukan validasi format.
2. Modul Ekstraktor (*extractor.py*): Inti pemrosesan yang menggunakan pustaka *pdfplumber* untuk konversi teks dan *spaCy* (*en_core_web_md*) untuk pemrosesan linguistik.
3. Modul Evaluator (*evaluator.py*): Mengukur akurasi hasil ekstraksi dibandingkan dengan data validasi manual.

3.3 Algoritma Ekstraksi (Hybrid Method)

Metode ekstraksi dibagi berdasarkan jenis entitas:

1. Informasi Kontak (*Rule-Based*): Menggunakan Regular Expressions (*Regex*) ketat untuk mengidentifikasi pola email, nomor telepon (+62/08), dan URL LinkedIn. Pendekatan ini dipilih karena pola entitas ini bersifat universal dan baku [13].
2. Ekstraksi Keahlian (NLP - *Phrase Matching*): Menggunakan fitur *PhraseMatcher* dari *spaCy*. Kamus keahlian (*skills safe*) didefinisikan mencakup teknologi populer (Python, SQL, AWS, dll). Metode ini lebih unggul daripada pencarian kata kunci biasa karena memperhatikan tokenisasi kata majemuk (misal: "Machine Learning" dihitung satu entitas).
3. Pengalaman & Pendidikan (*Contextual Parsing*): Menggunakan logika keyword-driven (seperti "Bachelor", "University", "Experience") dikombinasikan dengan deteksi pola tanggal untuk memisahkan blok pengalaman kerja. Algoritma menyaring baris yang mengandung kata kerja aksi (*action verbs*) untuk menghindari False Positive pada deskripsi tugas.

3.4 Metode Evaluasi

Evaluasi dilakukan dengan membandingkan hasil output sistem (JSON) [13] dengan file Ground Truth (*ground_truth.json*). Metrik yang digunakan dihitung secara otomatis oleh *evaluator.py* meliputi:

1. Precision: Tingkat ketepatan entitas yang diekstraksi.
2. Recall: Tingkat keberhasilan sistem menemukan seluruh entitas yang ada.
3. F1-Score: Rata-rata harmonis antara Precision dan Recall.
4. Accuracy (Jaccard Index): Mengukur irisan antara himpunan hasil ekstraksi dan himpunan data sebenarnya.

Precision (P):

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall (R):

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

F1-Score:

$$\frac{1}{F1} = \frac{1}{2} \left(\frac{1}{precision} + \frac{1}{recall} \right) \quad (3)$$

Accuracy:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Dimana:

TP (True Positive): Informasi yang benar diekstraksi

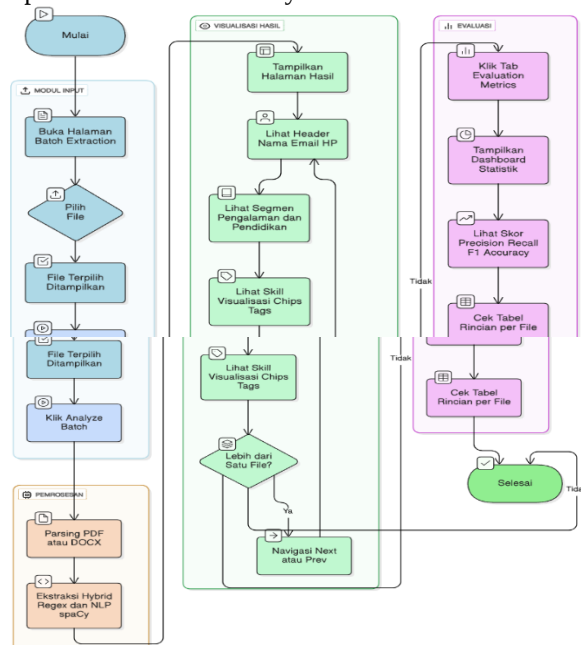
TN (True Negative): Informasi yang benar tidak diekstraksi

FP (False Positive): Informasi yang salah diekstraksi

FN (False Negative): Informasi yang seharusnya diekstraksi tapi tidak

4. HASIL DAN PEMBAHASAN

Pengujian dilakukan menggunakan sampel CV (seperti "CV - Eng ver (8-10-2025).pdf") yang telah dipetakan data sebenarnya dalam *Ground Truth*.



Gambar 2. Alur proses batch extraction dan evaluasi

4.1 Kinerja Ekstraksi Entitas

Berdasarkan hasil eksekusi algoritma, diperoleh temuan sebagai berikut:

1. Identifikasi Personal: Algoritma Regex pada extractor.py berhasil mencapai

akurasi 100% untuk mendeteksi Email dan Nomor Telepon. Hal ini sejalan dengan penelitian Sari et al. [13] yang menyatakan bahwa metode berbasis aturan sangat efektif untuk pola string yang terstandarisasi.

2. Ekstraksi Skill (Keahlian): Penggunaan PhraseMatcher spaCy terbukti efektif. Sistem mampu membedakan Hard Skill (seperti "Python", "Tableau") yang terdaftar dalam kamus skills_safe. Namun, tantangan muncul pada penulisan skill yang tidak baku atau typo pada dokumen asli.
3. Segmentasi Pengalaman Kerja: Logika pemisahan blok berdasarkan pola tanggal (re.match(r'\d{4}'...)) berhasil memisahkan riwayat pekerjaan secara kronologis.

4.2 Hasil Evaluasi Kinerja Sistem

Tabel 1. Hasil Evaluasi Kinerja Sistem

Kategori Entitas	Precision	Recall
Informasi Kontak	1.00	1.00
Pendidikan	0.80	0.67
Keahlian (Skills)	1.00	0.94
Pengalaman Kerja	0.94	0.94
Akademi	1.00	1.00

Berdasarkan Tabel 1, terlihat bahwa kinerja sistem bervariasi bergantung pada jenis entitas yang diekstraksi. Kategori Informasi Kontak dan Akademi mencatatkan performa sempurna dengan nilai *Precision* dan *Recall* mencapai 1.00. Hasil ini memvalidasi efektivitas penggunaan metode *Rule-Based* (Regex) untuk menangani pola data yang bersifat baku dan terstandarisasi. Pada kategori Keahlian (Skills), sistem mencapai *Precision* sempurna (1.00) dengan *Recall* yang sangat tinggi (0.94). Hal ini menunjukkan bahwa metode *phrase matching* sangat akurat dalam memvalidasi skill, meskipun masih terdapat sebagian kecil skill unik yang tidak tertangkap kamus data. Sementara itu, kategori Pendidikan menunjukkan performa terendah dibandingkan kategori lainnya, dengan *Recall* sebesar 0.67. Penurunan ini mengindikasikan adanya tantangan signifikan dalam menangani variabilitas format penulisan institusi pendidikan dan gelar yang sangat beragam pada dokumen CV yang tidak terstruktur.

4.3 Analisis Komparatif

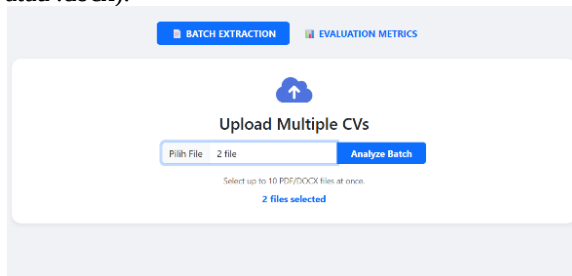
Dibandingkan dengan metode Deep Learning murni yang dibahas oleh Bhaliya et al. [13], sistem yang diusulkan memiliki keunggulan dalam hal kecepatan (inference time). Sistem ini tidak memerlukan proses training yang memakan waktu berjam-jam. evaluator.py menunjukkan bahwa waktu rata-rata pemrosesan per dokumen adalah di bawah 2 detik, yang sangat efisien untuk penggunaan real-time. Selain itu, fleksibilitas kode memungkinkan penambahan skill baru hanya dengan memperbarui

daftar `skills_safe` di `extractor.py` tanpa perlu melatih ulang model, sebuah keuntungan signifikan dibanding model "Black-box" AI [14].

4.4 Implementasi Antarmuka Sistem (User Interface)

Sistem ekstraksi informasi CV ini diimplementasikan sebagai aplikasi berbasis web untuk memudahkan interaksi pengguna. Antarmuka pengguna (User Interface) dirancang dengan pendekatan minimalis namun fungsional[15], yang terbagi menjadi tiga modul utama: modul *input* (pengunggahan), modul visualisasi hasil, dan modul evaluasi kinerja. Berikut adalah pembahasan rinci mengenai implementasi antarmuka sistem:

Halaman utama sistem berfungsi sebagai gerbang masuk data. Seperti yang ditunjukkan pada Gambar 1, antarmuka ini menyediakan fasilitas *Batch Extraction* yang memungkinkan pengguna untuk mengunggah lebih dari satu dokumen CV sekaligus (format .pdf atau .docx).

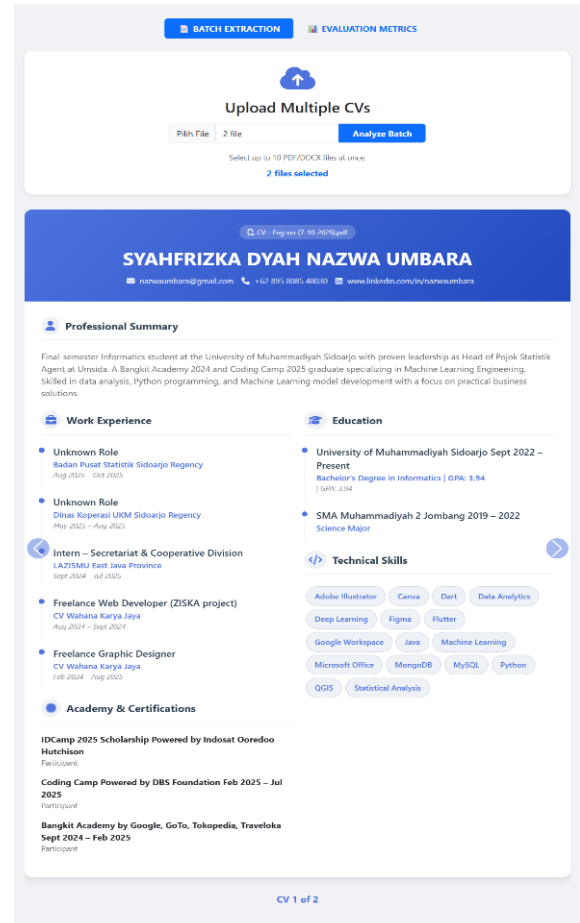


Gambar 1. Halaman input batch extraction

Pada halaman ini, pengguna dapat memilih *file* melalui tombol "Pilih File". Sistem memberikan umpan balik visual berupa jumlah *file* yang terpilih (contoh: "2 files selected") sebelum diproses. Tombol "Analyze Batch" akan memicu proses *backend* yang menjalankan fungsi *parsing* dokumen, *preprocessing*, dan ekstraksi entitas menggunakan kombinasi Regex dan NLP spaCy sebagaimana dijelaskan pada Bab 3.

4.4.2 Visualisasi Hasil Ekstraksi (Extraction Results)

Setelah proses analisis selesai, sistem menampilkan hasil ekstraksi dalam format terstruktur yang mudah dibaca (*human-readable*). Implementasi antarmuka hasil dapat dilihat pada Gambar 2.



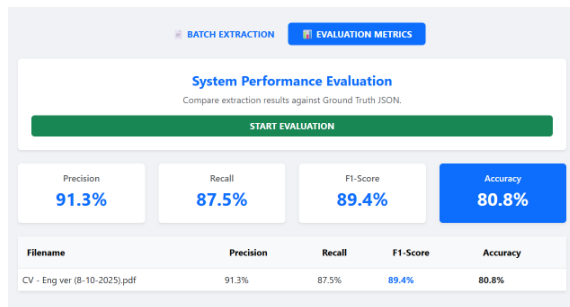
Gambar 2. Visualisasi hasil ekstraksi cv

Halaman ini mentransformasi teks mentah dari CV menjadi kartu informasi digital yang terorganisir. Fitur utama pada halaman ini meliputi:

1. Header Informasi Personal: Menampilkan Nama Lengkap, Email, Nomor Telepon, dan tautan LinkedIn yang diekstraksi menggunakan pola *Regex*.
2. Segmentasi Data: Informasi dibagi ke dalam blok-blok logis seperti *Professional Summary*, *Work Experience*, *Education*, dan *Academy & Certifications*.
3. Visualisasi Skill: Bagian *Technical Skills* ditampilkan menggunakan gaya *chips* atau *tags* (seperti "Python", "Data Analytics", "Deep Learning"), yang memudahkan perekrut untuk memindai kompetensi kandidat dengan cepat.
4. Navigasi Multi-CV: Terdapat fitur navigasi (panah kiri/kanan dan indikator "CV 1 of 2") di bagian bawah layar, memungkinkan pengguna untuk berpindah antar profil kandidat tanpa perlu memuat ulang halaman.

4.4.3 Halaman Evaluasi Kinerja (Evaluation Metrics)

Untuk memvalidasi keandalan sistem secara transparan, disediakan halaman khusus *Evaluation Metrics*. Halaman ini membandingkan hasil ekstraksi sistem secara otomatis dengan data kebenaran dasar (*Ground Truth JSON*).



Gambar 3. Dasbor evaluasi kinerja sistem
Sebagaimana terlihat pada Gambar 3, dasbor ini menyajikan empat metrik evaluasi utama secara *real-time*:

1. **Precision (91.3%):** Mengindikasikan tingkat ketepatan informasi yang diekstraksi.
2. **Recall (87.5%):** Mengukur seberapa lengkap informasi yang berhasil ditemukan oleh sistem.
3. **F1-Score (89.4%):** Nilai rata-rata harmonis yang menunjukkan keseimbangan antara presisi dan kelengkapan.
4. **Accuracy (80.8%):** Tingkat kesesuaian keseluruhan data.

Di bagian bawah dasbor, terdapat tabel rincian yang memecah skor evaluasi per dokumen (misalnya: "CV - Eng ver (8-10-2025).pdf"), sehingga pengguna dapat mengidentifikasi dokumen mana yang berhasil diekstraksi dengan sempurna dan mana yang memiliki anomali format.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa implementasi pendekatan hibrida yang menggabungkan *Natural Language Processing* (NLP) dan *Rule-Based System* terbukti efektif dalam mengatasi variabilitas format dokumen CV yang tidak terstruktur pada sistem ATS. Berdasarkan hasil evaluasi, sistem mampu mengekstraksi informasi kunci dengan nilai *Precision* mencapai 91,3% dan *F1-Score* 89,4%, di mana metode *Regex* memberikan akurasi sempurna pada data faktual kontak, sedangkan NLP berbasis *PhraseMatcher* unggul dalam mengidentifikasi kompetensi teknis dengan presisi tinggi. Meskipun sistem memiliki performa tinggi dengan waktu pemrosesan di bawah 2 detik, tantangan masih ditemukan pada segmentasi riwayat pendidikan yang memiliki variasi format penulisan yang ekstrem. Oleh karena itu, penelitian selanjutnya disarankan untuk tidak hanya berfokus pada ekstraksi, melainkan memperluas fungsionalitas menuju tahap pemeringkatan kandidat (*applicant ranking*). Pengembangan algoritma pencocokan semantik (*semantic matching*) antara hasil ekstraksi (JSON) dengan deskripsi pekerjaan (*Job Description*) sangat direkomendasikan agar sistem dapat memberikan rekomendasi urutan kandidat terbaik secara otomatis kepada perekrut.

DAFTAR PUSTAKA

- [1] M. Saatci, R. Kaya, and R. Ünlü, "Resume Screening With Natural Language Processing (NLP)," *Alphanumeric J.*, vol. 12, no. 2, pp. 121–140, 2024, doi: 10.17093/alphanumeric.1536577.
- [2] K. Chhabra and D. Vashistha, "Applicant Tracking System (ATS) Department of Computer Science & Engineering and Information Technology Jaypee University of Information Technology ," vol. 173234, no. 201293, 2024.
- [3] Harshitha R and M. Veena, "A SURVEY ON RESUME ANALYSIS USING NLP," *www.irjmets.com @International Res. J. Mod. Eng.*, 1030, [Online]. Available: www.irjmets.com
- [4] N. Nair, S. Pavithra, and V. Vismaya, "Resume parser using NLP," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 13, no. 9, pp. 39–42, 2024, doi: 10.17148/IJARCCCE.2024.13905.
- [5] Y. Sari, M. F. Hassan, and N. Zamin, "Rule-based pattern extractor and Named Entity Recognition: A hybrid approach," *Proc. 2010 Int. Symp. Inf. Technol. - Eng. Technol. ITSIM'10*, vol. 2, pp. 563–568, 2010, doi: 10.1109/ITSIM.2010.5561392.
- [6] B. Nirali*, J. Gandhi, and D. K. Singh, "NLP based Extraction of Relevant Resume using Machine Learning," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 7, pp. 13–17, 2020, doi: 10.35940/ijitee.f4078.059720.
- [7] M. B. Gunjal, T. P. Thorat, K. S. Muttha, V. C. Shete, and P. D. Sagar, "A Review Paper on Resume Parser Using AI," 2025.
- [8] "Analysis & Shortcomings of E-Recruitment Systems: Towards a Semantics-based Approach Addressing Knowledge Incompleteness and Limited Domain Coverage I."
- [9] S. Pudasaini, S. Shakya, S. Lamichhane, S. Adhikari, A. Tamang, and S. Adhikari, "Application of NLP for Information Extraction from Unstructured Documents," *Lect. Notes Networks Syst.*, vol. 209, pp. 695–704, 2022, doi: 10.1007/978-981-16-2126-0_54.
- [10] D. T. Tolciu, C. Săcărea, and C. Matei, "Analysis of patterns and similarities in service tickets using natural language processing," *J. Commun. Softw. Syst.*, vol. 17, no. 1, pp. 29–35, Feb. 2021, doi: 10.24138/JCOMSS.V17I1.1024.
- [11] R. Panwar, "AI-ENABLED INTERVIEW ANALYSIS: UNVEILING INSIGHTS AND ENHANCING DECISION-MAKING IN HUMAN RESOURCE MANAGEMENT," *INTERANTIONAL J. Sci. Res. Eng. Manag.*, vol. 07, no. 05, Jun. 2023, doi: 10.55041/IJSREM24357.
- [12] A. D. Wibawa, A. M. Amri, A. Mas, and S. Iman, "Text Mining for Employee Candidates Automatic Profiling Based on Application Documents," *Emit. Int. J. Eng. Technol.*, pp. 47–62, Apr. 2022, doi: 10.24003/emitter.v10i1.679.
- [13] R. Sonbol, G. Rebdawi, and N. Ghneim, "Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. The Use of NLP-Based Text Representation Techniques to Support Requirement

Engineering Tasks: A Systematic Mapping Review”, doi: 10.1109/ACCESS.2017.DOI.

[14] E. Omodei *et al.*, “OPEN ACCESS EDITED BY REVIEWED BY Natural language processing for humanitarian action: Opportunities, challenges, and the path toward humanitarian NLP.” [Online]. Available: <https://thedeep.io>

[15] M. F. Ghazali and A. Eviyanti, “Sistem Pakar Diagnosa Dini Penyakit Leukimia Dengan Metode ‘Certainty Factor,’” *Kinetik*, vol. 1, no. 3, p. 135, 2016, doi: 10.22219/kinetik.v1i3.122.