

## SISTEM *SCREENING* CV OTOMATIS BERBASIS NLP DAN *RULE-BASED* UNTUK SELEKSI KANDIDAT BERBASIS WEB

Rikardo<sup>1</sup>, Rafi Maulana Ash Shidqi<sup>2</sup>, Ana Sujana<sup>3</sup>, Sigit Wibawa<sup>4</sup>, Untung Rohwadi<sup>5</sup>,  
Felix Wuryo Handono<sup>6</sup>

Email: [rikardoo2212@gmail.com](mailto:rikardoo2212@gmail.com)<sup>1</sup>, [rafi.maulanash@gmail.com](mailto:rafi.maulanash@gmail.com)<sup>2</sup>,  
[anasujana99@gmail.com](mailto:anasujana99@gmail.com)<sup>3</sup>, [sigit.stb@bsi.ac.id](mailto:sigit.stb@bsi.ac.id)<sup>4</sup>, [untung.unr@bsi.ac.id](mailto:untung.unr@bsi.ac.id)<sup>5</sup>,  
[felix@bsi.ac.id](mailto:felix@bsi.ac.id)<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Universitas Bina Sarana Informatika

### ABSTRAK

Peningkatan volume pelamar dalam era digital menuntut otomatisasi proses seleksi kandidat yang efisien, akurat, dan objektif. Penelitian ini mengembangkan sistem *screening* CV otomatis berbasis web yang mengintegrasikan *Natural Language Processing* (NLP) dan pendekatan *rule-based* untuk meningkatkan efisiensi proses rekrutmen di divisi Human Resources. Sistem dirancang menggunakan Python dengan memanfaatkan pustaka Gradio untuk antarmuka pengguna, PyMuPDF untuk ekstraksi teks PDF, spaCy untuk *Named Entity Recognition*, dan *Regular Expressions* untuk pencocokan pola keahlian. Algoritma hybrid menggabungkan ekstraksi otomatis data kritis (nama, pengalaman kerja, keterampilan) melalui NLP dengan penilaian berbasis aturan eksplisit yang ditentukan HR. Sistem menggunakan *weighted scoring model* dengan tiga komponen utama: skor pengalaman (40%), skor *hard skill* wajib (40%), dan skor *hard skill* tambahan (20%). Pengujian pada data simulasi menunjukkan sistem mampu mengklasifikasikan kandidat menjadi tiga kategori rekomendasi (Utama, Dipertimbangkan, Tidak Cocok) dengan *threshold* skor 75, 50, dan 0. Hasil penelitian menunjukkan sistem berhasil mengotomatisasi proses seleksi awal, mengurangi waktu *screening* manual, meningkatkan konsistensi penilaian, serta memberikan dasar objektif untuk pengambilan keputusan rekrutmen sambil tetap mempertahankan *oversight* manusia dalam penilaian akhir.

**Kata Kunci:** *Screening* CV Otomatis, *Natural Language Processing*, *Rule-Based*, Sistem Rekomendasi, Rekrutmen Berbasis AI.

### ABSTRACT

*The increasing volume of job applicants in the digital era demands automation of the selection process to ensure efficiency, accuracy, and objectivity. This research develops*

*an automated résumé screening system based on web technologies. It integrates Natural Language Processing (NLP) and rule-based approaches to enhance Human Resources recruitment efficiency. The system uses Python, Gradio for the user interface, PyMuPDF for PDF text extraction, spaCy for Named Entity Recognition, and Regular Expressions for skill pattern matching. A hybrid algorithm extracts critical data (name, work experience, skills) with NLP and applies HR-determined scoring rules. The scoring model weights experience (40%), mandatory hard skills (40%), and additional hard skills (20%). Testing on simulated data shows that the system classifies candidates into Primary, Considered, and Not Suitable, with score thresholds of 75, 50, and 0, respectively. Results indicate successful automation of initial selection, reduced screening time, improved consistency, and greater objectivity. Human oversight is maintained during final evaluations.*

**Keywords:** *Automated CV Screening, Natural Language Processing, Rule-Based, Recommendation System, AI-Driven Recruitment.*

## 1. PENDAHULUAN

Perkembangan dunia kerja di era digital telah mendorong perusahaan untuk beradaptasi dengan teknologi dalam setiap aspek rekrutmen. Salah satu tantangan utama HRD adalah meningkatnya jumlah pelamar yang menyebabkan proses penyeleksian CV secara manual menjadi sangat memakan waktu, rentan kesalahan, dan berpotensi menimbulkan bias subjektif. Kebutuhan akan otomatisasi dalam seleksi awal kandidat kini menjadi semakin mendesak seiring tuntutan efisiensi dan akurasi proses rekrutmen yang lebih baik. Di era digital ini, urgensi penggunaan teknologi otomatis dalam proses seleksi semakin mengemuka karena besarnya volume data pelamar yang harus diolah secara manual oleh HRD, sehingga dibutuhkan sistem yang mampu mempercepat dan mengefisienkan proses seleksi awal secara objektif, salah satunya dengan mengimplementasikan *Natural Language Processing (NLP)* dan *rule-based*.

*Natural Language Processing (NLP)* merupakan cabang kecerdasan buatan yang banyak diadopsi untuk mengekstraksi informasi penting dari berbagai dokumen teks, termasuk CV pelamar kerja. Melalui teknik NLP, sistem dapat mengidentifikasi data kunci seperti pendidikan, pengalaman, dan keterampilan secara otomatis sehingga mempercepat proses penyaringan awal sebelum penilaian lanjutan oleh HR. Pendekatan ini selaras dengan tren digitalisasi proses bisnis di sektor sumber daya manusia. Selain NLP, metode *rule-based* sering diterapkan untuk memformulasikan kriteria seleksi kandidat secara eksplisit, misalnya penentuan syarat minimum pendidikan, pengalaman

kerja, atau keahlian spesifik. Integrasi kedua metode NLP untuk ekstraksi data dan *rule-based* untuk pengambilan keputusan dapat meningkatkan efisiensi serta akurasi sistem *screening* CV karena keputusan seleksi mengikuti kriteria yang telah ditentukan perusahaan.

Berbagai riset telah membahas metode otomatisasi seleksi dengan pendekatan yang beragam. Affandi & Hermanto misalnya, menggunakan algoritma *K-Nearest Neighbor* berbasis NLP untuk *screening* CV otomatis yang mampu mempercepat proses seleksi dan memberikan penyaringan berbasis skor. Penelitian lain dari Afdi Fauzul Bahar mengembangkan sistem berbasis machine learning menggunakan metode *Random Forest*, *Naive Bayes*, dan *Logistic Regression* tanpa *rule-based* untuk membandingkan akurasi dan kecepatan seleksi dokumen pelamar. Di tingkat internasional, Kogan dkk membahas pemanfaatan berbagai model *deep learning* untuk ekstraksi data dari *resume* pada seleksi lowongan IT. Beragamnya pendekatan tersebut menunjukkan bahwa pemilihan metode sangat berpengaruh terhadap kinerja aplikasi *screening* CV yang dihasilkan, baik dari sisi akurasi maupun efisiensi.

**Tabel 1. Perbandingan penelitian terkait sistem *screening* CV otomatis**

No	Peneliti / Tahun	Metode Utama	Rule-based	Output Utama	Kelebihan / Fokus	Kekurangan / Celah
1	Affandi & Hermanto [1]	NLP + KNN	Tidak	Skor kesesuaian kandidat	Mempercepat seleksi dengan pemodelan klasifikasi berbasis teks CV	Tidak ada aturan eksplisit dari HR, fleksibilitas kriteria terbatas
2	Afdi Fauzul Bahar [2]	<i>Random Forest</i> , <i>Naive Bayes</i> , <i>Logistic Regression</i>	Tidak	Perbandingan akurasi algoritma untuk seleksi CV	Menguji beberapa model ML untuk mencari akurasi terbaik	Tidak mengintegrasikan preferensi HR ke dalam <i>rules</i> ; fokus pada performa model
3	Kogan dkk [3]	<i>Deep learning</i> untuk ekstraksi resume	Tidak	Ekstraksi fitur dari <i>resume</i> lowongan IT	Mampu menangani fitur kompleks dan konteks teks	Kompleks, butuh data besar dan komputasi tinggi, tidak fokus ke <i>rules</i> HR
4	Penelitian ini	NLP + <i>Rule-based</i> + <i>weighted scoring</i>	Ya	Skor dan kategori rekomendasi kandidat (Utama, Dipertimbangkan, Tidak Cocok)	Menggabungkan ekstraksi otomatis dan aturan eksplisit HR; fleksibel pada kriteria posisi, pengalaman, <i>skill</i>	Pengujian masih terbatas dan <i>threshold</i> belum dikalibrasi pada data historis perusahaan

Ringkasan perbedaan utama antara penelitian ini dengan penelitian terkait ditunjukkan pada Tabel 1. Tabel tersebut menegaskan posisi penelitian ini yang mengombinasikan NLP dan *rule-based* dengan model pembobotan skor yang eksplisit ditentukan oleh HR, berbeda dengan penelitian sebelumnya yang umumnya berfokus pada pemodelan *machine learning* tanpa integrasi rules HR. *Research gap* yang diidentifikasi adalah minimnya studi dan implementasi aplikasi web *screening* CV yang terintegrasi, responsif, dan adaptif terhadap kebutuhan praktis HRD, terutama dalam konteks penyesuaian kriteria seleksi yang dapat berubah sewaktu-waktu mengikuti kebutuhan posisi dan kebijakan organisasi.

Berdasarkan latar belakang dan *research gap* tersebut, penelitian ini mengangkat beberapa rumusan masalah: (1) bagaimana merancang sistem *screening* CV otomatis berbasis NLP dan *rule-based* yang efektif sesuai kriteria HR; (2) bagaimana membuktikan peningkatan efisiensi serta akurasi dari sistem yang dikembangkan dibanding proses manual; dan (3) apa saja tantangan dalam penerapan model ini secara *real-world* dalam bentuk aplikasi web. Dengan rumusan masalah tersebut, penelitian ini bertujuan untuk mengembangkan serta menguji sistem otomatis *screening* CV berbasis Python yang mengintegrasikan NLP dan *rule-based*, mempermudah dan mempercepat proses seleksi kandidat di HR, serta mengurangi risiko *human error* dan meningkatkan objektivitas seleksi awal. Dengan demikian, penelitian diharapkan memberikan manfaat bagi HR dalam mempercepat proses rekrutmen sekaligus menjadi referensi pengembangan lebih lanjut sistem *screening* CV berbasis kecerdasan buatan di Indonesia.

## 2. METODE PENELITIAN

### Tahapan Penelitian

Penelitian dilakukan dengan tahapan sebagai berikut:

1. **Analisis Kebutuhan:** Identifikasi kebutuhan proses seleksi CV pada pengguna HR, meliputi kriteria seleksi (pendidikan, pengalaman, keterampilan) yang akan digunakan sebagai *rules* dalam sistem.
2. **Pengumpulan Dataset CV:** Dataset CV diperoleh dari sumber internal atau simulasi data pelamar kerja.
3. **Perancangan Sistem:** Sistem dirancang berbasis web dengan *interface* untuk HR, agar dapat mengunggah CV dan memasukkan kriteria seleksi. Pola ini sejalan dengan praktik pengembangan aplikasi *screening* dan *e-recruitment* modern yang

mengintegrasikan antarmuka web dengan modul pemrosesan AI di sisi *backend* untuk memudahkan penggunaan oleh praktisi HR non-teknis.

4. **Pengembangan Algoritma NLP dan *Rule-based*:** NLP digunakan untuk mengekstrak data penting dari file CV, seperti *text parsing*, pencarian *keyword*, dan identifikasi bagian dokumen (pendidikan, pengalaman kerja, skill). *Rule-based* digunakan untuk menilai dan membandingkan hasil ekstraksi dengan kriteria yang dimasukkan. Pendekatan *hybrid* seperti ini menggabungkan ekstraksi informasi berbasis NLP dengan aturan eksplisit HR banyak direkomendasikan dalam penelitian otomasi *resume screening* karena memberikan keseimbangan antara fleksibilitas dan transparansi keputusan sistem.

5. **Implementasi Sistem:**

Pengembangan dilakukan menggunakan bahasa Python di Google Colab, memanfaatkan pustaka NLP (misal: spaCy, NLTK, re), serta *library web framework* ringan seperti Streamlit/Flask untuk membangun antarmuka HR.

6. **Pengujian Sistem:**

Sistem diuji dengan mengunggah sejumlah CV, di mana keluaran berupa skor kesesuaian kandidat dengan kriteria HR. Pengujian mencakup fungsionalitas (input data, proses ekstraksi, pemberian skor, output hasil), serta kemudahan penggunaan oleh HR. Pola evaluasi ini selaras dengan panduan evaluasi sistem rekrutmen berbasis AI yang menekankan uji akurasi hasil seleksi sekaligus aspek *usability* bagi pengguna HR sebagai end-user utama.

## Metode Pengumpulan Data

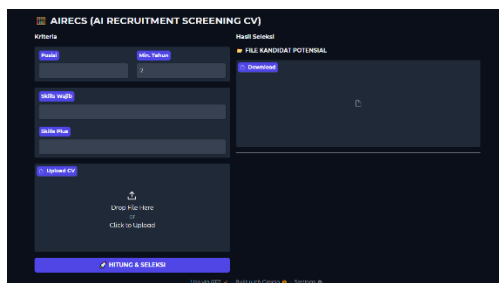
Data utama penelitian berupa file CV pelamar yang dikirim oleh kandidat kepada HR. Penggunaan data ini digunakan pada penelitian pengembangan sistem pendukung keputusan rekrutmen ketika data riil bersifat sensitif atau terbatas, selama struktur dan variasi datanya merepresentasikan kondisi lapangan.

## Perangkat Pengembangan

Implementasi sistem dilakukan menggunakan bahasa pemrograman Python karena dukungan pustaka (*library*) yang ekstensif untuk pengolahan teks dan data. Berikut adalah beberapa pustaka yang digunakan dalam penelitian ini adalah:

1. **Gradio (gradio):** Digunakan untuk merancang antarmuka pengguna (*Graphical*

*User Interface - GUI*) berbasis web. Pustaka ini memfasilitasi input kriteria dinamis (posisi, pengalaman, keahlian) serta visualisasi laporan hasil *screening* secara *real-time*.



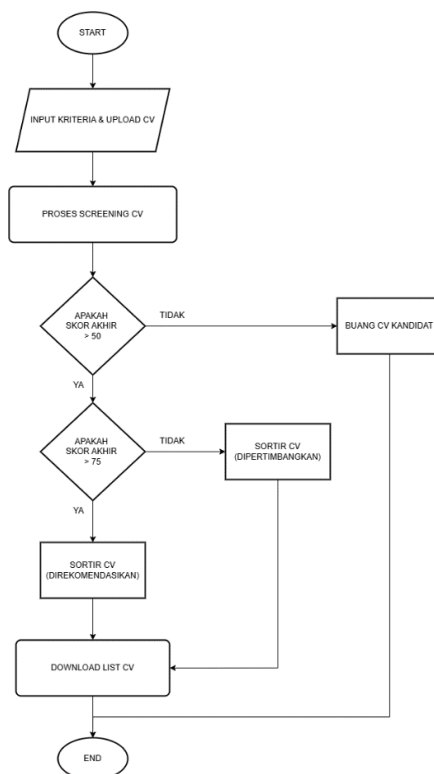
**Gambar 1. Tampilan Antarmuka Web**

Gambar ini menunjukkan antarmuka utama sistem *screening* CV, di mana HR dapat memasukkan kriteria seleksi berupa posisi, minimal tahun pengalaman, daftar *skills* wajib, dan *skills plus* pada panel kiri. Pada panel bawah disediakan fitur unggah CV, sedangkan panel kanan menampilkan hasil seleksi berupa daftar file kandidat potensial yang dapat diunduh setelah tombol “Hitung & Seleksi” dijalankan.

2. **PyMuPDF (fitz):** Berfungsi sebagai modul ekstraksi teks dari dokumen PDF. Dalam penelitian ini, fitur `get_text("blocks")` digunakan secara spesifik untuk menangani tata letak dokumen yang kompleks (seperti format dua kolom pada CV kreatif) dengan melakukan pengurutan blok teks berdasarkan koordinat vertikal dan horizontal (*block sorting*).
3. **spaCy (spacy):** Digunakan untuk pemrosesan bahasa alami (NLP). Model *pre-trained en\_core\_web\_sm* diimplementasikan untuk tugas *Named Entity Recognition* (NER), khususnya untuk mengidentifikasi entitas nama pelamar (*PERSON*) dari teks yang tidak memiliki pola baku.
4. **Regular Expressions (re):** Merupakan komponen inti dari metode *Rule-Based*. Modul ini digunakan untuk pencocokan pola (*pattern matching*) yang deterministik guna mendeteksi kata kunci keahlian (*skills*) dan pola tanggal pengalaman kerja.
5. **Datetime (datetime):** Digunakan untuk operasi aritmatika temporal. Modul ini mengonversi string tanggal yang diekstrak menjadi objek waktu untuk menghitung durasi pengalaman kerja secara presisi (selisih bulan dan tahun).

## Diagram Alir (*Flowchart*)

*Flowchart* ini digunakan untuk memberikan gambaran menyeluruh mengenai tahapan kerja mulai dari penerimaan file CV, proses pra-pemrosesan, ekstraksi informasi, perhitungan skor, hingga tahap seleksi akhir. Dengan adanya diagram ini, pembaca dapat memahami mekanisme sistem secara lebih jelas, terstruktur, dan mudah diikuti.



**Gambar 2. Diagram Alir (*Flowchart*)**

## Alur Pemrosesan Data (*System Logic*)

Proses *screening* otomatis dilakukan melalui beberapa tahapan algoritma sebagai berikut:

### 1. Pra-pemrosesan Dokumen (*Preprocessing*)

Sistem menerima input berupa file PDF. Menggunakan PyMuPDF, teks diekstrak dengan logika pengurutan blok (*spatial sorting*) untuk memastikan alur baca (*reading order*) yang benar, terutama pada dokumen non-linear (format kolom). Teks kemudian dinormalisasi menjadi huruf kecil (*lowercasing*) dan spasi berlebih dibersihkan.

### 2. Ekstraksi Informasi Hibrida

Sistem membagi tugas ekstraksi berdasarkan karakteristik data:

- a. **Ekstraksi Nama (Pendekatan NLP):** Mengingat nama orang tidak memiliki pola sintaksis tetap, sistem menggunakan model NER spaCy untuk mendeteksi entitas *PERSON* pada 1000 karakter pertama dokumen.
- b. **Ekstraksi Keahlian (Pendekatan *Keyword Matching*):** Sistem mencocokkan daftar keahlian (*skills*) yang diinputkan HR dengan teks CV menggunakan pencarian *word boundary* untuk menghindari kesalahan deteksi parsial (misalnya membedakan “Java” dengan “Javascript”), sesuai praktik *rule-based extraction* berbasis pola teks [9].

### 3. Algoritma Kalkulasi Pengalaman Kerja

Berbeda dengan metode ekstraksi angka sederhana, sistem menerapkan algoritma kalkulasi durasi berbasis tanggal (*Date-Difference Calculation*) dengan tahapan:

1. **Deteksi Pola:** Menggunakan *Regular Expressions* untuk menemukan pola tanggal seperti [Bulan] [Tahun] - [Bulan] [Tahun] (contoh: "Jan 2020 - Mar 2022") atau [Tahun] - [Tahun].
2. **Konversi Temporal:** Mengubah teks bulan (Bahasa Indonesia/Inggris) menjadi integer (1-12). Jika ditemukan kata kunci "*Present*", "*Now*", atau "*Sekarang*" pada tanggal akhir, sistem menggunakan tanggal waktu saat ini (*current timestamp*).
3. **Kalkulasi Durasi:** Menghitung selisih total bulan ( $M_{total}$ ) menggunakan rumus:

$$M_{total} = ((Y_{end} - Y_{start}) \times 12) + (M_{end} - M_{start}) \quad 1)$$

Total pengalaman dalam tahun ( $E_{years}$ ) diperoleh dari:

$$E_{years} = \frac{\sum M_{total}}{12} \quad 2)$$

## Mekanisme Penilaian (*Scoring*) dan Seleksi

Untuk memberikan rekomendasi yang objektif, sistem menggunakan model pembobotan (*weighted scoring model*).

### 1. Variabel Penilaian

- a. **Skor Pengalaman ( $S_{exp}$ ):** Bernilai 100 jika durasi pengalaman kandidat  $\geq$  kriteria minimal. Jika kurang, dinilai proporsional.
- b. **Skor *Hard Skill* Wajib ( $S_{wajib}$ ):** Persentase jumlah *skill* wajib yang ditemukan berbanding jumlah yang diminta.
- c. **Skor *Hard Skill* Tambahan ( $S_{plus}$ ):** Persentase jumlah *skill* tambahan yang ditemukan.

### 2. Formulasi Skor Akhir

Skor akhir (*FinalScore*) dihitung dengan memberikan bobot prioritas pada pengalaman dan *skill* wajib:

$$\begin{aligned} FinalScore &= (0.4 \times S_{exp}) \\ &+ (0.4 \times S_{wajib}) \\ &+ (0.2 \times S_{plus}) \quad 3) \end{aligned}$$

### 3. Logika *Filtering* (Penyaringan)

Sistem menerapkan ambang batas (*threshold*) untuk menentukan status rekomendasi dan menyaring file yang layak unduh. Pendekatan klasifikasi tiga level berbasis rentang skor ini mengikuti praktik umum model penilaian kinerja dan seleksi yang membagi skor menjadi kategori tinggi, sedang, dan rendah, sementara nilai batas spesifik ditetapkan peneliti agar sesuai dengan konteks rekrutmen yang dikaji.

- **Rekomendasi Utama (Download List):** Jika  $FinalScore \geq 75$ .
- **Dipertimbangkan (Download List):** Jika  $50 \geq FinalScore < 75$ .
- **Tidak Cocok (Disaring/Dibuang):** Jika  $FinalScore < 50$ .

## 3. HASIL DAN PEMBAHASAN

Sistem *screening* CV yang dikembangkan berbasis Python di Google Colab dirancang untuk membantu HR menyeleksi kandidat berdasarkan posisi yang dibuka, lama pengalaman kerja, serta pemenuhan *skill* wajib dan *skill* tambahan. CV diunggah dalam format PDF, termasuk CV yang dibuat menggunakan aplikasi desain seperti

Photoshop atau Canva dengan *layout* variatif. Sistem mengekstrak teks dari file PDF tersebut, kemudian melakukan pemrosesan untuk menemukan informasi terkait posisi yang dilamar, riwayat pengalaman, serta daftar *skill* yang tercantum dalam CV. Informasi ini selanjutnya dibandingkan dengan kriteria yang telah diinput HR untuk menghasilkan skor kesesuaian kandidat.

Pada tahap pengujian, beberapa CV sampel diunggah dengan kriteria seleksi yang sama, misalnya: posisi “Staff Admin”, minimal pengalaman 2 tahun, *skill* wajib “*Ms Office*” dan “*Filing*”, serta *skill* tambahan seperti “Bahasa Inggris” dan “ERP/SAP”. Sistem menghasilkan output berupa ringkasan yang menampilkan nama kandidat, lama pengalaman kerja, jumlah *skill* wajib yang terpenuhi, jumlah *skill* tambahan yang sesuai, dan skor akhir yang digunakan HR sebagai dasar rekomendasi.

**Tabel 2. Simulasi *Screening* CV dengan posisi Staff Admin**

No	Kandidat	S_exp (Thn)	S_wajib	S_plus	FinalScore	Status
1	Rizky Pratama	5	2 dari 2	2 dari 2	98	Rekomendasi Utama
2	Siti Aminah	3	2 dari 2	1 dari 2	85	Rekomendasi Utama
3	Aditya Wijaya	4	2 dari 2	1 dari 2	88	Rekomendasi Utama
4	Putri Rahayu	2	1 dari 2	2 dari 2	72	Direkomendasikan
5	Dimas Saputra	6	2 dari 2	0 dari 2	78	Rekomendasi Utama
6	Ani Suryani	1	2 dari 2	1 dari 2	80	Rekomendasi Utama
7	Bayu Nugroho	0.5	1 dari 2	0 dari 2	45	Tidak Cocok
8	Maya Indah	3.5	2 dari 2	2 dari 2	94	Rekomendasi Utama
9	Eko Prasetyo	2	1 dari 2	1 dari 2	65	Direkomendasikan
10	Dewi Lestari	4	1 dari 2	0 dari 2	55	Direkomendasikan
11	Farhan Hakim	1.5	2 dari 2	0 dari 2	70	Direkomendasikan

12	Gita Savitri	2.5	2 dari 2	1 dari 2	83	Rekomendasi Utama
13	Hendra K.	0	0 dari 2	1 dari 2	30	Tidak Cocok
14	Irma Yunita	3	2 dari 2	0 dari 2	75	Direkomendasikan
15	Jati Perkasa	7	2 dari 2	2 dari 2	100	Rekomendasi Utama
16	Kurnia Sari	2	2 dari 2	1 dari 2	82	Rekomendasi Utama
17	Luthfi Aziz	1	1 dari 2	0 dari 2	48	Tidak Cocok
18	Mega Utami	5	1 dari 2	1 dari 2	68	Direkomendasikan
19	Nanda Putra	3	1 dari 2	2 dari 2	74	Direkomendasikan
20	Olivia Z.	2.5	2 dari 2	2 dari 2	92	Rekomendasi Utama
21	Pandu D.	4	0 dari 2	1 dari 2	40	Tidak Cocok
22	Qori Sandi	1.5	2 dari 2	1 dari 2	79	Rekomendasi Utama
23	Riana Sari	2	1 dari 2	0 dari 2	52	Direkomendasikan
24	Satria Baja	0.5	0 dari 2	0 dari 2	20	Tidak Cocok
25	Tiara Andini	3	2 dari 2	1 dari 2	85	Rekomendasi Utama
26	Udin Sedunia	10	1 dari 2	0 dari 2	60	Direkomendasikan
27	Vina P.	2	2 dari 2	0 dari 2	72	Direkomendasikan
28	Wahyu H.	4.5	2 dari 2	1 dari 2	89	Rekomendasi Utama
29	Xena W.	1	0 dari 2	2 dari 2	42	Tidak Cocok
30	Yosep S.	3.5	2 dari 2	2 dari 2	95	Rekomendasi Utama

Tabel di atas merupakan rangkuman dari proses seleksi administrasi (CV *Screening*) terhadap 30 kandidat untuk posisi Staff Admin. Penilaian dilakukan secara sistematis berdasarkan tiga parameter utama: **S\_exp** (tahun pengalaman), **S\_wajib**

(kemampuan MS Office & *filing*), dan **S\_plus** (kemampuan bahasa Inggris & software ERP).

Berdasarkan total skor yang diperoleh, kandidat dikelompokkan ke dalam tiga kategori utama:

- **Rekomendasi Utama (Skor > 75):** Terdiri dari 14 kandidat yang memiliki kombinasi pengalaman kerja yang stabil dan penguasaan *skill* wajib maupun tambahan yang sangat baik. Kelompok ini diprioritaskan untuk segera dipanggil ke tahap wawancara.
- **Direkomendasikan (Skor 50 - 75):** Terdiri dari 10 kandidat yang memenuhi kriteria dasar namun mungkin memiliki kekurangan pada pengalaman durasi kerja atau *skill* tambahan tertentu. Kelompok ini diposisikan sebagai kandidat cadangan.
- **Tidak Cocok (Skor < 50):** Terdiri dari 6 kandidat yang tidak memenuhi ambang batas kualifikasi minimum yang ditetapkan untuk posisi Admin.

Secara keseluruhan, sistem yang dibangun sudah berhasil mengotomatisasi proses *screening* awal berdasarkan posisi, pengalaman, dan pemenuhan *skill* wajib serta *skill* tambahan dengan tampilan hasil yang ringkas dan mudah dibaca HR. Hal ini membantu mengurangi waktu yang dibutuhkan untuk membaca CV satu per satu, meningkatkan konsistensi penilaian, dan memberikan dasar objektif dalam pengambilan keputusan awal sebelum tahap interview.

## 4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem *screening* CV otomatis berbasis web dengan pendekatan hibrida yang mengintegrasikan *Natural Language Processing* (NLP) untuk ekstraksi data dan mekanisme *rule-based* untuk penilaian. Berdasarkan hasil pengembangan dan pengujian, dapat disimpulkan bahwa:

- **Efektivitas dan Efisiensi:** Sistem mampu mengotomatisasi proses seleksi awal dengan mengekstraksi informasi dari CV format PDF berbahasa Indonesia, termasuk desain variatif dari perangkat seperti Canva atau Photoshop. Penggunaan sistem ini secara signifikan mengurangi waktu *screening* manual dan menekan potensi bias subjektivitas dalam penilaian awal.
- **Model Penilaian Transparan:** Melalui metode *weighted scoring* (Pengalaman Kerja 40%, *Hard Skill* Wajib 40%, dan *Hard Skill* Tambahan 20%), sistem berhasil

mengklasifikasikan kandidat ke dalam tiga kategori: **Rekomendasi Utama** (skor  $\geq 75$ ), **Dipertimbangkan** (skor 50-74), dan **Tidak Cocok** (skor  $< 50$ ). Pendekatan ini menawarkan keterjelasan logika penilaian yang lebih tinggi dibandingkan model *machine learning* murni.

- **Kontribusi Praktis:** Sistem ini memberikan kontrol penuh kepada HR untuk menyesuaikan parameter kriteria tanpa memerlukan data latih dalam jumlah besar. Namun, sistem diposisikan sebagai **alat bantu pengambilan keputusan (*decision support system*)**, di mana pengawasan manusia (*human oversight*) tetap krusial untuk melakukan interpretasi konteks dan penilaian mendalam sebelum keputusan final diambil.

Meskipun menunjukkan hasil yang menjanjikan, penelitian ini memiliki keterbatasan pada cakupan format dokumen, skala data pengujian, dan validasi ambang batas (*threshold*). Untuk pengembangan lebih lanjut, disarankan poin-poin berikut:

1. **Ekspansi Teknis Dokumen:** Memperluas kemampuan sistem agar mendukung format dokumen lain seperti DOCX dan gambar (menggunakan teknologi OCR), serta mendukung pemrosesan CV dalam bahasa asing lainnya.
2. **Validasi Skala Besar:** Melakukan pengujian menggunakan *dataset* riil dengan volume ribuan CV untuk mengevaluasi akurasi, performa, dan skalabilitas sistem dalam kondisi rekrutmen dunia nyata.
3. **Optimasi Algoritma:** Mengintegrasikan metode *machine learning* untuk mengoptimalkan pembobotan dan *threshold* skor secara adaptif berdasarkan pola data historis rekrutmen di berbagai sektor industri.
4. **Mitigasi Bias dan Fairness:** Menambahkan fitur deteksi bias otomatis dalam proses ekstraksi (seperti bias gender, usia, atau etnis) untuk menjamin keadilan (*fairness*) dalam proses seleksi.
5. **Integrasi Sistem:** Mengintegrasikan sistem ini dengan *Human Resource Information System* (HRIS) perusahaan untuk menciptakan alur kerja *end-to-end*, mulai dari penyaringan awal hingga penjadwalan wawancara secara otomatis.

Implikasi praktis dan teoretis dari penelitian ini dapat dilihat dari dua sisi yang saling melengkapi. Secara praktis, sistem *screening* CV berbasis NLP dan *rule-based* yang dikembangkan memberikan alat bantu konkret bagi HR untuk mempercepat proses

seleksi awal kandidat, karena CV dengan format variatif dapat diproses secara otomatis dan langsung dikonversi menjadi skor serta rekomendasi. Hal ini mengurangi ketergantungan pada pembacaan manual yang memakan waktu, meningkatkan konsistensi penilaian antar kandidat, dan memungkinkan HR mengalihkan fokus dari pekerjaan administratif berulang ke aktivitas yang lebih strategis seperti wawancara mendalam, asesmen lanjutan, dan pengambilan keputusan akhir. Selain itu, fleksibilitas aturan penilaian (posisi, pengalaman, *skill* wajib, dan *skill* tambahan) membuat sistem mudah disesuaikan dengan kebutuhan rekrutmen yang berbeda-beda di tiap perusahaan atau posisi.

Dari sisi teoretis, penelitian ini memperkuat gagasan bahwa pendekatan hibrida antara NLP dan *rule-based* efektif digunakan untuk tugas ekstraksi informasi dan penilaian kandidat dalam konteks *screening* CV, tanpa selalu bergantung pada model *machine learning* yang kompleks dan membutuhkan data latih besar. Penerapan *Named Entity Recognition* untuk nama, dikombinasikan dengan *keyword matching* dan pola tanggal untuk pengalaman dan *skill*, menunjukkan bahwa desain arsitektur ekstraksi yang memisahkan tipe data bebas dan berpola dapat menghasilkan sistem yang relatif akurat sekaligus mudah dijelaskan. Model pembobotan yang memadukan pengalaman, *skill* wajib, dan *skill* tambahan juga dapat menjadi rujukan bagi penelitian selanjutnya sebagai kerangka dasar penilaian kandidat, yang dapat dikembangkan lebih lanjut dengan menambahkan variabel baru (seperti *soft skill* atau kualitas portofolio), menguji variasi bobot, ataupun mengintegrasikannya dengan model AI yang lebih canggih untuk membandingkan kinerja dan tingkat interpretabilitasnya.

## DAFTAR PUSTAKA

- R. M. Affandi and D. Hermanto, "Screening CV Otomatis Dalam Natural Language Processing," vol. 1, no. 1, pp. 13–21, 2025.
- A. F. Bahar, "Model Otomatisasi Seleksi Resume Lamaran Kerja Menggunakan Machine Learning."
- N. Vanetik, A. Kolesnev, and G. Kogan, "NLP-based Screening for IT Job Vacancies : A Case Study," 2023.
- N. Sianipar, "Jurnal Penelitian Nusantara Peran Teknologi AI dalam Transformasi Fungsi Manajemen Sumber Daya Manusia Menulis : Jurnal Penelitian Nusantara," vol. 1,

no. November, pp. 31–32, 2024.

D. K. S. Reddy and A. S. Kumar, “RESUME SCREENING SYSTEM USING NATURAL LANGUAGE,” vol. 16, no. 06, pp. 487–506, 2025.

Sewang, A. N. A. Ainun, and Misrah, “Artificial intelligence dalam Rekrutmen dan Seleksi Karyawan : Manfaat dan Tantangannya,” vol. 3, 2025.

M. Yamin, “INTEGRASI KECERDASAN BUATAN DALAM PROSES REKRUTMEN SDM : SEBUAH TINJAUAN SISTEMATIS INTEGRATION OF ARTIFICIAL INTELLIGENCE IN THE HUMAN RESOURCE RECRUITMENT PROCESS : A SYSTEMATIC REVIEW Pendahuluan,” vol. 4, no. 1, 2025.

L. Hafidz, “Artificial Intelligence: Contoh Penggunaan AI di Bidang HR,” talentics.id. [Online]. Available: <https://www.talentics.id/resources/blog/artificial-intelligence-contoh-penggunaan-ai-di-bidang-hr/>

L. Zuliasyari, “Ekstraksi Data Tidak Terstruktur dengan NLP Machine Learning.” [Online]. Available: <https://dqqlab.id/ekstraksi-data-tidak-terstruktur-dengan-nlp-machine-learning>

A. Fatah, M. Faricho, A. Hidayat, M. F. Adhim, and T. Inayati, “Integrasi Teknologi Informasi Dalam Rekrutmen Dan Seleksi Tenaga Kerja : Literatur Review,” vol. 2, no. 3, pp. 549–558, 2025, doi: 10.62387/naafijurnalilmiahmahasiswa.v2i3.207.