



# **PROPOSAL PROYEK AKHIR**

## **Penerapan *Large Language Model* (LLM) untuk Rekomendasi Lowongan Pekerjaan Berdasarkan Analisis CV Pengguna**

**ARDIANSYAH INDRA FEBRIANTO**

NRP. 3322600014

### **DOSEN PEMBIMBING**

Tri Hadiyah Muliawati, S.ST., M.Kom.  
NIP. 199210122018032001

Renovita Edelani, S.ST., M.Tr. Kom.  
NIP. 199510142022032008

**PROGRAM STUDI SARJANA TERAPAN SAINS DATA TERAPAN  
DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER  
POLITEKNIK ELEKTRONIKA NEGERI SURABAYA  
2025**

“Halaman ini sengaja dikosongkan”

## DAFTAR ISI

<b>DAFTAR GAMBAR/TABEL .....</b>	<b>v</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 LATAR BELAKANG .....	1
1.2 PERMASALAHAN .....	2
1.3 BATASAN MASALAH.....	3
1.4 TUJUAN.....	3
1.5 MANFAAT .....	3
1.6 SISTEMATIKA PENULISAN.....	4
<b>BAB 2 KAJIAN PUSTAKA .....</b>	<b>7</b>
2.1 DESKRIPSI PERMASALAHAN .....	7
2.2 TEORI PENUNJANG .....	8
2.2.1 Python .....	8
2.2.2 Streamlit .....	9
2.2.3 <i>Named Entity Recognition (NER)</i> .....	9
2.2.4 <i>Natural Language Processing (NLP)</i> .....	10
2.2.5 <i>Large Language Models (LLM)</i> .....	11
2.3 PENELITIAN TERKAIT .....	12
2.3.1 Job Recommendation System Based on Skill Sets.....	12
2.3.2 Learning-Based Matched Representation System for Job Recommendation.....	13
2.3.3 NLP-Based Bi-Directional Recommendation System.....	13
2.3.4 The Multi Agent System for Job Recommendation .....	14
2.3.5 Tripartite Vector Representations for Better Job Recommendation.....	14
<b>BAB 3 DESKRIPSI SISTEM .....</b>	<b>19</b>
3.1 DESKRIPSI SOLUSI .....	19
3.2 DESAIN SISTEM .....	20

3.2.1 Pelatihan Model.....	20
3.2.2 Percobaan Model.....	27
3.2.3 Activity Diagram.....	31
3.2.3.1 Kolom Pengguna.....	31
3.2.3.2 Kolom Streamlit.....	32
3.2.3.3 Kolom Large Language Model .....	33
3.2.4 Mockup Aplikasi .....	33
<b>DAFTAR PUSTAKA .....</b>	<b>34</b>

## DAFTAR GAMBAR/TABEL

<b>Tabel 2. 1</b> Perbandingan Penelitian Terkait.....	16
<b>Gambar 3. 1</b> Desain Sistem dari solusi yang ditawarkan.....	20
<b>Gambar 3. 2</b> Proses Pelatihan Model Job Matcher.....	20
<b>Gambar 3. 3</b> Dataset Pelatihan Model dari Kaggle .....	22
<b>Gambar 3. 4</b> Templat CV ATS MS. Word .....	23
<b>Gambar 3. 5</b> Penulisan Lowongan Pekerjaan pada Job Portal.....	24
<b>Gambar 3. 6</b> Proses Percobaan Model Job Matcher .....	27
<b>Gambar 3. 7</b> Activity Diagram Sistem Rekomendasi Lowongan Pekerjaan .....	31
<b>Gambar 3. 8</b> Gambar Tampilan Aplikasi Sistem .....	33
Rekomendasi Lowongan Pekerjaan .....	33

“Halaman ini sengaja dikosongkan”

# BAB 1

## PENDAHULUAN

### 1.1 LATAR BELAKANG

Seiring dengan pesatnya perkembangan teknologi informasi, Artificial Intelligence (AI) semakin banyak diterapkan dalam berbagai bidang, termasuk dalam proses penerimaan tenaga kerja. Salah satu tantangan utama dalam dunia kerja saat ini adalah banyaknya jumlah pencari kerja yang kesulitan menemukan lowongan yang sesuai dengan keterampilan dan pengalaman yang mereka miliki. Proses pencocokan antara pencari kerja dan perusahaan seringkali memakan waktu lama dan tidak efisien karena masih banyak dilakukan secara manual.

Menurut Future of Jobs Report 2025 yang dirilis oleh World Economic Forum (WEF), diperkirakan hingga tahun 2030 akan terjadi pergeseran besar di pasar kerja global, dengan sekitar 170 juta pekerjaan baru tercipta, namun 92 juta pekerjaan juga akan tergantikan, menghasilkan pertumbuhan bersih sekitar 78 juta pekerjaan atau sekitar 7% dari total pekerjaan saat ini. Pergeseran ini merupakan dampak dari transformasi struktural pasar tenaga kerja yang dipicu oleh tren makro dan adopsi teknologi baru.

Di Indonesia, Kementerian Ketenagakerjaan (Kemnaker) secara rutin mengadakan Jobfair nasional, termasuk Jobfair Kemnaker Indonesia 2025, sebagai upaya mempertemukan pencari kerja dan perusahaan secara langsung. Namun, meskipun Jobfair menyediakan banyak peluang, masih terdapat tantangan signifikan seperti ketidaksesuaian antara profil pencari kerja dengan lowongan yang tersedia, antrian panjang, dan keterbatasan waktu dalam memilih pekerjaan yang tepat. Selain itu, persaingan talenta yang semakin ketat, perubahan kebutuhan keterampilan akibat perkembangan teknologi, serta ekspektasi kandidat yang berubah menambah kompleksitas proses rekrutmen di era ini. Kondisi ini menyebabkan banyak pencari kerja belum dapat memanfaatkan kesempatan Jobfair secara optimal, sehingga potensi pencocokan kerja yang efektif belum maksimal.

Di era digital ini, penggunaan Large Language Model (LLM) menawarkan solusi inovatif untuk menjawab tantangan tersebut. Dengan kemampuan pemrosesan bahasa alami (NLP) yang canggih, LLM dapat menganalisis isi CV pengguna secara otomatis, memahami keterampilan serta pengalaman yang mereka miliki, dan memberikan rekomendasi lowongan pekerjaan yang paling relevan dengan cepat. Sistem ini memungkinkan pencari kerja memperoleh daftar lowongan yang sesuai

hanya dengan menggunakan CV mereka, sehingga menghemat waktu dan tenaga dibandingkan dengan pencarian kerja secara konvensional.

Banyak pencari kerja yang tidak mengetahui lowongan pekerjaan yang sesuai dengan keahlian mereka atau terkadang melewatkannya peluang penting karena keterbatasan informasi. Dengan adanya sistem rekomendasi berbasis LLM, permasalahan ini dapat diminimalisir karena sistem dapat secara otomatis menyaring dan menampilkan pekerjaan yang relevan berdasarkan profil masing-masing individu. Selain itu, perusahaan juga akan mendapatkan manfaat karena dapat lebih mudah menemukan kandidat yang cocok untuk posisi yang dibutuhkan.

Penerapan LLM dalam sistem rekomendasi lowongan pekerjaan juga dapat mengatasi kekeliruan pencocokan kualifikasi pendaftar dalam seleksi lowongan pekerjaan di awal. Dengan menggunakan analisis data dan pemahaman kontekstual yang lebih baik, model ini dapat memberikan rekomendasi yang lebih objektif dan akurat dibandingkan metode tradisional yang masih bergantung pada kata kunci tanpa mempertimbangkan konteks pengalaman dan keahlian secara menyeluruh.

## 1.2 PERMASALAHAN

Berdasarkan uraian latar belakang di atas, dapat diidentifikasi dan dirumuskan permasalahan inti yang melandasi urgensi penelitian dan pengembangan sistem penerapan large language model pada rekomendasi lowongan pekerjaan, permasalahan tersebut diantaranya:

- Banyaknya lowongan yang tersedia di website lowongan pekerjaan dan pengguna membutuhkan waktu lama untuk bisa mengurutkan lowongan yang paling sesuai dengan kompetensi mereka.
- Berbagai situs lowongan kerja masih menggunakan filter dasar seperti lokasi, industri, dan gaji, sehingga belum mampu menyaring lowongan berdasarkan keterampilan spesifik yang dibutuhkan. Hal ini menyebabkan pencari kerja kesulitan menemukan posisi yang sesuai karena adanya kesenjangan keterampilan (skill gap), serta proses pencocokan kerja yang masih kurang efisien karena dilakukan secara manual atau dengan sistem yang belum akurat.
- Kurangnya teknologi yang mampu memahami konteks dari pengalaman kerja dan keahlian individu secara mendalam untuk memberikan rekomendasi yang lebih tepat.

### **1.3 BATASAN MASALAH**

Untuk memastikan bahwa penelitian ini tetap fokus dan terarah, serta untuk memberikan pemahaman yang jelas mengenai ruang lingkup dan kapabilitas sistem yang dikembangkan, maka perlu ditetapkan beberapa batasan masalah. Adapun batasan-batasan masalah dalam penelitian ini adalah sebagai berikut:

- Sistem hanya memproses Curriculum Vitae (CV) dalam format PDF digital (berbasis teks) dan ditulis dalam Bahasa Inggris. Pembatasan ini disebabkan oleh spesifikasi model NLP dan NER yang dioptimalkan untuk bahasa Inggris dan metode ekstraksi teks langsung yang tidak mendukung dokumen hasil pindaian (scan) atau format lain.
- Lingkup Domain Pekerjaan yang digunakan untuk rekomendasi lowongan pekerjaan pada sistem difokuskan di bidang informasi dan teknologi (IT).
- Analisis sistem terbatas hanya pada konten tekstual yang terdapat di dalam dokumen CV. Sistem tidak menganalisis data dari sumber eksternal seperti portofolio GitHub, profil LinkedIn, atau informasi lain di luar dokumen yang diunggah.

### **1.4 TUJUAN**

Berdasarkan permasalahan yang telah diketahui, melalui sub-bab ini akan dijabarkan tujuan utama yang hendak dicapai melalui penelitian dan pengembangan sistem penerapan large language model pada rekomendasi lowongan pekerjaan. Beberapa tujuannya sebagai berikut:

- Mengembangkan sistem rekomendasi lowongan pekerjaan yang mampu menganalisis dan memahami isi CV secara mendalam menggunakan teknologi LLM.
- Mempermudah pencari kerja dalam menemukan peluang yang sesuai dengan keterampilan dan pengalaman mereka tanpa harus melakukan pencarian manual yang memakan waktu.
- Mengatasi kekeliruan dalam pencocokan kualifikasi dengan kemampuan pencari kerja dalam proses seleksi awal dengan memberikan rekomendasi pekerjaan yang lebih objektif berdasarkan analisis terhadap CV pencari kerja.

### **1.5 MANFAAT**

Melalui penelitian dan pengembangan sistem rekomendasi lowongan pekerjaan berbasis Large Language Model (LLM), diharapkan tercipta manfaat yang tidak hanya berdampak secara praktis

bagi pencari kerja, tetapi juga memberikan kontribusi teoritis terhadap pengembangan teknologi informasi dan sistem pendukung keputusan di bidang ketenagakerjaan. Adapun manfaat yang diharapkan antara lain sebagai berikut:

- Meningkatkan peluang mendapatkan pekerjaan yang tepat sesuai dengan kompetensinya, sehingga kesempatan panggilan kerja lebih besar dibandingkan dengan pencarian konvensional.
- Memperluas cakupan peluang karier dalam jangka panjang dengan membantu pencari kerja menemukan lowongan yang relevan meskipun menggunakan istilah yang berbeda dari standar umum, berkat kemampuan pemahaman konteks dari Sistem LLM.
- Memberikan kontribusi empiris terhadap pemanfaatan Large Language Model (LLM) dan Natural Language Processing (NLP) dalam pengembangan career support systems, terutama pada aspek otomatisasi analisis dokumen profesional seperti CV dan deskripsi pekerjaan.

## 1.6 SISTEMATIKA PENULISAN

Adapun sistematika penulisan untuk Proposal Buku Proyek Akhir ini adalah:

### Bab 1 Pendahuluan

Bab ini menjelaskan tentang latar belakang yang menguraikan tantangan dalam proses rekrutmen tenaga kerja di era digital , perumusan masalah yang menyoroti inefisiensi sistem pencarian kerja saat ini, serta tujuan penelitian yang berfokus pada pengembangan sistem rekomendasi berbasis LLM. Selain itu, dipaparkan juga manfaat yang diharapkan bagi pencari kerja dan dunia akademik serta struktur penulisan proposal ini.

### Bab 2 Kajian Pustaka

Bab ini menyajikan deskripsi mendalam mengenai permasalahan ketidaksesuaian antara profil pencari kerja dan lowongan pekerjaan. Selanjutnya, dibahas teori-teori penunjang utama yang relevan dengan pengembangan sistem, seperti bahasa pemrograman Python, framework Streamlit untuk membangun antarmuka aplikasi, kemampuan NER dapat mengidentifikasi entitas, Natural Language Processing (NLP) untuk pemrosesan bahasa, dan Large Language Model (LLM) sebagai inti dari sistem rekomendasi. Bab ini ditutup dengan ulasan terhadap lima

penelitian terdahulu yang relevan sebagai dasar untuk pengembangan dan inovasi.

### **Bab 3 Desain Sistem**

Bab ini menguraikan arsitektur dan alur kerja sistem yang diusulkan. Bagian ini dimulai dengan deskripsi solusi yang menjelaskan bagaimana sistem memanfaatkan LLM untuk menganalisis CV dan mencocokkannya dengan lowongan pekerjaan dari berbagai platform secara *real-time*. Selanjutnya, dipaparkan desain sistem secara rinci, yang terdiri dari tiga komponen utama, yaitu *input* (pengguna mengunggah CV), proses (ekstraksi data melalui *library* PyPDF dan NLP, *preprocessing*, analisis oleh LLM, dan pencocokan lowongan kerja), serta *output* (penyajian hasil dalam bentuk tabel rekomendasi).

### **Bab 4 Eksperimen dan Analisis**

Bab ini akan merinci metodologi pengujian serta analisis hasil dari sistem yang dikembangkan. Bagian ini akan menjelaskan tentang skenario eksperimen yang dilakukan, dataset CV dan lowongan pekerjaan yang digunakan untuk pengujian, serta metrik evaluasi yang dipakai untuk mengukur performa sistem, seperti presisi dan relevansi rekomendasi. Lebih lanjut, bab ini akan menyajikan hasil dari eksperimen tersebut dan melakukan analisis mendalam untuk menilai efektivitas model LLM dalam memberikan rekomendasi pekerjaan yang akurat.

### **Bab 5 Penutup**

Bab ini merupakan bagian akhir dari laporan yang merangkum keseluruhan penelitian. Bagian ini akan berisi kesimpulan yang menjawab rumusan masalah dan tujuan penelitian berdasarkan hasil analisis yang telah dilakukan pada bab sebelumnya. Selain itu, bab ini juga akan menyajikan saran untuk pengembangan sistem di masa mendatang, seperti penambahan fitur baru atau penggunaan model yang lebih canggih untuk meningkatkan akurasi dan fungsionalitas sistem.

“Halaman ini sengaja dikosongkan”

## **BAB 2**

### **KAJIAN PUSTAKA**

Penelitian ini di latar belakangi oleh permasalahan pada bidang lowongan pekerjaan, di mana proses pencarian kerja masih sering kali memakan waktu dan kurang efisien bagi para pencari kerja. Salah satu tantangan utama dalam proses ini adalah sulitnya dalam mencocokkan lowongan kerja yang relevan dengan keahlian yang dimiliki oleh pelamar pekerjaan. Meskipun terdapat banyak platform pencarian kerja, sebagian besar masih mengandalkan pencocokan berbasis kata kunci dan filter dasar seperti lokasi atau gaji, sehingga belum mampu memahami konteks isi dari CV pengguna secara menyeluruh. Oleh karena itu, solusi yang diusulkan dalam penelitian ini adalah pengembangan sistem rekomendasi lowongan pekerjaan berbasis CV pengguna dengan memanfaatkan teknologi Large Language Model (LLM).

#### **2.1 DESKRIPSI PERMASALAHAN**

Perkembangan teknologi informasi yang pesat telah mengubah lanskap dunia kerja, terutama dalam hal pencarian dan penawaran lowongan pekerjaan. Banyak pencari kerja mengalami kesulitan menemukan posisi yang sesuai dengan keterampilan dan pengalaman mereka, sementara perusahaan juga menghadapi tantangan dalam menemukan kandidat yang tepat. Proses pencocokan yang masih banyak dilakukan secara manual sering kali tidak efisien dan memakan waktu, sehingga menimbulkan ketidaksesuaian antara kebutuhan pasar kerja dan ketersediaan tenaga kerja.

Menurut Future of Jobs Report 2025 yang dirilis oleh World Economic Forum, hingga tahun 2030 akan terjadi pergeseran besar di pasar kerja global, dengan terciptanya sekitar 170 juta pekerjaan baru dan hilangnya 92 juta pekerjaan lama, menghasilkan pertumbuhan bersih sekitar 78 juta pekerjaan atau sekitar 7% dari total pekerjaan saat ini. Pergeseran ini merupakan dampak transformasi struktural yang dipicu oleh tren makro dan adopsi teknologi baru, yang juga memengaruhi kebutuhan keterampilan dan pola pencarian kerja di Indonesia.

Permasalahan utama yang muncul adalah ketidaksesuaian antara profil pencari kerja dengan lowongan yang tersedia, yang menyebabkan banyak pencari kerja tidak dapat memanfaatkan peluang secara optimal. Dampak dari kondisi ini meliputi peningkatan angka pengangguran, ineffisiensi sumber daya waktu dan tenaga, serta persaingan talenta yang semakin ketat akibat perubahan kebutuhan keterampilan. Selain itu, proses seleksi yang masih bergantung pada metode tradisional sering kali

menghasilkan pencocokan yang kurang objektif dan akurat, sehingga memperburuk masalah ketidaksesuaian tersebut.

Selain tantangan teknis dari platform yang ada, terdapat pula ketidaksesuaian dari sisi konten dan format CV. Berdasarkan survei pendahuluan yang dilakukan terhadap 3 praktisi HRD di industri ekspedisi, asuransi, dan telekomunikasi, ditemukan bahwa elemen yang paling krusial saat menyaring CV melalui format ATS adalah pengalaman kerja sebelumnya, skill teknis, dan kesesuaian kata kunci dengan deskripsi pekerjaan. Survei ini juga memberikan validasi untuk rekomendasi utama yang perlu diperhatikan pencari kerja dalam menyusun CV ATS. Para praktisi HRD menyarankan agar CV disesuaikan secara spesifik dengan kata kunci yang tertera pada lowongan yang dituju, serta disajikan dalam format yang ringkas, padat informasi, singkat, dan jelas.

Sebagai solusi inovatif, penerapan Large Language Model (LLM) dalam sistem rekomendasi lowongan pekerjaan menawarkan kemajuan signifikan. Dengan kemampuan pemrosesan bahasa alami yang canggih, LLM dapat secara otomatis menganalisis isi CV, memahami konteks keterampilan dan pengalaman, serta memberikan rekomendasi lowongan yang paling relevan dengan cepat dan akurat. Sistem ini tidak hanya menghemat waktu dan tenaga pencari kerja, tetapi juga membantu perusahaan menemukan kandidat yang tepat, sehingga meningkatkan efektivitas dan efisiensi proses rekrutmen di era digital saat ini.

## 2.2 TEORI PENUNJANG

### 2.2.1 Python

Python adalah bahasa pemrograman tingkat tinggi yang dikenal karena sintaksisnya yang sederhana dan mirip dengan bahasa Inggris, sehingga mudah dipahami dan dipelajari, terutama oleh pemula. Salah satu kekuatannya adalah kemampuannya untuk ditulis seperti pseudo-code, yang memungkinkan pengembang fokus pada penyelesaian masalah daripada memahami struktur bahasa itu sendiri. Python bersifat open source dan gratis digunakan, memungkinkan siapa pun untuk mengakses, memodifikasi, dan mendistribusikan ulang kode program dengan bebas. Dengan sifatnya yang portable dan lintas platform, Python dapat dijalankan di berbagai sistem operasi seperti Windows, Linux, Mac OS, hingga sistem embedded seperti Raspberry Pi, tanpa perlu mengubah kode jika tidak bergantung pada spesifikasi platform tertentu. Selain itu, Python bersifat interpreted, yang berarti dapat langsung dijalankan dari source code tanpa proses kompilasi rumit seperti pada bahasa C atau C++. [1]

Python menyediakan pustaka standar yang sangat besar, mencakup berbagai bidang seperti operasi string, internet, alat layanan web, antarmuka sistem operasi, dan protokol. Sebagian besar tugas pemrograman yang sering digunakan telah disediakan dalam bentuk skrip di dalamnya, sehingga mengurangi panjang kode yang perlu ditulis saat menggunakan python [2]. Melalui kemampuan yang dimilikinya, menjadikan python sebagai bahasa pemrograman untuk metode-metode yang digunakan, seperti Artificial Intelligence hingga Large Language Model.

### **2.2.2 Streamlit**

Streamlit adalah framework open-source berbasis Python yang dirancang untuk memudahkan pembuatan aplikasi web interaktif, terutama digunakan oleh peneliti untuk membagikan data berskala besar dari studi ilmiah. Framework ini memungkinkan pembuatan web-app secara cepat tanpa perlu keahlian dalam pengembangan front-end seperti HTML, JavaScript, atau React, sehingga cocok bagi peneliti yang ingin menyebarluaskan metode komputasi dan hasil eksperimen kepada komunitas ilmiah secara lebih interaktif dan mudah diakses. Streamlit semakin populer karena kesederhanaannya dalam mengintegrasikan komponen visualisasi dan analisis data ke dalam antarmuka web hanya dengan beberapa baris kode Python. [3]

Streamlit memberikan berbagai keuntungan dan kemudahan yang menjadikannya sangat cocok digunakan bersama teknologi seperti Large Language Model (LLM), terutama dalam pengembangan aplikasi berbasis data dan visualisasi interaktif. Streamlit memungkinkan penyajian model dan hasil analisis secara cepat dan efisien tanpa memerlukan keahlian front-end, karena sudah dilengkapi dengan berbagai widget bawaan seperti upload file, slider, input teks, checkbox, dan radio button. Aplikasi ini bersifat open-source, ringan, dan ramah pengguna, serta dapat dijalankan menggunakan Python versi 3.7 ke atas melalui editor seperti Anaconda, PyCharm, atau Visual Studio Code. Streamlit juga memungkinkan pembaruan tampilan secara dinamis setiap kali ada interaksi pengguna karena seluruh skrip dijalankan ulang dari awal. Dengan kemampuan menampilkan visualisasi hasil analisis data menggunakan library seperti NumPy dan Pandas, Streamlit sangat mendukung integrasi dengan model pembelajaran mesin seperti kombinasi LSTM dan GRU, sehingga menjadikannya pilihan ideal untuk membangun antarmuka interaktif bagi sistem berbasis LLM. [4]

### **2.2.3 Named Entity Recognition (NER)**

*Named Entity Recognition* (NER) adalah tugas fundamental dalam *natural language processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan entitas dalam teks ke dalam kategori yang telah

ditentukan, seperti nama orang, lokasi, dan organisasi. Konsep ini pertama kali diperkenalkan dalam *Message Understanding Conference* (MUC-6) dan berfungsi sebagai langkah awal yang krusial untuk berbagai tugas NLP tingkat lanjut, termasuk *information retrieval*, sistem tanya jawab, *machine translation*, dan ekstraksi relasi. Cara kerja NER telah berevolusi dari waktu ke waktu; metode awal mengandalkan pendekatan berbasis aturan (*rule-based*) yang dirumuskan secara manual oleh para ahli. Seiring berkembangnya teknologi, pendekatan berbasis *machine learning* menjadi lebih umum, di mana model dilatih menggunakan data berlabel skala besar. Dalam pendekatan modern, NER sering kali dianggap sebagai tugas klasifikasi token (*token classification*), di mana model *encoder-only* seperti BERT digunakan untuk mengklasifikasikan setiap token dalam teks. Selain itu, pendekatan terbaru juga memperlakukan NER sebagai proses generatif, memanfaatkan kemampuan pemahaman bahasa dari model bahasa generatif untuk menghasilkan daftar entitas secara langsung.

Kelebihan utama NER yang membuatnya sangat cocok untuk analisis CV adalah kemampuannya untuk secara otomatis mengekstrak informasi kunci yang terstruktur dari dokumen teks yang tidak terstruktur seperti CV. Dalam konteks CV, NER dapat secara akurat mengidentifikasi dan mengkategorikan beragam entitas, seperti nama kandidat (Person), nama perusahaan atau institusi pendidikan (Organization), lokasi (Location), serta entitas lain yang lebih spesifik seperti keterampilan, produk, atau bahkan jabatan (Miscellaneous). Sistem NER yang modern, terutama yang menggunakan model bahasa generatif, tidak hanya sekadar mencocokkan kata kunci, tetapi juga memanfaatkan kemampuan pemahaman bahasa yang mendalam untuk menafsirkan konteks. Hal ini memungkinkan ekstraksi informasi yang jauh lebih akurat dan relevan dari narasi dalam CV, yang kemudian menjadi data penting untuk diolah lebih lanjut dalam sistem pencocokan lowongan kerja atau analisis profil kandidat. [5,6]

#### **2.2.4 Natural Language Processing (NLP)**

*Natural Language Processing* (NLP) adalah bidang interdisipliner yang bertujuan untuk memungkinkan komputer melakukan tugas-tugas yang bermanfaat yang melibatkan bahasa manusia. Ini mencakup berbagai tugas seperti memungkinkan komunikasi antara manusia dan mesin, meningkatkan komunikasi antarmanusia, dan memproses teks atau ucapan. Karakteristik utama dari sistem pemrosesan bahasa adalah penggunaan pengetahuan tentang bahasa, yang membedakannya dari aplikasi pemrosesan data lainnya. Pengetahuan ini mencakup berbagai tingkatan, mulai dari fonetik (bunyi bahasa), morfologi (komponen kata), sintaks (struktur kalimat), semantik (makna), pragmatik (tujuan penutur), hingga wacana (unit linguistik yang lebih besar dari satu ujaran). Sebagian besar tugas dalam pemrosesan ucapan

dan bahasa dapat dipandang sebagai penyelesaian ambiguitas pada salah satu tingkatan ini, di mana beberapa struktur linguistik alternatif dapat dibangun untuk suatu masukan.

Keunggulan utama NLP terletak pada kemampuannya untuk menciptakan aplikasi yang revolusioner dan bermanfaat dalam berbagai bidang. Misalnya, agen percakapan atau sistem dialog memungkinkan pengguna untuk berinteraksi dengan mesin guna membuat reservasi perjalanan atau mendapatkan informasi keberangkatan dan kedatangan. Dalam industri otomotif, teknologi pengenalan ucapan dan sintesis ucapan memungkinkan pengemudi untuk mengontrol sistem kendaraan dengan suara. Selain itu, NLP menjadi dasar bagi layanan pencarian informasi lintas bahasa dan penerjemahan otomatis, yang memungkinkan pengguna untuk mencari dan membaca informasi dari web dalam berbagai bahasa. Teknologi ini juga digunakan dalam sistem penilaian esai otomatis, agen virtual interaktif untuk pendidikan, dan analisis teks untuk mengukur opini publik dari berbagai sumber daring. Peningkatan sumber daya komputasi dan ketersediaan data yang masif dari internet terus mendorong pengembangan aplikasi pemrosesan bahasa dan ucapan yang semakin canggih. [7]

### **2.2.5 *Large Language Models (LLM)***

*Large Language Models (LLM)* adalah model yang dapat diinterpretasikan sebagai penerjemah atau compiler yang mampu mengubah spesifikasi dari bahasa alami menjadi sebuah program yang ditulis dalam bahasa pemrograman tertentu. Teknologi yang mendasari LLM modern adalah arsitektur Transformer, yang memiliki elemen inti berupa mekanisme attention. Dengan kemampuan ini, LLM dianggap dapat menyediakan sebuah lapisan abstraksi bahasa tambahan di atas bahasa pemrograman tingkat tinggi yang ada saat ini, serupa dengan bagaimana bahasa tingkat tinggi menjadi abstraksi untuk kode assembly.[8]

Keunggulan utama LLM terletak pada kemampuannya untuk meningkatkan akurasi dan kinerja secara signifikan dalam tugas-tugas yang kompleks dan terspesialisasi, bahkan di luar domain bahasa murni, seperti yang ditunjukkan pada tugas registrasi citra medis. Model seperti LLaMA 2, yang merupakan LLM dengan parameter dalam jumlah besar yang dilatih pada data korpus berskala besar, dapat diadaptasi secara efisien dengan hanya sedikit penambahan jumlah parameter [9]. LLaMA 2, dilatih pada 2 triliun token data dan menunjukkan kinerja yang tinggi di berbagai tugas NLP. Namun, perkembangannya berlanjut dengan kehadiran LLaMA 3, yang dilatih pada dataset yang jauh lebih masif, yaitu 15 triliun token. Peningkatan data pelatihan sebesar 7 kali lipat ini memberikan LLaMA 3 kemampuan

pemahaman bahasa yang lebih bermuansa dan akurat untuk rentang topik yang lebih luas.

Untuk sistem rekomendasi lowongan pekerjaan melalui analisis CV, pemilihan LLaMA 3 sebagai model dasar menawarkan keunggulan yang signifikan dibandingkan pendahulunya. Pertama, LLaMA 3 memiliki *context window* yang diperluas hingga 8.000 token, dua kali lipat dari LLaMA 2. Hal ini sangat krusial karena memungkinkan model untuk memproses dokumen CV yang panjang dan kompleks secara keseluruhan tanpa kehilangan informasi penting, sehingga pemahaman terhadap profil kandidat menjadi lebih holistik. Kedua, kemampuan penalaran (*reasoning*) dan penanganan tugas multi-langkah pada LLaMA 3 telah ditingkatkan secara signifikan. Ini memungkinkan model untuk tidak hanya mencocokkan kata kunci, tetapi juga menafsirkan hubungan kontekstual antara pengalaman kerja di CV dengan persyaratan lowongan, sehingga menghasilkan rekomendasi yang lebih cerdas dan akurat. [10]

## 2.3 PENELITIAN TERKAIT

### 2.3.1 Job Recommendation System Based on Skill Sets [11]

Penelitian oleh G. Mahalakshmi dkk. pada tahun 2022 mengembangkan sistem rekomendasi pekerjaan yang berfokus pada analisis mendalam terhadap set keterampilan pengguna dari resume mereka. Sistem ini dirancang untuk mengatasi kesulitan yang dihadapi lulusan baru dalam memilih jalur karier yang tepat dengan menyediakan rekomendasi pekerjaan yang dipersonalisasi, lengkap dengan skor kecocokan dan saran keterampilan untuk pengembangan diri. Dengan menggunakan dataset pekerjaan dari Kaggle dan resume yang dikumpulkan, metode berbasis konten ini menerapkan pra-pemrosesan NLP (*Porter Stemmer* dan *Stopwords*), mengubah teks menjadi vektor dengan *tf-idf*, dan menghitung kecocokan menggunakan *Cosine Similarity* yang terbukti paling akurat dibandingkan metode lain.

Hasilnya, sistem ini mampu menyajikan daftar pekerjaan yang diurutkan secara hierarkis berdasarkan relevansinya dengan profil pengguna, beserta visualisasi skor dalam bentuk diagram lingkaran. Fitur utamanya adalah kemampuan untuk merekomendasikan 5 keterampilan teratas yang perlu dipelajari untuk 5 pekerjaan yang paling cocok, memberikan panduan konkret bagi pengguna untuk meningkatkan daya saing mereka di pasar kerja. Sistem ini dirancang agar dapat diakses oleh siapa saja tanpa perlu login atau biaya langganan, dengan tujuan untuk mengurangi tingkat pengangguran dengan mencocokkan keterampilan individu dengan peluang yang ada.

### **2.3.2 Learning-Based Matched Representation System for Job Recommendation [12]**

Pada tahun 2022, sebuah penelitian oleh Suleiman Ali Alsaif dan timnya memperkenalkan sistem rekomendasi pekerjaan berbasis konten yang dirancang untuk membantu pencari kerja menemukan lowongan yang sesuai dengan resume mereka. Dengan mengambil data deskripsi pekerjaan dari Indeed di kota-kota besar Arab Saudi, sistem ini menggunakan teknik NLP untuk membersihkan data, word2vec untuk mengubah teks menjadi vektor, dan membandingkan dua metrik kemiripan: *Jaccard Coefficient* dan *Cosine Similarity* untuk mencocokkan keterampilan.

Hasilnya membuktikan bahwa *Cosine Similarity* jauh lebih efektif, memberikan akurasi sistem sebesar 86%, dibandingkan dengan 61% dari *Jaccard Coefficient*. Sistem ini berhasil mencocokkan 137 dari 159 resume dengan benar dan menunjukkan presisi yang sangat tinggi untuk beberapa profil pekerjaan seperti Data Scientist. Selain fungsi utamanya, sistem ini juga menyediakan analisis visual tentang tren keterampilan dan pasar kerja, menawarkan alat yang komprehensif bagi pencari kerja untuk menavigasi karier mereka.

### **2.3.3 NLP-Based Bi-Directional Recommendation System [13]**

Pada tahun 2022, Suleiman Ali Alsaif dan rekan-rekannya mempublikasikan sebuah penelitian tentang Sistem Rekomendasi Bi-directional Berbasis NLP yang dirancang untuk mempertemukan pencari kerja dengan lowongan yang sesuai dan perekrut dengan kandidat yang relevan. Dengan menggunakan dataset resume dari GitHub dan deskripsi pekerjaan dari sa.indeed.com, sistem ini menerapkan pendekatan *Content-Based Filtering*. Prosesnya melibatkan ekstraksi entitas penting seperti "keterampilan" menggunakan spaCy (NER), mengubahnya menjadi vektor dengan Word2Vec, dan menghitung kecocokan menggunakan *Cosine Similarity* untuk menjembatani kesenjangan antara bahasa yang digunakan dalam resume dan lowongan pekerjaan.

Hasil penelitian menunjukkan bahwa sistem ini sangat efektif, dengan akurasi keseluruhan mencapai 80% dalam mencocokkan profil pekerjaan dengan benar. Model ini menunjukkan kinerja yang sangat tinggi dalam mengidentifikasi entitas dari teks, dengan akurasi hingga 100% untuk kategori "Pendidikan". Sistem ini berhasil memberikan rekomendasi yang sangat akurat untuk profil pekerjaan dengan keterampilan yang spesifik (seperti PHP Developer dengan presisi 1.0), meskipun menghadapi tantangan pada profil dengan keterampilan yang tumpang tindih, seperti Data Engineer. Secara keseluruhan, penelitian ini berhasil menunjukkan bahwa pendekatan

NLP dapat secara signifikan meningkatkan akurasi dan efisiensi dalam proses rekrutmen online dua arah.

### **2.3.4 The Multi Agent System for Job Recommendation [14]**

Penelitian yang dilakukan oleh Meilany Nonsi Tentua dkk. pada tahun 2020 mengusulkan sebuah Sistem Multi-Agen (MAS) untuk rekomendasi pekerjaan guna mengatasi masalah banyaknya informasi dari berbagai portal lowongan di Indonesia. Menggunakan platform JADE, sistem ini mengerahkan agen-agen komputer otonom yang dibagi menjadi dua peran: agen pengumpul informasi yang memantau setiap portal pekerjaan (seperti Indeed.com dan karir.com) dan agen pengguna yang melayani pencari kerja. Agen pengumpul informasi secara otomatis mengambil data lowongan dan menyimpannya dalam satu database terpusat, sementara agen pengguna mencocokkan profil pengguna dengan data yang tersedia untuk memberikan rekomendasi.

Berdasarkan hasil implementasinya, sistem ini terbukti berhasil menjalankan fungsinya. Sistem mampu mengumpulkan informasi pekerjaan dari berbagai sumber, menyortir data untuk menghilangkan penawaran yang berulang, dan memberikan hasil pencarian yang sesuai dengan profil yang diinput oleh pengguna. Dengan demikian, sistem rekomendasi berbasis multi-agen ini dapat secara efektif membantu pencari kerja menemukan lowongan yang cocok dan relevan di tengah melimpahnya informasi.

### **2.3.5 Tripartite Vector Representations for Better Job Recommendation [15]**

Dalam penelitian tahun 2019, Mengshu Liu dan rekannya dari CareerBuilder memperkenalkan metode rekomendasi pekerjaan yang lebih baik dengan menciptakan tripartite vector representation, yaitu representasi data yang menggabungkan tiga aspek kunci: judul pekerjaan, keterampilan, dan lokasi. Menggunakan data dari CareerBuilder.com, metodologi ini dimulai dengan mempelajari vektor untuk judul dan keterampilan secara bersamaan dalam satu ruang laten menggunakan tiga graf informasi (job-job, skill-skill, job-skill). Selanjutnya, teknik retrofitting digunakan untuk menyempurnakan vektor judul berdasarkan keterampilan yang terkait pada suatu lowongan, sebelum akhirnya digabungkan dengan vektor lokasi 3D yang dinormalisasi dari data geospasial.

Hasilnya menunjukkan peningkatan yang sangat signifikan dalam relevansi rekomendasi. Model yang menyertakan lokasi secara eksplisit berhasil mengurangi rata-rata jarak geografis pekerjaan yang direkomendasikan hingga 90%, dengan rata-rata jarak hanya 90,4 mil. Meskipun ada sedikit penurunan pada tingkat kecocokan judul sebagai trade-

off untuk akurasi lokasi, model ini tetap lebih unggul dibandingkan metode dasar lainnya. Penelitian ini membuktikan bahwa dengan mengintegrasikan judul, keterampilan, dan lokasi secara efektif, sistem dapat memberikan rekomendasi pekerjaan yang tidak hanya relevan dari segi kualifikasi, tetapi juga sangat peka terhadap lokasi bagi pencari kerja.

**Tabel 2. 1 Perbandingan Penelitian Terkait**

Judul Jurnal	Tahun	Dataset	Preprocessing	Metode yang digunakan	Hasil Kuantitatif
<b>Job Recommendation System Based on Skill Sets</b>	2022	<ul style="list-style-type: none"> <li>- <b>Job Dataset:</b> 13.001 deskripsi pekerjaan dari Kaggle dan Google .</li> <li>- <b>Resume Dataset:</b> 101 resume .</li> <li>- <b>Skill Dataset:</b> 32 set keterampilan pekerjaan dari Google</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Stop Words Removal:</b> Menghapus kata-kata umum yang tidak bermakna .&amp;</li> <li>- <b>Porter Stemmer:</b> Mengubah kata menjadi bentuk akarnya</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Content-Based Filtering:</b> Merekomendasikan pekerjaan berdasarkan kesamaan konten .</li> <li>- <b>TF-IDF Vectorizer:</b> Mengubah teks menjadi matriks fitur numerik .</li> <li>- <b>Cosine Similarity:</b> Mengukur kesamaan antara resume dan deskripsi pekerjaan .</li> <li>- <b>Rekomendasi Keterampilan:</b> Menyarankan keterampilan yang perlu ditingkatkan</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Tidak disebutkan secara eksplisit.</b> Hasil dievaluasi secara kualitatif dengan membandingkan output dari tiga fungsi similaritas (Cosine, Euclidean, Jaccard) dengan peringkat yang diberikan oleh seorang ahli. Disimpulkan bahwa <b>Cosine Similarity</b> memberikan hasil yang paling mendekati peringkat ahli</li> </ul>
<b>Learning-Based Matched Representation System for Job Recommendation</b>	2022	Data lowongan pekerjaan di- <i>scrap</i> dari sa.indeed.com untuk beberapa jenis pekerjaan di Dammam, Jeddah, dan Riyadah	<ul style="list-style-type: none"> <li>- <b>Pembersihan Teks:</b> Menghapus tag, tokenisasi, <i>lemmatization</i>, dan menghapus <i>stop words</i></li> </ul>	<ul style="list-style-type: none"> <li>- <b>Content-Based Filtering:</b> Menganalisis konten untuk merekomendasikan item .</li> <li>- <b>Word Embedding (Word2vec):</b> Mengubah teks menjadi representasi vektor .</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Perbandingan Akurasi:</b> Cosine Similarity (CS) secara signifikan lebih baik dengan akurasi <b>0.86 (86%)</b>, dibandingkan Jaccard Similarity (JC) dengan akurasi <b>0.61 (61%)</b> .</li> </ul>

				<ul style="list-style-type: none"> <li>- <b>Perbandingan Similarity:</b> Membandingkan <b>Jaccard Similarity (JC)</b> dan <b>Cosine Similarity (CS)</b> untuk pencocokan</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Presisi (dengan CS):</b> Presisi mencapai 1.0 untuk <i>Data Scientist</i>, 0.91 untuk <i>Sales</i>, dan 0.81 untuk <i>Network Security Engineer</i></li> </ul>
<b>NLP-Based Bi-Directional Recommendation System</b>	2022	<p><b>Lowongan Pekerjaan:</b> Data di-scrap dari sa.indeed.com di kota-kota besar Arab Saudi.</p> <p><b>Resume:</b> 138 resume untuk training dan 25 untuk testing dari repositori GitHub (DataTurks-Engg)</p>	<ul style="list-style-type: none"> <li>- <b>Pembersihan Teks:</b> Menghapus tag HTML, karakter non-ASCII, tanda baca, dan <i>stop words</i>.</li> <li>- <b>Lemmatization:</b> Mengubah kata ke bentuk dasarnya</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Bi-directional Recommendation:</b> Merekomendasikan pekerjaan ke pencari kerja dan resume ke perekrut .</li> <li>- <b>NER (Named Entity Recognition):</b> Menggunakan spaCy untuk mengekstrak entitas seperti Keterampilan, Lokasi, dan Pendidikan.</li> <li>- <b>Word2vec &amp; Cosine Similarity:</b> Untuk mencocokkan antara resume dan pekerjaan</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Akurasi Model NER:</b> Akurasi untuk entitas <i>Skills</i> adalah 99.08%, <i>Name</i> 99.88%, <i>Location</i> 99.77%, dan <i>Education</i> 100% .</li> <li>- <b>Akurasi Sistem:</b> Akurasi keseluruhan sistem adalah 0.8 (80%). Sebanyak 20 dari 25 resume berhasil diklasifikasikan dengan benar</li> </ul>
<b>The Multi Agent System for Job Recommendation</b>	2020	Data dikumpulkan dari berbagai portal pekerjaan online seperti indeed.com, jobsdb.com, dan karir.com	<ul style="list-style-type: none"> <li>- <b>Pengumpulan &amp; Penyaringan:</b> Mengumpulkan informasi dari berbagai portal dan menyaring pekerjaan yang sama</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Multi-Agent System (MAS):</b> Menggunakan platform JADE .</li> <li>- <b>Information Gathering Agent:</b> Bertugas</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Tidak disebutkan secara eksplisit.</b> Jurnal ini lebih fokus pada arsitektur dan implementasi sistem. Disimpulkan bahwa sistem dapat memberikan hasil pencarian yang sesuai dengan input pengguna, namun tidak</li> </ul>

			agar hanya ditampilkan sekali	mengumpulkan informasi pekerjaan dari portal online .  - <b>User Agent:</b> Bertugas mencari pekerjaan di database berdasarkan profil pengguna	ada metrik kuantitatif seperti akurasi atau presisi yang disajikan
<b>Tripartite Vector Representations for Better Job Recommendation</b>	2019	Data pekerjaan dan pengguna dari CareerBuilder.com, mencakup 300.000 lowongan pekerjaan aktual	<p>- <b>Parsing:</b> Mengekstrak Judul Pekerjaan, Keterampilan, dan Lokasi dari lowongan dan resume .</p> <p>- <b>Normalisasi:</b> Vektorisasi dan normalisasi informasi lokasi (latitude &amp; longitude) menjadi vektor 3D</p>	<p>- <b>Representation Learning:</b> Membuat representasi vektor gabungan untuk judul dan keterampilan menggunakan tiga graf informasi (job-job, skill-skill, job-skill) dengan <i>Bayesian Personalized Ranking</i> (BPR) .</p> <p><b>Retrofitting:</b> Menggabungkan dan menyempurnakan vektor judul dan keterampilan .</p> <p><b>FAISS:</b> Pencarian kesamaan (similarity search) yang efisien untuk vektor</p>	<p><b>Jarak Rekomendasi:</b> Model <i>Retrofitter-loc</i> mengurangi rata-rata jarak hingga 90.417 mil (penurunan 90%) .</p> <p><b>Kecocokan Judul:</b> Model <i>Retrofitter-loc</i> mencapai tingkat kecocokan judul 14.1%, lebih baik dari model dasar (sekitar 11%). Model <i>Retrofitter-no loc</i> mencapai 68.9%</p>

## **BAB 3**

### **DESKRIPSI SISTEM**

#### **3.1 DESKRIPSI SOLUSI**

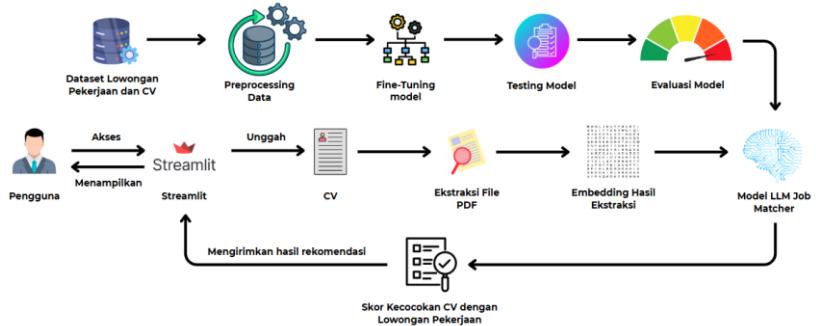
Solusi yang diusulkan untuk mengatasi permasalahan pencocokan lowongan pekerjaan adalah dengan mengembangkan sebuah sistem rekomendasi pekerjaan cerdas yang dirancang untuk mengatasi inefisiensi dan ketidakakuratan dalam proses pencocokan kerja saat ini. Sistem ini memanfaatkan kekuatan Large Language Model (LLM) untuk menganalisis konten Curriculum Vitae (CV) secara mendalam, memahami konteks kualifikasi, keterampilan, dan pengalaman pengguna untuk memberikan rekomendasi lowongan yang sangat relevan.

Alur kerja dimulai saat pengguna mengunggah CV mereka melalui antarmuka web interaktif yang dibangun dengan Streamlit. Setelah diterima, sistem secara otomatis mengekstrak seluruh teks dari dokumen PDF. Teks mentah ini kemudian diproses melalui serangkaian tahapan Natural Language Processing (NLP), yaitu dimulai dengan informasi kunci seperti keterampilan (skills) diekstraksi menggunakan Named Entity Recognition (NER), berikutnya hasil ekstraksi tersebut dibersihkan dan distandardkan melalui proses normalisasi untuk memastikan konsistensi data dan teks yang sudah bersih diubah menjadi representasi vektor numerik melalui proses embedding.

Inti dari solusi ini adalah penggunaan Model LLM Job Matcher yang telah dioptimalkan. Embedding dari profil pengguna dikirim ke model ini, yang kemudian melakukan proses pencocokan canggih dengan membandingkannya terhadap data lowongan pekerjaan yang diambil secara real-time dari berbagai platform melalui API. Model akan menghitung skor kuantitatif untuk setiap lowongan, yang merepresentasikan tingkat kecocokan antara profil pengguna dan persyaratan pekerjaan. Hasil akhir berupa daftar rekomendasi yang telah diurutkan berdasarkan skor tertinggi kemudian disajikan kembali kepada pengguna melalui antarmuka Streamlit, lengkap dengan detail dan tautan langsung ke lowongan tersebut. Dengan demikian, solusi ini tidak hanya menghemat waktu pencari kerja tetapi juga meningkatkan peluang mereka untuk menemukan pekerjaan yang tepat sesuai kompetensi.

### 3.2 DESAIN SISTEM

Dalam proyek akhir ini, desain sistem di bawah ini berdasarkan Gambar 3.1



Gambar 3.1 Desain Sistem dari solusi yang ditawarkan

Gambar 3.1 merupakan diagram alur sistem rekomendasi lowongan pekerjaan berbasis CV yang mengintegrasikan proses fine-tuning model LLM, ekstraksi data, serta antarmuka pengguna melalui Streamlit. Sistem ini dimulai dari proses pengumpulan dataset CV dan deskripsi pekerjaan, yang kemudian diproses melalui tahapan preprocessing dan fine-tuning model LLM hingga siap digunakan untuk pencocokan data. Pengguna dapat mengakses sistem melalui antarmuka Streamlit untuk mengunggah CV berformat PDF, yang selanjutnya akan diekstrak dan diubah menjadi representasi embedding. Embedding tersebut kemudian dibandingkan dengan data lowongan yang ada menggunakan model LLM Job Matcher untuk menghasilkan skor kecocokan dan rekomendasi pekerjaan. Integrasi antara proses pembelajaran LLM dengan alur pengguna memungkinkan sistem memberikan hasil pencocokan yang relevan dan akurat, sehingga mendukung efisiensi dalam proses pencarian kerja secara otomatis dan cerdas.

#### 3.2.1 Pelatihan Model



Gambar 3.2 Proses Pelatihan Model Job Matcher

## **1. Dataset Lowongan Pekerjaan dan CV**

Tahap ini adalah titik awal dari seluruh proses pengembangan model. Sebuah Dataset yang terdiri dari kumpulan *Curriculum Vitae* (CV) dan deskripsi lowongan pekerjaan dikumpulkan sebagai bahan baku utama. Dataset ini dirancang untuk merepresentasikan data dunia nyata yang akan dihadapi oleh sistem nantinya. Kualitas dan keragaman data dalam dataset ini sangat krusial, karena akan menjadi dasar bagi model untuk belajar dan diuji kemampuannya.

Pengumpulan dataset ini bisa melibatkan pengambilan data dari berbagai sumber, seperti portal pekerjaan publik (misalnya Indeed.com), repositori data (seperti Kaggle), atau kumpulan resume dari platform seperti GitHub. Tujuan utamanya adalah memiliki korpus data yang cukup besar dan representatif untuk melatih model agar dapat mengenali berbagai pola dan istilah yang umum digunakan dalam CV dan lowongan pekerjaan.

Secara khusus, CV yang berbasis ATS (Applicant Tracking System) adalah dokumen yang dirancang untuk dapat terbaca dan terstruktur dengan baik oleh sistem otomatis. CV jenis ini umumnya menggunakan format standar seperti teks bersih (tanpa tabel kompleks atau desain grafis) dan menyoroti bagian penting seperti pengalaman kerja sebelumnya, skill teknis, dan kesesuaian kata kunci dengan deskripsi pekerjaan. Dengan demikian, sistem rekomendasi yang dikembangkan diharapkan mampu memproses CV ATS ini secara efektif, mengekstrak informasi penting, dan mencocokkannya dengan kebutuhan dari deskripsi lowongan pekerjaan.

NER Annotated CVs

Data Card Code (1) Discussion (0) Suggestions (0)

**cv (1).pdf** (42.55 kB)

About this file Suggest Edits

This file does not have a description yet.

**Data Explorer**  
672.08 MB

- ResumesJsonAnnotated
- ResumesJsonAnnotated
- ResumesPDF
- cv (1).pdf
- cv (10).pdf
- cv (100).pdf
- cv (1000).pdf
- cv (1001).pdf
- cv (1002).pdf
- cv (1003).pdf
- cv (1004).pdf
- cv (1005).pdf
- cv (1006).pdf
- cv (1007).pdf
- cv (1008).pdf
- cv (1009).pdf
- cv (101).pdf
- cv (1010).pdf
- cv (1011).pdf
- cv (1012).pdf
- cv (1013).pdf
- cv (1014).pdf

**Gambar 3.3** Dataset Pelatihan Model dari Kaggle

(Sumber: <https://www.kaggle.com/datasets/mehyarmlaweh/ner-annotated-cvs>)

Format data disesuaikan dengan kebutuhan analisis teks dan proses ekstraksi informasi, seperti menggunakan file PDF atau JSON dengan atribut yang terstandarisasi. Sebagai contoh, untuk CV ATS, struktur data dapat mencakup kolom seperti *nama kandidat, riwayat pendidikan, pengalaman kerja, keahlian teknis, sertifikasi, dan ringkasan profil*, sementara untuk deskripsi lowongan pekerjaan mencakup *judul posisi, deskripsi pekerjaan, kualifikasi, lokasi, dan industri*. Secara keseluruhan, jumlah data yang digunakan mencapai lebih dari 10.064 pasang data CV, baik dari data sintetis maupun hasil ekstraksi sumber publik. Struktur ini dirancang agar model dapat memahami dan mencocokkan informasi secara efektif. Format yang digunakan merupakan format yang direkomendasikan oleh HRD agar terbaik oleh sistem serta sesuai standar rekrutmen yang umum digunakan. Format ini dapat ditemukan pada penggunaan Templat Microsoft Word. Dengan contoh, sebagai berikut:

<p><b>Shir Rosenstein</b></p> <p>PHLEBOTOMIST</p> <hr/> <p><b>EDUCATION</b></p> <p>Bellows College   Associates in Phlebotomy Aug 2004 — May 2005</p> <p>Completed courses in all areas to Phlebotomy, Venipuncture &amp; Techniques, Phlebotomy Techniques, Laboratory Procedures and Equipment, and Medical Law and Ethics as part of associates degree in Phlebotomy program.</p> <hr/> <p><b>PROFESSIONAL EXPERIENCE</b></p> <p>Phlebotomist   Lumen Healthcare Jan 2005 — present</p> <p>This is the place for a summary of your key responsibilities and most stellar accomplishments</p> <p>Phlebotomist   Wholeness Healthcare Feb 2004 — Jan 2005</p> <p>This is the place for a summary of your key responsibilities and most stellar accomplishments</p> <p>Phlebotomist   Swiss Scrips Healthcare Dec 2003 — Jun 2004</p> <p>This is the place for a summary of your key responsibilities and most stellar accomplishments</p>	<p><b>CONTACT INFO</b></p> <p>4321 Middle Ave Seattle, WA 98141 Tel: 206-555-1234 sheatoni@example.com</p> <hr/> <p><b>SKILLS &amp; ABILITIES</b></p> <p>Performing venipunctures and phlebotomy procedures Professionally managing working in a medical laboratory Excellent customer service Thorough knowledge of anatomy and patient privacy regulations</p> <hr/> <p><b>Rewards</b></p> <p>Recognized by supervisor and colleagues for excellent performance and dedication to detail. Successfully implemented a new system at pharmaceutical firm in national health organization.</p>	<p><b>DONNA ROBBINS</b></p> <p>4547 Main Street, Detroit, MI 48127   (012) 555-0100   donna@example.com   www.great辭辭ness.com</p> <p>An analytical, organized and detail-oriented accountant with GAAP expertise and experience in the full spectrum of public accounting. Collaborative team player with ownership mentality and a track record of delivering the highest quality strategic solutions to resolve challenges and propel business growth.</p> <p><b>EXPERIENCE</b></p> <p><b>20XX – Present</b> Accountant   Reuter Research   San Francisco, CA</p> <p>Working in a mid-sized public accounting firm to provide professional accounting services for individual and business clients. Provide full range of services, include income tax preparation, audit support, preparative of financial statements, pro forma budgeting, general ledger accounting, and bank reconciliation.</p> <p><b>20XX – 20XX</b> Bookkeeper   Reuter Real Estate   Berkeley, CA</p> <p>Responsible for a real estate investment company. Maintained financial books, tracked expenses, prepared and submitted invoices, and processed payroll.</p> <p><b>December 20XX – April 20XX</b> Accounting Intern   Olson Marine LM   Valencia, CA</p> <p>Assisted with payroll and Pensions service management for 150+ employees. Prepared invoices for more than 200 clients. Assisted with call management, records organization and preparation, and other office duties to support finance and accounting operations.</p> <p><b>EDUCATION</b></p> <p><b>June 20XX</b> Bachelor of Science in Accounting, Minor in Business Administration   Bellows College</p> <ul style="list-style-type: none"> <li>▪ Distinguished member of university's Accounting Society</li> <li>▪ Relevant coursework: Advanced Financial Accounting and Reporting</li> <li>▪ GPA: 3.8</li> </ul> <p><b>SKILLS</b></p> <ul style="list-style-type: none"> <li>▪ Microsoft Office Dynamics</li> <li>▪ Cashflow planning &amp; management</li> <li>▪ State &amp; Federal tax codes</li> <li>▪ Bookkeeping</li> <li>▪ Exceptional communication</li> <li>▪ Fluent in German</li> </ul>
<p><b>Andree Rocher</b></p> <p>PROFESSIONAL TITLE</p> <p><b>Philadelphia, PA   705.555.0121   andree@example.com</b></p> <p><b>Objective</b></p> <p>To obtain a challenging data scientist position in a dynamic and innovative organization where I can use my technical and analytical skills.</p> <hr/> <p><b>Experience</b></p> <p>Fluorite Corp   Data Scientist <span style="float: right;">20XX – 20XX</span></p> <ul style="list-style-type: none"> <li>▪ Increased customer retention by 20%</li> <li>▪ Reduced healthcare charges by 50%</li> <li>▪ Developed machine learning models for prediction</li> <li>▪ Drive business decisions and reduced operational costs</li> </ul> <p>Perfume Labs   Junior Data Scientist <span style="float: right;">20XX – 20XX</span></p> <ul style="list-style-type: none"> <li>▪ Cleaned and preprocessed data</li> <li>▪ Conducted statistical tests for decision support</li> <li>▪ Developed machine learning models for prediction</li> <li>▪ Collected and ran A/B tests. Increased click-through</li> </ul> <p><b>Education</b></p> <p>Jasper University, M.S. Data Science <span style="float: right;">20XX</span></p> <p>Major: Data Science   Minor: Machine Learning</p> <p>Believe College, B.S. Mathematics <span style="float: right;">20XX</span></p> <p>Major: Mathematics   Minor: Computer Science</p> <p><b>Skills &amp; Abilities</b></p> <ul style="list-style-type: none"> <li>▪ Management</li> <li>▪ Problem solving</li> <li>▪ Communication</li> <li>▪ Leadership</li> </ul>	<p><b>DANIELLE BRASSEUR</b></p> <p>4547 1st Avenue, Carroll City, NV 89111   (001) 555-0000   danielle@example.com   www.brassurus.com</p> <p>Dynamics and detail-oriented accountant with expertise in GAAP and comprehensive public accounting experience. Thrived in assisting department managers, and tracking business growth through effective collaboration and leadership.</p> <hr/> <p><b>EDUCATION</b></p> <p>Bachelor of Science in Accounting, Minor in Business Administration   Bellows College</p> <p>Reuter Research   Bachelor of Science in Accounting, Minor in Business Administration   Bellows College</p> <ul style="list-style-type: none"> <li>▪ Distinguished member of university's Accounting Society, GPA: 3.8</li> <li>▪ Relevant coursework: Advanced Financial Accounting and Reporting</li> </ul> <p><b>EXPERIENCE</b></p> <p>Accountant   Tilly's Network   Los Angeles, CA May 2005 – Present</p> <ul style="list-style-type: none"> <li>▪ Proven accounting services for individuals and businesses in a non-client basis.</li> <li>▪ Specialize in income tax preparation, audit support, and financial statement preparation.</li> <li>▪ Manage pre-formula budgeting, general ledger accounting, and bank reconciliation.</li> </ul> <p>Bookkeeper   Banner Real Estate   Berkeley, CA May 2004 – April 2005</p> <ul style="list-style-type: none"> <li>▪ Bookkeeping services for a real estate development company.</li> <li>▪ Maintained financial books, tracked expenses</li> <li>▪ Prepared invoices for over 200 clients</li> <li>▪ Assisted with call management, records organization and preparation, and other office duties</li> </ul> <p>Accounting Intern   Olson Marine LM   Valencia, CA December 2003 – April 2004</p> <ul style="list-style-type: none"> <li>▪ Assisted with payroll and Pensions service management for 150+ employees</li> <li>▪ Prepared invoices for over 200 clients</li> <li>▪ Assisted with call management, records organization and preparation, and other office duties</li> </ul>	<p><b>DONNA ROBBINS</b></p> <p>4547 Main Street, Detroit, MI 48127   (012) 555-0100   donna@example.com   www.great辭辭ness.com</p> <p>An analytical, organized and detail-oriented accountant with GAAP expertise and experience in the full spectrum of public accounting. Collaborative team player with ownership mentality and a track record of delivering the highest quality strategic solutions to resolve challenges and propel business growth.</p> <p><b>EXPERIENCE</b></p> <p><b>20XX – Present</b> Accountant   Reuter Research   San Francisco, CA</p> <p>Working in a mid-sized public accounting firm to provide professional accounting services for individual and business clients. Provide full range of services, include income tax preparation, audit support, preparative of financial statements, pro forma budgeting, general ledger accounting, and bank reconciliation.</p> <p><b>20XX – 20XX</b> Bookkeeper   Reuter Real Estate   Berkeley, CA</p> <p>Responsible for a real estate investment company. Maintained financial books, tracked expenses, prepared and submitted invoices, and processed payroll.</p> <p><b>December 20XX – April 20XX</b> Accounting Intern   Olson Marine LM   Valencia, CA</p> <p>Assisted with payroll and Pensions service management for 150+ employees. Prepared invoices for more than 200 clients. Assisted with call management, records organization and preparation, and other office duties to support finance and accounting operations.</p> <p><b>EDUCATION</b></p> <p><b>June 20XX</b> Bachelor of Science in Accounting, Minor in Business Administration   Bellows College</p> <ul style="list-style-type: none"> <li>▪ Distinguished member of university's Accounting Society</li> <li>▪ Relevant coursework: Advanced Financial Accounting and Reporting</li> <li>▪ GPA: 3.8</li> </ul> <p><b>SKILLS</b></p> <ul style="list-style-type: none"> <li>▪ Microsoft Office Dynamics</li> <li>▪ Cashflow planning &amp; management</li> <li>▪ State &amp; Federal tax codes</li> <li>▪ Bookkeeping</li> <li>▪ Exceptional communication</li> <li>▪ Fluent in German</li> </ul>

**Gambar 3. 4 Templat CV ATS MS. Word**

(Sumber: microsoft word)

Berdasarkan hasil survei yang dilakukan terhadap beberapa HRD dari berbagai perusahaan, ditemukan bahwa struktur informasi yang dicantumkan dalam deskripsi lowongan pekerjaan umumnya mengikuti pola yang konsisten dan dianggap paling efektif dalam menarik serta menyaring

kandidat yang sesuai. HRD menyebutkan bahwa elemen yang paling sering mereka sertakan adalah judul posisi, diikuti oleh gambaran umum peran dan tanggung jawab pekerjaan, serta kualifikasi wajib, seperti latar belakang pendidikan, pengalaman kerja, dan keterampilan teknis atau soft skill tertentu. Selain itu, HRD juga menambahkan informasi seperti lokasi penempatan, jenis pekerjaan (misalnya full-time atau remote), dan deskripsi singkat tentang perusahaan. Dalam beberapa kasus, disertakan pula kisaran gaji, benefit tambahan, serta kata kunci atau tag yang relevan untuk membantu visibilitas lowongan di platform pencarian kerja. Temuan dari survei ini menjadi dasar dalam menyusun struktur data deskripsi pekerjaan dalam dataset, agar lebih representatif terhadap praktik rekrutmen yang sebenarnya dilakukan oleh pihak perusahaan. Dan berikut merupakan beberapa contoh penulisan pada laman pencarian lowongan pekerjaan.

#### About the role

Join PT NNR RPX Global Logistics Indonesia' as a **SALES ASSISTANT MANAGER**. In this full-time role based in South Jakarta Jakarta, you will be responsible for supporting the sales team and contributing to the overall growth and success of the company within the Forwarding & Logistics industry.

#### What you'll be doing

- Assist Manager to manage Sales team in prospecting, qualifying, and closing new business opportunities
- Preparing proposals, managing customer relationships, and tracking sales activities
- Identifying and pursuing new business leads through market research, networking, and other sales and marketing activities
- Collaborating with the overseas networking for business expansion
- Ensure seamless delivery of services to customers by good cooperation with Operation Division
- Providing regular sales and performance reports to the management team
- Participating in the development and implementation of sales strategies and initiatives

#### What we're looking for

- Minimum 5 years of Sales experience in Freight Forwarding and Logistics industry is Mandatory
- Sales target oriented and ready to grab business promptly
- Well understanding about Incoterm and Forwarding knowledge
- Excellent problem-solving and analytical skills
- Familiarity with import/export and customs regulations and procedures
- Proficient in English speaking and writing and able to use Sales software
- A proven track record of achieving sales targets and contributing to the growth of an organisation

#### What we offer

At PT NNR RPX Global Logistics Indonesia', we are committed to creating a supportive and rewarding work environment for our employees. We offer a competitive salary, opportunities for career development, and a range of benefits to support your work-life balance.

#### About us

PT NNR RPX Global Logistics Indonesia' is a leading provider of comprehensive logistics services, specialising in import/export and customs solutions. With a strong presence across Indonesia and a reputation for excellence, we are dedicated to delivering innovative and reliable services to our clients.

### Gambar 3. 5 Penulisan Lowongan Pekerjaan pada Job Portal (Sumber:

<https://id.jobstreet.com/id/job/85174987?type=standard&ref=search-standalone&origin=jobCard#sol=38e024f07486b436a4a37f35ea03f57dbe88eac>)

## **2. Preprocessing Data**

Data mentah yang telah dikumpulkan dari dataset seringkali "kotor" dan tidak terstruktur, sehingga tidak bisa langsung digunakan untuk melatih model. Oleh karena itu, tahap Preprocessing Data sangat penting. Proses ini melibatkan serangkaian teknik pembersihan dan normalisasi untuk mengubah data mentah menjadi format yang bersih, konsisten, dan terstruktur agar siap diolah oleh model.

Tahapan preprocessing ini dapat mencakup beberapa hal, seperti pembersihan teks dari karakter atau tag yang tidak relevan (misalnya tag HTML), tokenisasi (memecah teks menjadi unit-unit kecil), penghapusan *stop words* (kata-kata umum yang tidak memiliki makna signifikan seperti "dan", "di", "atau"), serta *lemmatization* atau *stemming* untuk mengubah kata ke bentuk dasarnya. Tujuan akhirnya adalah untuk meningkatkan kualitas data, yang pada gilirannya akan meningkatkan efektivitas dan akurasi model selama proses pelatihan.

## **3. Fine-Tuning Model**

Tahap Fine-Tuning adalah proses di mana sebuah model bahasa besar (LLM) yang sudah ada (pre-trained) diadaptasi secara khusus untuk tugas pencocokan pekerjaan. Alih-alih melatih model dari nol yang akan memakan sumber daya sangat besar, pendekatan ini memanfaatkan pengetahuan umum yang sudah dimiliki oleh model seperti LLaMA 3, lalu melatihnya lebih lanjut menggunakan dataset CV dan lowongan pekerjaan yang telah diproses sebelumnya.

Seperti mana yang telah disampaikan mengenai kelebihan LLM, model LLM modern dapat diadaptasi secara efisien untuk tugas-tugas spesifik. Selama fine-tuning, model belajar mengenali nuansa, terminologi, dan pola-pola spesifik yang ada dalam domain rekrutmen. Proses ini "mengasah" kemampuan generalis model menjadi kemampuan spesialis yang sangat relevan untuk menganalisis CV dan mencocokkannya dengan deskripsi pekerjaan secara akurat.

## **4. Testing Model**

Setelah model selesai melalui tahap fine-tuning, langkah selanjutnya adalah Testing Model. Pada tahap ini, performa model diuji menggunakan sebagian dari dataset yang telah disisihkan khusus untuk pengujian (data yang belum pernah dilihat model selama pelatihan). Proses pengujian ini mensimulasikan bagaimana model akan bekerja di dunia nyata saat dihadapkan dengan data baru.

Skenario eksperimen yang dirancang akan dijalankan pada tahap ini. Model akan diberikan input dari dataset pengujian (misalnya, sebuah CV) dan diminta untuk menghasilkan output (rekomendasi pekerjaan). Hasil atau prediksi yang dibuat oleh model ini kemudian akan dicatat untuk dievaluasi pada tahap berikutnya. Tahap pengujian ini sangat penting untuk menilai sejauh mana model dapat menggeneralisasi pengetahuannya, bukan sekadar menghafal data latihan.

## 5. Evaluasi Model

Evaluasi Model adalah tahap di mana hasil prediksi dari fase pengujian diukur secara kuantitatif untuk menentukan seberapa baik performa model. Proses ini melibatkan perbandingan antara output yang dihasilkan oleh model dengan "jawaban yang benar" atau *ground truth* yang telah ditentukan sebelumnya untuk dataset pengujian.

Untuk mengukur kinerja ini, berbagai metrik evaluasi digunakan, seperti presisi dan relevansi. Metrik lain yang umum digunakan dalam penelitian terkait adalah akurasi, yang mengukur persentase prediksi yang benar. Hasil dari evaluasi ini memberikan gambaran yang objektif tentang efektivitas model dalam menjalankan tugasnya. Skor dari metrik-metrik inilah yang menjadi dasar untuk memutuskan apakah model sudah siap untuk digunakan atau masih memerlukan iterasi penyempurnaan lebih lanjut.

## 6. Model LLM Job Matcher

Model LLM Job Matcher adalah hasil akhir dari keseluruhan pipeline pengembangan dan pelatihan ini. Ini adalah versi model yang telah disempurnakan (fine-tuned), diuji, dan divalidasi kinerjanya. Model ini bukan lagi model LLM generik, melainkan sebuah model spesialis yang telah dioptimalkan untuk tugas spesifik, yaitu menganalisis CV dan memberikan rekomendasi pekerjaan yang relevan.

Model inilah yang kemudian akan diintegrasikan ke dalam sistem aplikasi *live* yang diakses oleh pengguna. Ketika seorang pengguna mengunggah CV mereka ke aplikasi Streamlit, "Model LLM Job Matcher" inilah yang akan bekerja di belakang layar untuk melakukan analisis dan menghasilkan rekomendasi. Model ini merupakan aset inti yang menjadi "otak" dari solusi cerdas yang ditawarkan.

### 3.2.2 Percobaan Model



Gambar 3. 6 Proses Percobaan Model Job Matcher

#### 1. Pengguna

Pengguna dalam sistem ini adalah para pencari kerja yang menghadapi berbagai tantangan dalam menemukan lowongan yang sesuai dengan kompetensi mereka. Mereka seringkali harus menghabiskan waktu yang lama untuk menyaring informasi secara manual dan berhadapan dengan sistem pencarian kerja konvensional yang filternya masih sangat dasar, seperti lokasi atau industri, tanpa mampu memahami konteks keahlian secara spesifik.

Kondisi ini menciptakan adanya kesenjangan keterampilan (*skill gap*) dan inefisiensi dalam proses pencocokan kerja. Oleh karena itu, pengguna sistem ini adalah individu yang membutuhkan solusi cerdas untuk menjembatani kesenjangan tersebut, dengan harapan dapat menemukan peluang kerja yang paling relevan dengan profil unik mereka secara cepat dan efisien.

#### 2. Akses & Menampilkan

Proses "Akses" dan "Menampilkan" menggambarkan interaksi dua arah yang dinamis antara pengguna dan sistem melalui antarmuka aplikasi. "Akses" adalah langkah inisiasi di mana pengguna membuka aplikasi untuk memulai sesi pencarian kerja. Ini adalah titik masuk di mana pengguna akan berinteraksi dengan semua fitur yang disediakan oleh sistem.

"Menampilkan" adalah langkah balasan dari sistem kepada pengguna, yang terjadi di akhir alur kerja. Setelah seluruh proses analisis selesai, sistem akan menampilkan hasil akhir berupa daftar rekomendasi pekerjaan yang telah diurutkan dan dipersonalisasi. Interaksi dua arah ini memastikan bahwa pengguna tidak hanya dapat memberikan input, tetapi

juga menerima output yang jelas dan bermanfaat sebagai solusi dari permasalahan mereka.

### 3. Streamlit

Streamlit adalah *framework open-source* berbasis Python yang dipilih untuk membangun antarmuka pengguna (UI) dari aplikasi web ini. Alasan utama penggunaannya adalah karena Streamlit memungkinkan pengembangan aplikasi web yang interaktif dengan cepat, bahkan tanpa memerlukan keahlian mendalam di bidang *front-end* seperti HTML atau JavaScript.

*Framework* ini sangat ideal untuk proyek berbasis data dan *machine learning* seperti ini, karena menyediakan berbagai *widget* bawaan yang fungsional, salah satunya adalah fitur untuk mengunggah file. Kemampuannya untuk secara dinamis memperbarui tampilan berdasarkan interaksi pengguna menjadikannya pilihan yang sangat cocok untuk menyajikan hasil analisis dari model LLM secara efektif dan interaktif.

### 4. Unggah CV

Langkah "Unggah CV" merupakan aksi inti yang dilakukan oleh pengguna untuk memberikan data utama kepada sistem. Dalam tahap ini, pengguna mengunggah dokumen *Curriculum Vitae* (CV) mereka, yang berfungsi sebagai sumber data primer yang tidak terstruktur. Dokumen inilah yang akan menjadi dasar dari seluruh proses analisis dan rekomendasi yang dipersonalisasi.

Tujuan dari sistem ini adalah untuk menganalisis dan memahami isi CV secara mendalam, termasuk kualifikasi, pengalaman kerja, dan riwayat pendidikan yang tercantum di dalamnya. Dengan mengunggah CV, pengguna secara efektif menyerahkan profil profesional mereka untuk diurai dan dipahami oleh teknologi LLM, sehingga sistem dapat mulai bekerja untuk menemukan peluang yang paling relevan.

### 5. Ekstraksi File PDF

Setelah CV berhasil diunggah, sistem segera memulai tahap teknis pertama, yaitu "Ekstraksi File PDF". Pada langkah ini, sistem memproses file PDF untuk menarik keluar seluruh konten teksual yang ada di dalamnya. Proses ini adalah bagian dari tahapan *input* dan *preprocessing* data, di mana dokumen yang tadinya dalam format file diubah menjadi data teks mentah yang siap untuk dianalisis lebih lanjut.

Alur kerja ini berfokus pada ekstraksi teks langsung dari PDF digital dengan menggunakan library PyPDF. Proses ekstraksi ini penting untuk memastikan bahwa semua informasi relevan dari CV mulai dari detail kontak hingga deskripsi pengalaman kerja dapat dibaca dan diproses oleh komponen NLP pada tahap selanjutnya.

## 6. Embedding Hasil Ekstraksi

Setelah teks berhasil diekstraksi, langkah selanjutnya adalah "Embedding". Ini adalah proses fundamental di mana data teks mentah yang tidak terstruktur diubah menjadi representasi numerik (vektor) yang padat makna. Model bahasa besar seperti LLM tidak dapat memproses teks secara langsung, mereka beroperasi pada angka. Proses *embedding* inilah yang menjembatani kesenjangan tersebut.

Setiap kata atau frasa dalam CV dipetakan ke dalam vektor yang menangkap hubungan semantik dan kontekstualnya. Ini adalah keunggulan utama dari sistem berbasis LLM, karena memungkinkan model untuk memahami "makna" di balik kualifikasi pengguna, bukan sekadar mencocokkan kata kunci secara harfiah. Hasil dari tahap ini adalah profil pengguna yang terstruktur secara numerik dan siap untuk dianalisis oleh inti pemrosesan model.

## 7. Proses Pelatihan Model LLM Job Matcher

"Proses Pelatihan Model LLM" adalah sebuah alur kerja *offline* yang krusial dan dilakukan selama fase pengembangan, terpisah dari interaksi pengguna secara langsung. Tahap ini melibatkan pengujian dan evaluasi model menggunakan dataset yang telah ditentukan, yang terdiri dari berbagai sampel CV dan deskripsi lowongan pekerjaan.

Tujuan dari proses ini adalah untuk mengukur dan meningkatkan kinerja model sebelum diterapkan dalam sistem produksi. Kinerja model dievaluasi menggunakan metrik kuantitatif seperti presisi dan relevansi untuk memastikan bahwa rekomendasi yang dihasilkan akurat dan dapat diandalkan. Hasil dari analisis eksperimental ini kemudian digunakan sebagai dasar untuk menyempurnakan model, sehingga mampu memberikan hasil yang optimal bagi pengguna akhir.

## 8. Model LLM Job Matcher

"Model LLM Job Matcher" adalah komponen inti dan "otak" dari keseluruhan sistem rekomendasi ini. Model ini merupakan sebuah *Large Language Model* yang telah dilatih untuk memiliki kemampuan pemahaman bahasa alami yang canggih. Tugas utamanya adalah menerima data

*embedding* dari CV pengguna dan melakukan analisis pencocokan yang mendalam terhadap database lowongan pekerjaan.

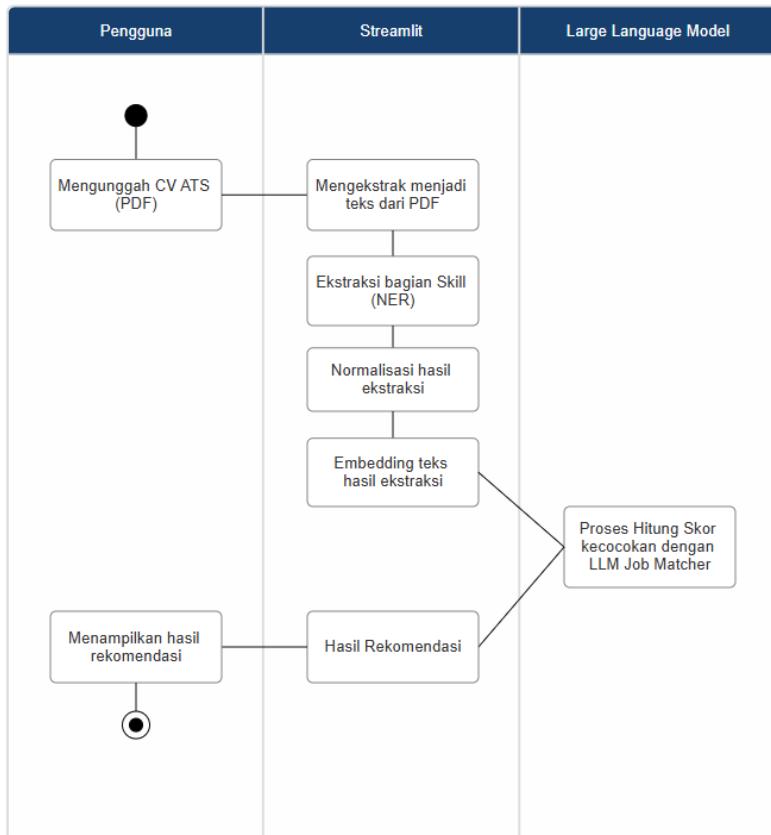
Tidak seperti sistem tradisional, model ini mampu memahami konteks, nuansa, dan hubungan antara berbagai keterampilan dan pengalaman yang tercantum dalam CV. Secara bersamaan, sistem ini juga terintegrasi dengan API dari platform pekerjaan untuk mendapatkan data lowongan secara *real-time*. Kombinasi inilah yang memungkinkan model untuk memberikan rekomendasi yang sangat relevan dan akurat.

## 9. Skor Kecocokan & Mengirimkan Hasil Rekomendasi

Tahap ini adalah puncak dari seluruh alur kerja, di mana hasil analisis dari Model LLM diubah menjadi output yang bermanfaat bagi pengguna. Model akan menghasilkan "Skor Kecocokan" untuk setiap lowongan pekerjaan, yang merepresentasikan tingkat relevansi antara profil pengguna dengan kualifikasi yang dibutuhkan.

Hasil-hasil ini kemudian diagregasi dan diurutkan, lalu dikirim kembali ke antarmuka Streamlit untuk ditampilkan kepada pengguna. Output akhir disajikan dalam format yang mudah dipahami, seperti tabel rekomendasi, yang tidak hanya menampilkan pekerjaan yang cocok tetapi juga dilengkapi dengan informasi detail dan tautan langsung ke laman asli lowongan tersebut. Dengan ini, siklus interaksi selesai, dan pengguna mendapatkan solusi konkret untuk membantu proses pencarian kerja mereka.

### 3.2.3 Activity Diagram



Gambar 3. 7 Activity Diagram Sistem Rekomendasi Lowongan Pekerjaan

#### 3.2.3.1 Kolom Pengguna

1. Mengunggah CV ATS (PDF) Langkah ini adalah titik awal dari seluruh alur kerja, di mana pengguna memulai interaksi dengan sistem. Pengguna secara aktif mengunggah dokumen *Curriculum Vitae* (CV) mereka dalam format PDF melalui antarmuka aplikasi. Tindakan ini merupakan *input* utama yang akan memicu semua proses analisis otomatis yang dirancang untuk mengatasi kesulitan pencari kerja dalam menemukan lowongan yang sesuai secara manual.

- Menampilkan hasil rekomendasi Ini adalah langkah akhir dari perspektif pengguna, di mana mereka menerima output dari sistem. Setelah semua proses analisis dan pencocokan selesai, hasil akhirnya ditampilkan kepada pengguna. Tujuannya adalah menyajikan daftar lowongan pekerjaan yang paling relevan dengan cara yang mudah dipahami, sehingga pengguna dapat dengan mudah mengambil tindakan selanjutnya. Proses ini menutup siklus interaksi, memberikan tanggapan langsung kepada pengguna.

### 3.2.3.2 Kolom Streamlit

- Mengekstrak menjadi teks dari PDF Setelah pengguna mengunggah CV, komponen aplikasi (Streamlit) akan memulai proses teknis pertama. Aktivitas ini melibatkan ekstraksi seluruh konten tekstual dari dokumen PDF. Langkah ini sangat penting untuk mengubah dokumen yang semula dalam format file menjadi data teks mentah yang dapat diproses oleh langkah-langkah *Natural Language Processing* (NLP) selanjutnya.
- Ekstraksi bagian Skill (NER) Setelah mendapatkan teks mentah, sistem menggunakan teknik *Named Entity Recognition* (NER). NER adalah tugas fundamental dalam NLP yang secara cerdas mengidentifikasi dan mengklasifikasikan entitas kunci dari dalam teks, seperti nama perusahaan, lokasi, dan yang terpenting, keterampilan (skills). Proses ini memungkinkan sistem untuk memahami isi CV secara terstruktur, tidak hanya sebagai kumpulan kata-kata.
- Normalisasi hasil ekstraksi Entitas yang telah diekstraksi, terutama keterampilan, kemudian melalui tahap normalisasi. Proses ini bertujuan untuk membersihkan dan menstandarkan teks agar lebih konsisten, misalnya dengan mengubah berbagai variasi kata menjadi bentuk dasarnya (*lemmatization*). Normalisasi sangat penting untuk memastikan bahwa "pemrograman," "memprogram," dan "program" dianggap sebagai konsep yang sama, sehingga meningkatkan akurasi pada tahap pencocokan.
- Embedding teks hasil ekstraksi Pada tahap ini, teks yang sudah bersih dan ternormalisasi diubah menjadi representasi numerik yang disebut *embedding*. Proses ini menerjemahkan kata dan frasa menjadi vektor yang dapat dipahami oleh *Large Language Model*. Vektor *embedding* ini menangkap makna dan konteks semantik dari setiap keterampilan, yang memungkinkan LLM melakukan analisis yang jauh lebih mendalam daripada sekadar pencocokan kata kunci.

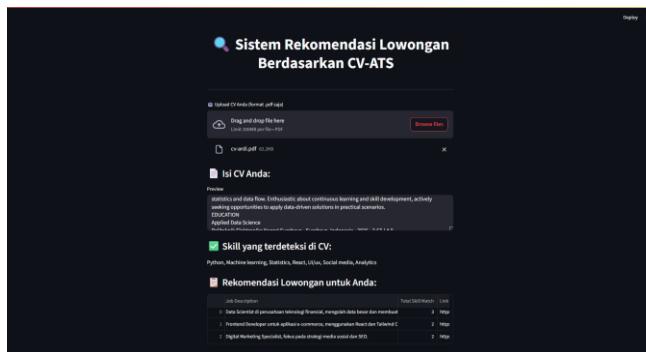
- Hasil Rekomendasi Setelah LLM selesai menghitung skor, hasilnya dikirim kembali ke aplikasi Streamlit. Komponen ini bertugas menerima data skor dan daftar pekerjaan yang sudah diurutkan dari LLM. Sebelum ditampilkan kepada pengguna, data ini mungkin akan diformat terlebih dahulu agar lebih mudah dibaca, misalnya disusun dalam bentuk tabel yang rapi.

### 3.2.3.3 Kolom Large Language Model

- Proses Hitung Skor kecocokan dengan LLM Job Matcher Ini adalah inti dari keseluruhan sistem, di mana kecerdasan buatan bekerja. Data *embedding* dari CV pengguna dikirim ke *Large Language Model* (LLM). Model ini kemudian melakukan proses pencocokan yang canggih, membandingkan profil pengguna dengan kualifikasi yang dibutuhkan oleh setiap lowongan pekerjaan yang datanya diambil secara *real-time*. Untuk setiap lowongan, LLM akan menghitung dan memberikan skor kuantitatif yang merepresentasikan tingkat kecocokan.

### 3.2.4 Mockup Aplikasi

Untuk memberikan gambaran awal mengenai tampilan dan alur interaksi pengguna dengan sistem, dibuatlah desain mockup sebagai representasi visual dari antarmuka aplikasi. Mockup ini berfungsi sebagai panduan awal dalam proses pengembangan. Tampilan dirancang agar mudah dipahami oleh pengguna dari berbagai latar belakang. Contoh tampilan antarmuka yang dimaksud ditunjukkan pada Gambar 3.8.



**Gambar 3. 8** Gambar Tampilan Aplikasi Sistem Rekomendasi Lowongan Pekerjaan

“Halaman ini sengaja dikosongkan”

## DAFTAR PUSTAKA

- [1] A.L.Sayeth Saabith, MMM.Fareez, T.Vinothraj, **Python current trend applications-an overview**, *International Journal of Advance Engineering and Research Development*, Vol. 06, No. 10, Hal. 7-8, 2019.
- [2] Harshita Sharma , Ravindra Soni, "**Python: An Appropriate Language For Real World Programming**", Iconic Research And Engineering Journals Volume 1 Issue 9 2018 Page 250-253.
- [3] Nápoles-Duarte JM, Biswas A, Parker MI, Palomares-Baez JP, Chávez-Rojo MA and Rodríguez-Valdez LM (2022), **Stmol: A component for building interactive molecular visualizations within streamlit web-applications**. *Front. Mol. Biosci.* 9:990846.
- [4] Widi Hastomo, Nur Aini, Adhitio Satyo Bayangkari Karno, L.M. Rasdi Rere, **Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management**, *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, Vol. 11, No. 2, Hal. 131-139, 2022.
- [5] Zhang, Ying., Xiao, Gang., **Named Entity Recognition Datasets: A Classification Framework**, *International Journal of Computational Intelligence Systems*, Vol. 17, No. 71, Hal. 1-17, Springer, 2024.
- [6] Deußer, Tobias., La., Hillebrand, Lars., Bauckhage, Christian., Sifa, Rafet., **Informed Named Entity Recognition Decoding for Generative Language Models**, *IEEE International Conference on Big Data (BigData)*, Hal. 1-13, 2024.
- [7] Jurafsky, Daniel, Martin, James H., **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**, Pearson Prentice Hall, Edisi Kedua, 2008.
- [8] Kanerika, **Llama 3 vs Llama 2: A Detailed Comparison**, Kanerika Blog, <https://kanerika.com/blogs/llama-3-vs-llama-2/>, Diakses tanggal 23 Juni, 2025.
- [9] Alsaif, Suleiman Ali, Sassi Hidri, Minyar, Eleraky, Hassan Ahmed, Ferjani, Imen, Amami, Rimah, **Learning-Based Matched Representation System for Job Recommendation**, *Computers*, Vol. 11, No. 161, Hal. 1-18, MDPI, 2022.
- [10] Liu, Mengshu, Wang, Jingya, Abdelfatah, Kareem, Korayem, Mohammed, **Tripartite Vector Representations for Better Job Recommendation**, *DI2KG '19*, Anchorage, Alaska, USA, 2019.

- [11] Ma, Mingrui, Yang, Yu, ***LLaMA-Reg: Using LLaMA 2 for Unsupervised Medical Image Registration***, arXiv preprint arXiv:2405.18774, 2024.
- [12] Alsaif, Suleiman Ali, Sassi Hidri, Minyar, Ferjani, Imen, Eleraky, Hassan Ahmed, Hidri, Adel, **NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters**, **Big Data and Cognitive Computing**, Vol. 6, No. 147, Hal. 1-17, MDPI, 2022.
- [13] Marcondes, Francisco S., Almeida, José João, Novais, Paulo, **Large Language Models: Compilers for the 4th Generation of Programming Languages?**, 12th Symposium on Languages, Applications and Technologies (SLATE 2023), Hal. 10:1-10:8, 2023.
- [14] Tentua, Meilany Nonsi, Azhari, Azhari, Musdholifah, Aina, **The Multi Agent System for Job Recommendation**, **The 7th International Conference on DV-Xa Method**, Hal. 1-7, 2020.
- [15] Mahalakshmi, G., Kumar, A. A., Senthilnayaki, B., Duraimurugan, J., **Job Recommendation System Based On Skill Sets**, **International Journal of Creative Research Thoughts (IJCRT)**, Vol. 10, No. 8, Hal. a770-a785, IJCRT.ORG, 2022.