

**SISTEM LAYANAN REKOMENDASI PEKERJAAN  
MENGUNAKAN ALGORITMA INDOBERT DAN COSINE  
SIMILARITY**

**LAPORAN TUGAS AKHIR**

Laporan ini disusun guna memenuhi salah satu syarat untuk menyelesaikan Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



**NELNA RUSTIA LAZUANTI**

**NIM 32602000052**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM SULTAN AGUNG  
SEMARANG**

**2024**

LEMBAR PENGESAHAN  
TUGAS AKHIR

SISTEM LAYANAN REKOMENDASI PEKERJAAN MENGGUNAKAN  
ALGORITMA INDOBERT DAN COSINE SIMILARITY

NELNA RUSTIA LAZUANTI  
NIM 32602000052

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir  
Program Studi Teknik Informatika  
Universitas Islam Sultan Agung  
Pada tanggal : 2 Desember 2024

TIM PENGUJI UJIAN SARJANA :

Imam Much Ibnu S, ST, M.Sc, Ph.D  
NIDN.0613037301

(Ketua Penguji)

Moch Taufik, ST, MIT  
NIDN.0622037502

(Anggota Penguji)

Badie'ah, ST, M.Kom  
NIDN.0619018701

(Pembimbing 1)

Sam Farisa Chaerul H, ST, M.Kom  
NIDN. 0628028602

(Pembimbing 2)

Semarang, 5 Desember 2024

Mengetahui,

Kaprodi Teknik Informatika  
Universitas Islam Sultan Agung

Moch Taufik, ST, MIT  
NIDN.0622037502

## SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Nelna Rustia Lazuanti

NIM : 32602000052

Judul Tugas Akhir : Sistem Layanan Rekomendasi Pekerjaan Menggunakan Algoritma IndoBERT dan Cosine Similarity

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 4 Desember 2024

Yang Menyatakan,

UNISSULA

جامعة سلطان أبوبنح الإسلامية



Nelna Rustia Lazuanti

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Nelna Rustia Lazuanti

NIM : 32602000052

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : Sistem Layanan Rekomendasi Pekerjaan Menggunakan Algoritma IndoBERT dan Cosine Similarity

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 4 Desember 2024



Nelna Rustia Lazuanti



## KATA PENGANTAR

Dengan nama Allah Yang Maha Pengasih dan Penyayang, segala puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan hidayah-nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “Sistem Layanan Rekomendasi Pekerjaan Menggunakan Algoritma IndoBERT dan *Cosine Similarity*” ini untuk memenuhi salah satu syarat menyelesaikan studi (S-1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang, dengan rasa rendah hati penulis menyampaikan rasa terimakasih yang sedalam-dalamnya kepada yang terhormat :

1. Rektor UNISSULA Bapak Prof.Dr.H.Gunarto, S.H.,M.H yang telah mengizinkan penulis mencari ilmu di kampus ini
2. Ibu Dr. Novi Marlyana,S.T., M.T., IPU., Asean Eng Selaku Dekan Fakultas Teknologi Industri.
3. Bapak Ir. Sri Mulyono, M.Eng. Selaku Kaprodi Program Studi Teknik Informatika.
4. Ibu Badie'ah, ST, M.Kom selaku dosen pembimbing I yang memberikan ilmu serta nasehat kepada penulis.
5. Bapak Sam Farisa Chairul H, ST, M.Kom selaku dosen pembimbing II yang membantu dan memberi ilmu kepada penulis.
6. Bapak dan Ibu yang telah memberikan kasih sayang dan dukungan kepada penulis.
7. Semua pihak yang ikut terlibat dan tidak dapat penulis sebutkan satu persatu, atas bantuan dan support yang telah diberikan kepada penulis.

Dengan segala kerendahan hati, penulis menyadari masih terdapat banyak kekurangan ilmu pengetahuan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi terbentuknya laporan yang lebih baik.

Semarang, 1 Desember 2024

Nelna Rustia Lazuanty



## DAFTAR ISI

<b>SISTEM LAYANAN REKOMENDASI PEKERJAAN MENGGUNAKAN ALGORITMA INDOBERT DAN COSINE SIMILARITY .....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>SURAT PERNYATAAN KEASLIAN TUGAS AKHIR .....</b>	<b>iii</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....</b>	<b>iv</b>
<b>KATA PENGANTAR .....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR TABEL.....</b>	<b>ix</b>
<b>DAFTAR GAMBAR.....</b>	<b>x</b>
<b>ABSTRAK .....</b>	<b>xi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	2
1.3 Pembatasan Masalah .....	2
1.4 Tujuan.....	3
1.5 Manfaat.....	3
1.6 Sistematika Penulisan.....	3
<b>BAB II TINJAUAN PUSTAKA DAN DASAR TEORI .....</b>	<b>5</b>
2.1 Tinjauan Pustaka .....	5
2.2 Dasar Teori.....	8
2.2.1. Sistem Rekomendasi .....	8
2.2.2. <i>Curriculum Vitae</i> (CV) .....	9
2.2.3. <i>Glints</i> .....	10
2.2.4. <i>Web Scraping</i> .....	10
2.2.5. <i>Text Preprocessing</i> .....	11
2.2.6. <i>Transformer</i> .....	12
2.2.7. BERT.....	15
2.2.8. IndoBERT .....	16
2.2.9. <i>Cosine Similarity</i> .....	17
2.2.10. Evaluasi Model.....	17

<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>17</b>
3.1    Metode Penelitian.....	17
3.1.1.    Studi Literatur .....	17
3.1.2.    Pengumpulan data .....	17
3.1.3.    Alur Perancangan Sistem .....	18
3.1.4.    Gambaran Sistem .....	19
3.1.5.    Perancangan <i>User Interface</i> .....	21
3.1.6.    Kebutuhan Perangkat Lunak .....	24
<b>BAB IV HASIL DAN ANALISIS PENELITIAN .....</b>	<b>26</b>
4.1    Hasil Perancangan Model.....	26
4.1.1. Deskripsi <i>Dataset</i> .....	26
4.1.2. Data <i>Preprocessing</i> .....	29
4.1.1.1. <i>Case Folding</i> .....	29
4.1.1.2. <i>Tokenizing</i> .....	30
4.1.1.3. <i>Stopword Removal</i> .....	30
4.1.1.4. <i>Stemming</i> .....	31
4.1.3. Hasil Perhitungan Pencocokan .....	32
4.1.4. Hasil Pelatihan Model.....	34
4.1.5. Hasil Evaluasi .....	34
4.2    Hasil Implementasi Sistem.....	36
4.3    Hasil Pengujian Sistem.....	39
4.3.1. Pengujian Performa Sistem.....	39
4.3.2. Pengujian <i>Black Box</i> .....	40
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>26</b>
<b>DAFTAR PUSTAKA .....</b>	<b>36</b>



## DAFTAR TABEL

Tabel 2. 1 Tabel Tinjauan Pustaka.....	6
Tabel 4. 1 Tabel CV Pelamar.....	27
Tabel 4. 2 Tabel hasil Similarity .....	32
Tabel 4. 3 Hasil evaluasi AP@K .....	34
Tabel 4. 4 Tabel hasil Precision dan Recall .....	35



## DAFTAR GAMBAR

Gambar 2. 1 Arsitektur Transformers(Vaswani dkk., 2017) .....	13
Gambar 2. 2 Multi-head Attention pada Transformers(Vaswani dkk., 2017) .....	14
Gambar 2. 3 Arsitektur Pre-Training dan Fine-Tuning pada BERT(Devlin dkk., 2019) .....	15
Gambar 3. 1 Preprocessing .....	18
Gambar 3. 2 Flowchart alur sistem .....	20
Gambar 3. 3 Gambaran beranda pada sistem.....	21
Gambar 3. 4 Interface pengisian form pelamar.....	22
Gambar 3. 5 Interface upload CV .....	23
Gambar 3. 6 Halaman hasil rekomendasi sistem .....	23
Gambar 3. 7 Instalasi Library Pandas .....	24
Gambar 3. 8 Instalasi library Fitz.....	25
Gambar 3. 9 Instalasi Library Sastrawi.....	25
Gambar 4. 1 Dataset hasil web scraping lowongan pekerjaan.....	26
Gambar 4. 2 Tampilan data kandidat .....	28
Gambar 4. 3 Data sebelum dilakukan case folding.....	29
Gambar 4. 4 Data setelah dilakukan case folding .....	29
Gambar 4. 5 Data setelah dan sebelum proses tokenizing .....	30
Gambar 4. 6 Data sebelum dilakukan proses stopword removal .....	31
Gambar 4. 7 Data setelah dilakukan proses stopword removal .....	31
Gambar 4. 8 Data setelah dan sebelum dilakukan proses stemming .....	32
Gambar 4. 10 Halaman utama.....	37
Gambar 4. 11 Tampilan implementasi halaman untuk upload CV user .....	37
Gambar 4. 12 Tampilan implementasi halaman pengisian form data diri .....	37
Gambar 4. 13 Tampilan hasil rekomendasi pekerjaan .....	38

## ABSTRAK

Sistem rekomendasi merupakan sebuah *website* yang bertujuan untuk merekomendasikan sebuah lowongan pekerjaan sesuai dengan harapan pengguna. Untuk bisa merekomendasikan sebuah lowongan maka dibutuhkan data kecocokan antara lowongan pekerjaan yang tersedia dengan data diri pelamar. Sistem ini akan menganjurkan pengguna untuk mengisikan data diri dalam bentuk form atau CV. Dengan adanya data user ini maka sistem akan melakukan sebuah embedding menggunakan metode IndoBERT yang akan merubah teks menjadi sebuah vektor yang kemudian data pelamar dengan lowongan pekerjaan akan di uji kecocokan nya menggunakan *Cosine Similarity* untuk dapat merekomendasikan sebuah hasil lowongan pekerjaan yang sesuai dengan data user. Dengan menggunakan *Cosine Similarity* akan menguji kecocokan yang jika hasilnya mendekati angka 1 maka bisa dinyatakan bahwa data semakin cocok. Hasil evaluasi menggunakan AP @K menunjukkan bahwa rata rata precision yaitu 0,733 dan rata rata recallnya yaitu 0,0156 yang mengartikan bahwa pemodelan cukup baik dalam merekomendasikan lowongan pekerjaan yang sesuai data diri pelamar.. Uji performa juga perlu dilakukan setelah sistem berhasil, dengan tujuan agar performa sistem tidak mengalami kendala bug ataupun lainnya. Kecepatan koneksi yang digunakan user juga akan mempengaruhi kecepatan sistem dalam merekomendasikan lowongan pekerjaan kepada pengguna.

**Kata Kunci** : Rekomendasi Pekerjaan, IndoBERT, *Cosine Similarity*

## ABSTRACT

*A recommendation system is a website that aims to recommend job vacancies according to user expectations. To be able to recommend a vacancy, match data is needed between the available job vacancies and the applicant's personal data. This system will prepare users to fill in personal data in the form or CV. With this user data, the system will carry out an embedding using the IndoBERT method which will convert the text into a vector and then the applicant data with job vacancies will be tested for suitability using Cosine Similarity to be able to recommend a job vacancy result that matches the user data. Using Cosine Similarity will test the suitability, if the result is close to 1, it can be stated that the data is increasingly suitable. The evaluation results using AP @K show that the average precision is 0.733 and the average recall is 0.0156, which means that the modeling is quite good in recommending job vacancies that match the applicant's personal data. Performance tests also need to be carried out after the system is successful, with the aim of ensuring that System performance does not experience bugs or other problems. The connection speed used by the user will also affect the speed of the system in providing job vacancies to users.*

**Keywords** : Job Recommendation. IndoBERT, *Cosine Similarity*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pertumbuhan industri teknologi informasi terus berkembang dengan pesat, pengembangan sistem akan dibutuhkan agar mempermudah masyarakat umum. Dengan semakin berkembangnya tingkat kependudukan akan berpengaruh kepada banyaknya permintaan tenaga kerja dan jumlah pekerjaan yang dibutuhkan. Proses pencarian pekerjaan telah mengalami transformasi yang signifikan di era digital yang semakin berkembang. Berbagai *platform online* yang menawarkan layanan pencarian pekerjaan, menjadikan internet menjadi salah satu sumber utama bagi mereka yang mencari pekerjaan. Meskipun demikian, masalah baru muncul bersamaan dengan kemajuan ini. Salah satunya adalah kesulitan bagi pencari kerja untuk menemukan pekerjaan yang sesuai dengan minat dan preferensi mereka.

Sistem rekomendasi pekerjaan sendiri ialah sistem yang akan membantu pencari kerja menemukan pekerjaan yang paling sesuai dengan minat, kualifikasi, dan preferensi mereka. Sistem ini menganalisis data pencari kerja dan lowongan pekerjaan dengan berbagai teknik, dan kemudian memberikan rekomendasi pekerjaan yang sesuai (Habibi dkk., 2022).

Salah satu metode yang menarik perhatian penulis dalam pengembangan sistem rekomendasi adalah IndoBERT (Indonesian BERT), sebuah model bahasa berbasis *Transformer* yang telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami dalam bahasa Indonesia. Dengan menggunakan IndoBERT dalam konteks sistem rekomendasi pekerjaan, kualitas rekomendasi dapat ditingkatkan dengan mempertimbangkan konteks dan karakteristik bahasa Indonesia secara lebih baik (Saadah dkk., 2022).

Pada sistem ini mengambil dataset pada *platform online* yaitu Glints, pada Glints ini terdapat database mengenai lowongan pekerjaan. Dataset diambil dengan menggunakan proses yaitu *web scraping* yang kemudian akan diolah menggunakan metode IndoBERT untuk menyajikan informasi yang berkesinambungan dengan CV pelamar.

Sistem ini dirancang sebagai platform online yang memungkinkan pengguna mencari pekerjaan dengan lebih efisien dan akurat. Tujuan penulis adalah untuk meningkatkan kualitas rekomendasi pekerjaan yang disesuaikan dengan preferensi dan kebutuhan individu dalam lingkungan negara Indonesia.

## 1.2 Perumusan Masalah

Bagaimana performa metode IndoBERT dan *Cosine Similarity* ini dalam proses rekomendasi pekerjaan dan seberapa efektif sistem layanan rekomendasi pekerjaan menggunakan metode IndoBERT dan *Cosine Similarity* dalam menyajikan rekomendasi pekerjaan yang sesuai.

## 1.3 Pembatasan Masalah

Adapun batasan masalah pada tugas akhir ini adalah :

1. *Dataset* yang digunakan diambil dari sebuah platform yang digunakan untuk mencari lowongan pekerjaan yaitu glints. Platform glints tersebut merupakan situs yang menyediakan database pekerjaan dan konten yang informatif mengenai karir dan saran pencarian pekerjaan.
2. Setelah dataset didapatkan pada platform glint akan melalui proses *Web Scraping*, proses *Web Scraping* ini dilakukan untuk mengambil database berupa judul pekerjaan, deskripsi pekerjaan, informasi perusahaan, kualifikasi dan persyaratan pekerjaan. Data yang diambil berjumlah 3 dengan kategori *web development*, *mobile development* dan *data analyst*. Dan masing masing kategori jenis pekerjaan diambil 50 data sehingga seluruh data berjumlah 150 dataset.
3. Sistem yang dibuat hanya untuk *platform website*.
4. IndoBERT hanya digunakan untuk embedding pada tugas akhir ini.

1

---

<sup>1</sup> <https://glints.com/id>



#### 1.4 Tujuan

Tujuan tugas akhir ini adalah mengembangkan sistem rekomendasi menggunakan metode IndoBERT dan *Cosine Similarity* sehingga sistem tersebut akan merekomendasikan data lowongan pekerjaan dengan kategori *Web Development, Mobile Development, Data Analis* yang sesuai pada data yang user kirimkan.

#### 1.5 Manfaat

Pada penelitian ini diharapkan dapat memberi manfaat yang besar kepada pengguna untuk mempermudah dalam pencarian lowongan yang sesuai dengan data diri pelamar, sehingga diharapkan bisa meningkatkan *value* diterimanya pelamar pada lowongan yang direkomendasikan.

#### 1.6 Sistematika Penulisan

Adapun sistematika penulisan yang akan digunakan. penulis dalam menyusun laporan tugas akhir ini adalah sebagai berikut :

**BAB I : PENDAHULUAN**

Pada bab ini penulis akan menjelaskan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan serta sistematika penulisan penelitian.

**BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI**

Pada bab ini penulis akan mencantumkan penelitian-penelitian terdahulu dan dasar teori yang berguna untuk membantu memahami teori yang berkaitan dengan metode IndoBERT dan *Cosine Similarity* pada penelitian ini.

**BAB III : METODE PENELITIAN**

Pada bab ini penulis akan menjelaskan tahapan-tahapan penelitian dimulai dari pengumpulan data hingga pada proses uji kesamaan dataset.

**BAB IV : HASIL DAN ANALISIS PENELITIAN**

Pada bab ini akan menjelaskan mengenai hasil penelitian yaitu

uji kesamaan menggunakan metode IndoBERT dan *Cosine Similarity*.

## BAB V : KESIMPULAN DAN SARAN

Pada bab ini penulis akan memberikan kesimpulan dari proses penelitian dari awal hingga akhir.



## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Sistem rekomendasi merupakan sistem yang akan mengolah data untuk memberikan rekomendasi kepada pengguna. Jika pengguna memiliki profil jangka panjang yang menggambarkan karakteristik pribadi mereka, sistem rekomendasi memberikan rekomendasi yang relevan berdasarkan deskripsi, profil, atau informasi lain yang sebanding (Habibi dkk., 2022). Pada sistem rekomendasi pekerjaan ini akan mengolah data (*Curriculum Vitae*) CV yang dikirimkan untuk dapat diolah dan dicocokkan pada data lowongan pekerjaan sehingga sistem dapat memberikan informasi pekerjaan yang relevan kepada pelamar.

Pada penelitian (Koloman dkk., 2023) dapat disimpulkan bahwa memberikan rekomendasi pekerjaan yang sesuai dengan latar belakang dan keterampilan pengguna adalah manfaat besar yang timbul dari adanya sistem rekomendasi pekerjaan bagi para pencari pekerjaan di industri IT. Sistem dapat dengan mudah menemukan peluang pekerjaan yang sesuai dengan keterampilan mereka, sehingga menghemat waktu dan upaya dalam pencarian pekerjaan. Selain itu, sistem ini juga dapat membantu mereka dalam mengidentifikasi area keterampilan yang perlu ditingkatkan untuk meningkatkan prospek karir mereka.

Sistem ini nantinya akan menggunakan metode IndoBERT. IndoBERT merupakan salah satu metode *deep learning* untuk memahami dan menghasilkan teks bahasa Indonesia telah ditunjukkan, karena model bahasa *transformer* ini dilatih pada korpus teks bahasa Indonesia yang luas. IndoBERT ini dapat melakukan berbagai tugas, seperti klasifikasi teks, pembuatan teks, dan menjawab pertanyaan, digunakan untuk menguji kemampuan ini (Saadah dkk., 2022).

Pengimplementasian *web scraping* dalam pengambilan dataset merupakan cara yang efektif dikarenakan pada penelitian (Elfirdaus dkk., 2023) disimpulkan bahwa *web scraping* dapat mempercepat sebuah proses pengumpulan data dalam jumlah yang besar dan waktu yang singkat. *Web scraping* akan memastikan akurasi data yang dikumpulkan untuk mengekstrak data dari halaman web. Pengumpulan data menggunakan *web scraping* ini memungkinkan mengambil

data yang tidak tersedia secara langsung melalui API. Pada tabel 2.1 merupakan tabel hasil rangkuman dari jurnal-jurnal terkait dengan metode IndoBERT dan *Cosine Similarity*.

Tabel 2. 1 Tabel Tinjauan Pustaka

No	Judul	Penulis	Model	Hasil
1	Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom's Taxonomy	(Baharuddin dkk., 2023)	IndoBERT	Hasil penelitian ini menunjukkan bahwa IndoBERT menghasilkan data yang lebih personal dan akurat presentase sebesar 87,3%.
2	Penerapan Algoritma <i>Cosine Similarity</i> pada Sistem Klasifikasi Dokumen Skripsi	(Wahyuni dkk, 2019)	<i>Cosine Similarity</i>	Dalam penelitian ini uji kemiripan dokumen <i>Cosine Similarity</i> memiliki tingkat akurasi dengan presentase sebesar 89% yang artinya memiliki tingkat akurasi yang tinggi.

No	Judul	Penulis	Model	Hasil
3	Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering	(Parwita, 2019)	IndoBERT	Penelitian ini menggunakan dataset e-commerce untuk mendeteksi aspek review e-commerce, dan akurasi model yang menggunakan IndoBERT sebagai word embedding mencapai 94,86%.
4	Learning-Based Matched Representation System for Job Recommendation	(Alsaif dkk., 2022)	Bert dan <i>Cosine Similarity</i>	Dalam penelitian ini menunjukkan bahwa penggunaan BERT dan <i>Cosine Similarity</i> sangat akurat dalam mencocokkan resume dengan deskripsi pekerjaan. Untuk beberapa kategori pekerjaan, akurasi rata-rata dilaporkan sekitar 87%.



No	Judul	Penulis	Model	Hasil
5	The performance of BERT as data representation of text clustering	(Subakti dkk, 2021)	Bert	Pada penelitian ini menunjukkan penggunaan BERT dalam akan meningkatkan akurasi dan relevansi hingga 20% untuk memberikan rekomendasi.

## 2.2 Dasar Teori

### 2.2.1. Sistem Rekomendasi

Sistem rekomendasi merupakan sebuah sistem dengan tujuan untuk memberikan rekomendasi sesuai dengan kebutuhan pengguna. Sistem rekomendasi dapat memprediksi item mana yang akan disukai atau dibutuhkan pengguna dengan menggunakan data perilaku pengguna atau data pengguna. Pada sistem rekomendasi juga memiliki beberapa jenis :

#### a. *Collaborative Filtering*

Salah satu metode dalam sistem rekomendasi adalah *Collaborative Filtering* (CF), yang membuat suatu rekomendasi kepada pengguna berdasarkan pada pola perilaku dan preferensi pengguna lain yang sebanding. *Collaborative filtering* pada dasarnya merupakan metode penyaringan atau pengevaluasian item dengan memanfaatkan pendapat orang lain. Dalam proses ini, metode ini menyaring data berdasarkan tingkah laku karakteristik user sehingga dapat memberikan informasi baru kepada user lain (Muarif dkk., 2022)

#### b. *Content-based Filtering*

Salah satu metode dalam sistem rekomendasi yang disebut *content-based filter* menggunakan informasi atau atribut dari item untuk membuat rekomendasi kepada pengguna. Metode ini didasarkan pada gagasan bahwa pengguna cenderung menyukai item dengan fitur atau karakteristik yang sebanding dengan item yang mereka sukai sebelumnya. Prosesnya mencakup

memeriksa konten atau fitur suatu barang, seperti teks, gambar, atau fitur lainnya, dan menemukan kesamaan antara barang yang disarankan dan barang yang telah disukai oleh pengguna (Nastiti, 2019).

c. *Hybrid Recommendation*

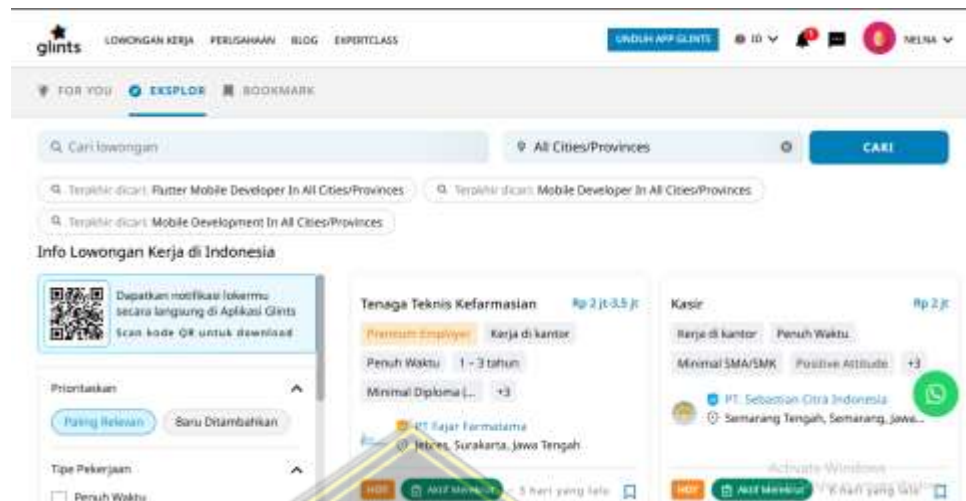
Algoritma ini menggabungkan *collaborative filtering* dan *content-based filtering* untuk menghasilkan rekomendasi yang lebih akurat.

**2.2.2. Curriculum Vitae (CV)**

*Curriculum Vitae* (CV) adalah dokumen yang berisi informasi tentang pengalaman hidup seseorang yang bertujuan untuk memberikan gambaran lengkap tentang diri seseorang. CV terdiri dari data diri, latar belakang pendidikan, kursus-kursus, kemampuan atau pengalaman kerja, pengalaman kerja, dan informasi lainnya yang diperlukan untuk melamar pekerjaan, melanjutkan studi, atau untuk keperluan lain yang penting dalam proses seleksi pekerjaan. CV harus dibuat dengan cara yang tepat dan lengkap. Melalui CV pihak perusahaan dapat mengetahui kemampuan atau skill yang dimiliki pelamar sehingga pihak perusahaan bisa melakukan seleksi.

Sistem rekomendasi pekerjaan menggunakan CV sebagai sumber informasi penting untuk membuat rekomendasi pekerjaan yang lebih akurat, personal, dan kredibel. Mereka dapat membantu pengguna menemukan pekerjaan yang sesuai dengan keahlian, pengalaman, minat, dan tujuan karir mereka. Sehingga penulisan dan pembuatan CV sangat dibutuhkan untuk mempermudah pelamar mendapatkan pekerjaan yang sesuai dengan keahlian.

### 2.2.3. Glints



Gambar 2. 1 Tampilan Website Glints

*Glints* merupakan sebuah website yang didirikan pada tahun 2014 dan telah berkembang menjadi platform terkemuka di Asia Tenggara dengan jutaan pengguna dan ribuan perusahaan, platform ini memiliki sebuah fitur untuk bisa memberikan rekomendasi pekerjaan kepada pengguna *glint* sesuai dengan kriteria yang sesuai dengan keinginan pengguna.

*Glints* ini dapat memberikan rekomendasi tempat kerja atau perusahaan terbaik sesuai dengan rating dan *review* perusahaan. Alasan penulis memilih *glint* untuk menjadikan *platform* untuk pengambilan data menggunakan *web scraping* dikarenakan *platform* *glint* mudah untuk mengakses lowongan pekerjaan tanpa harus melakukan login.

### 2.2.4. Web Scraping

*Web scraping* adalah proses otomatisasi untuk mengekstrak data atau informasi dari halaman web. Proses ini melibatkan penggunaan program komputer atau bot untuk menjelajahi halaman web dan menemukan data tertentu dari elemen HTML, seperti teks, gambar, atau tabel. Data yang diekstrak kemudian dapat disimpan dalam format yang dapat digunakan lebih lanjut, seperti spreadsheet atau database (Flores dkk., 2020).

Analisis pasar, penelitian pesaing, pemantauan harga, dan pembuatan indeks pencarian adalah beberapa contoh penggunaan *web scraping*. Namun perlu diingat bahwa *web scraping* harus dilakukan dengan mematuhi etika dan kebijakan privasi, serta hak cipta dan aturan penggunaan yang berlaku pada situs website yang disasar. Beberapa situs web mungkin memiliki aturan atau prosedur yang mencegah atau membatasi akses ke konten mereka melalui *web scraping*.

### 2.2.5. Text Preprocessing

Salah satu langkah penting dalam sistem layanan rekomendasi pekerjaan, yang menggunakan IndoBERT, adalah preprocessing teks. Tujuan dari tahap ini adalah untuk membersihkan dan memformat teks yang diambil dari website pekerjaan sehingga model IndoBERT dapat memprosesnya dengan baik dan menghasilkan vektor yang akurat. Pemrosesan awal teks sangat penting untuk menghemat waktu dan sumber daya serta meningkatkan kualitas dan akurasi hasil analisis teks.

Dalam penelitian (Muarif dkk., 2022) melakukan text preprocessing akan melalui beberapa tahapan antara lain :

- a. *Cleansing* adalah proses awal untuk membersihkan data sehingga menghasilkan data teks yang memiliki kualitas tinggi sehingga dapat dianalisis dengan lebih akurat.
- b. *Case folding* adalah salah satu tahap dalam preprocessing teks yang bertujuan untuk mengubah semua huruf dalam teks menjadi huruf kecil. Hal ini dilakukan untuk mengurangi kerumitan dalam pemrosesan teks, karena dengan mengabaikan perbedaan besar-kecilnya huruf, model dapat dengan lebih mudah menemukan pola dan makna dalam teks. Misalnya, kata "Halo" dan "halo" akan dianggap sama setelah proses *case folding*.
- c. *Tokenizing* adalah proses pembagian teks menjadi bagian-bagian kata yang disebut "token." Mengambil kata yang dianggap penting dari hasil *tokenizing* atau membuang kata yang dianggap tidak penting dalam proses *preprocessing* teks.

d. *Stopword Removal*

*Stopword* adalah langkah umum dalam tahap *preprocessing* teks untuk berbagai aplikasi NLP. Dengan menghapus *stopwords*, kita dapat meningkatkan efisiensi dan akurasi model NLP serta mempermudah analisis teks. Namun, keputusan untuk menggunakan *stopword removal* harus didasarkan pada tujuan akhir dan jenis analisis yang akan dilakukan.

e. *Stemming*

Proses *preprocessing* teks yang dikenal sebagai *stemming* bertujuan untuk mengubah kata menjadi bentuk dasarnya (*stem*). Dengan *stemming*, tujuan utama adalah untuk mengurangi jumlah kata yang berbeda yang ada dalam dataset dan meningkatkan kemampuan model pembelajaran mesin untuk mengenali kata-kata yang memiliki makna yang sama.

**2.2.6. Transformer**

*Transformers* adalah arsitektur jaringan saraf yang kuat dan serbaguna untuk berbagai tugas pemrosesan bahasa natural. *Transformers* dapat memproses data secara paralel, mempelajari hubungan antar bagian yang jauh, dan membuat interpretasi lebih mudah. *Transformers* tidak menggunakan lapisan *rekursif* atau konvolusional, melainkan mengandalkan mekanisme *attention*.

Dengan menggunakan mekanisme "*attention*", *transformer* berkonsentrasi pada elemen input yang paling terkait dengan pekerjaan yang sedang dilakukan. Mekanisme *self-attention* digunakan oleh transformator untuk mempelajari hubungan antar kata dalam sebuah kalimat. Mekanisme ini memungkinkan model untuk memahami bagaimana kata-kata saling terkait satu sama lain dan bagaimana maknanya berubah tergantung pada konteks kalimat. *Self-Attention Mechanism* dapat dihitung dengan menggunakan rumus (1) :

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Keterangan :

Q = Kueri

K = Kunci

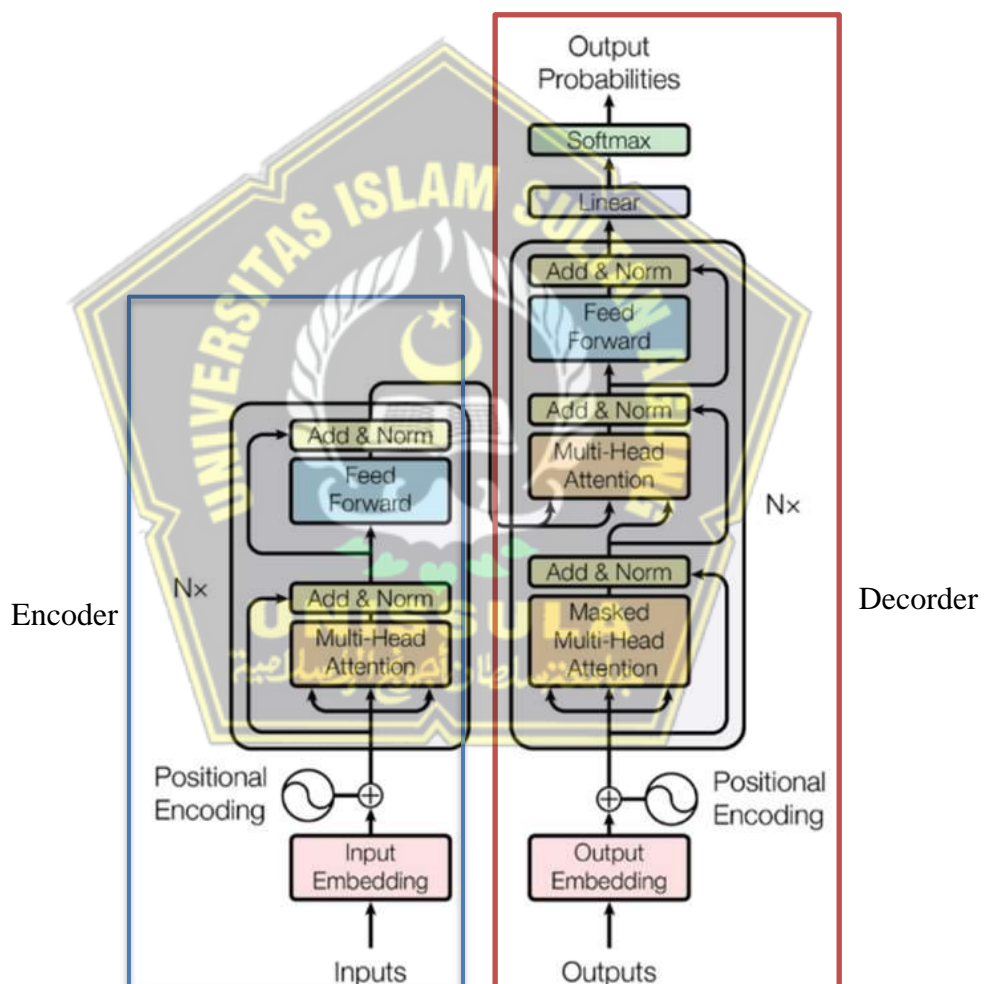


$V$  = Nilai

$d_k$  = Kueri dan Kunci dimensi

$d_v$  = Nilai dari dimensi

Dua bagian utama *transformers* adalah *encoder* dan *decoder*. *Encoder* memproses input dan menghasilkan representasi konteks. *Decoder* kemudian menggunakan representasi ini untuk menghasilkan output (Zeng dkk., 2022). Untuk arsitektur *transformers* dapat dilihat pada gambar 2.1.



Gambar 2. 2 Arsitektur Transformers(Vaswani dkk., 2017)

Arsitektur *transformers* pada gambar 2.2 terdapat 2 bagian utama dari *transformer* yaitu *encoder* dan *decoder*.

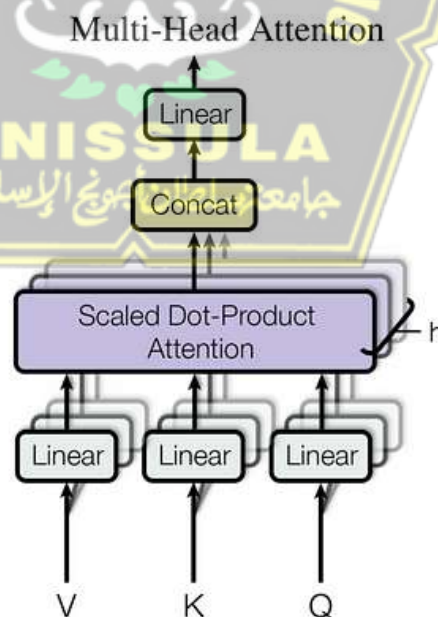
### 2.2.6.1 Encoder dan Decoder

Pada gambar 2.2 terdapat lambang  $N$  yang merupakan simbol dari jumlah lapisan *encoder* dan *decoder*. *Encoder* dan *decoder* memiliki beberapa lapisan identik yang biasanya berjumlah 6 atau 12 lapisan. Gambar 2.1 tersebut pada *encoder* memiliki  $N = 6$  lapisan yang identik, pada setiap lapisan terdapat dua sub-lapisan. Lapisan pertama yaitu mekanisme *multi-head self-attention* dan lapisan kedua yaitu *feed-forward*.

Pada *decoder* terdapat lapisan yang sama yaitu  $N = 6$ , yang memiliki 2 sub-lapisan yang sama pada *encoder* tetapi *decoder* menyisipkan sub-lapisan yang ketiga untuk melakukan *multi-head attention*.

### 2.2.6.2 Multi-head Attention

*Multi-Head Attention* dapat menangkap berbagai aspek informasi dari input sequence. *Multi-Head Attention* memungkinkan model *Transformer* untuk fokus pada berbagai aspek data input secara bersamaan, meningkatkan fleksibilitas dan kemampuan representasinya. Ini memungkinkan model untuk mengidentifikasi hubungan yang kompleks dalam data.



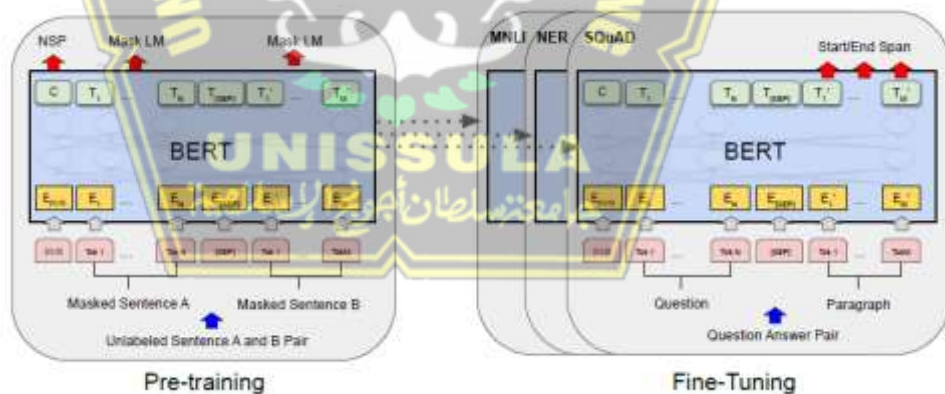
Gambar 2. 3 Multi-head Attention pada *Transformers*(Vaswani dkk., 2017)

Pada gambar 2.3 tersebut merupakan tahapan dari *Multi-Head Attention*. *Multi-Head Attention* menerima tiga input vektor yaitu vektor nilai (*Value*),

vektor kunci (*Key*), dan vektor kuery (*Query*). *Value* (V), *Key* (K), dan *Query* (Q) ialah representasi yang dihasilkan oleh input embedding. Masing-masing dari ketiga representasi ini akan diperoleh dengan mengalikan embedding input dengan matriks berat yang berbeda setelah melalui proyeksi linear. Pada *Scaled Dot-Product Attention* kueri (Q) dan kunci (K) digunakan untuk menghitung *score Attention* melalui *Dot Product*, *Scaling* dan *Softmax* yang akan menghasilkan *output Attention* untuk setiap “*Head*”. *Output* dari seluruh heads akan digabungkan menjadi satu vektor panjang pada tahapan *Concat*. Hasil dari *Concat* akan diproyeksikan kembali melalui lapisan linear akhir untuk bisa mendapatkan representasi akhir dari *Multi-Head Attention* .

### 2.2.7. BERT

BERT (*Bidirectional Encoder Representations from Transformers*) merupakan metode yang dibuat untuk memungkinkan pelatihan dua arah dari teks yang tidak berlabel dengan mengkondisikan bersama di semua lapisan konteks. Model BERT menggunakan *encoder* transformator dua arah *multi-layer*. BERT bekerja dalam dua tahap yaitu *pre-training* dan *fine-tuning* (Saadah dkk., 2022).



Gambar 2. 4 Arsitektur *Pre-Training* dan *Fine-Tuning* pada BERT (Devlin dkk., 2019)

Pada gambar 2.4 merupakan arsitektur dari BERT. Selama tahap *pre-training*, model dilatih dengan menggunakan data yang tidak berlabel dan dilatih dengan berbagai tugas (Baharuddin dkk., 2023). Pada *pre-training* BERT dilatih dengan data tes berjumlah besar yang tidak berlabel dengan dua tugas utama yaitu :

- a. *Masked Language Modeling* (MLM) yang bertujuan untuk memprediksi kata asli yang sudah digantikan berdasarkan pada konteks kata-kata sebelumnya. Dalam proses MLM ini akan menginput sepasang kalimat, kemudian melalui tahap encoding.
- b. *Next Sentence Prediction* (NSP) yang memiliki tugas untuk memprediksi apakah dua kalimat yang diinputkan berurutan atau tidak. Proses tahapan NSP dengan cara dua kalimat diinputkan ke dalam model kemudian di-encode setelah itu model akan mencoba memprediksi apakah kalimat kedua merupakan kalimat selanjutnya dari kalimat pertama.

Pada tahap fine-tuning, model BERT dimulai dengan parameter yang telah dilatih sebelumnya, dan kemudian semua parameter disetel menggunakan data berlabel (Dharmawan dkk., 2022). Pada dasarnya, fine-tuning dilakukan dalam model BERT dengan menambahkan token khusus untuk menentukan fungsi model yang dimaksudkan. Setelah *pre-training* BERT akan diadaptasi untuk berbagai tugas yang spesifik seperti klasifikasi teks, *question answering*, dan lainnya dengan memberi pelatihan tambahan pada dataset.

### 2.2.8. IndoBERT

Pada sistem ini akan menggunakan metode indoBERT. IndoBERT dan BERT memiliki arsitektur yang sama. Pada indoBERT terdapat Indo4B yang dapat digunakan untuk mengubah teks menjadi vektor-vektor. Penggunaan Embedding yang lebih canggih seperti indo4B dapat meningkatkan tingkat akurasi (Koto dkk., 2020).

Setelah *preprocessing*, IndoBERT digunakan untuk mengolah teks dan menghasilkan vektor representasi kontekstual dari teks tersebut dengan tahapan antara lain :

1. Membuat Algoritma Mesin Pembelajaran : Menghitung kesamaan antara profil pelamar dan lowongan pekerjaan dengan menggunakan algoritma *Cosine Similarity*.
2. Melakukan Pelatihan: Melakukan pelatihan algoritma pembelajaran mesin dengan menggunakan data yang telah diproses untuk memperoleh model yang akurat.

3. Penggunaan IndoBERT: Pengolahan teks dan pembuatan representasi kontekstual dapat dilakukan dengan menggunakan IndoBERT.

Dengan memanfaatkan IndoBERT, sistem dapat memahami dan menganalisis profil pencari kerja serta deskripsi pekerjaan, dan menemukan kesamaan antara keduanya dengan menggunakan *cosine similarity*. Ini memungkinkan sistem untuk menyaring dan menyarankan pekerjaan yang paling relevan berdasarkan profil pencari kerja.

### 2.2.9. Cosine Similarity

*Cosine Similarity* adalah metrik yang digunakan untuk mengukur kemiripan dua vektor dalam arah atau orientasi, mengabaikan perbedaan dalam skala atau besaran. Kedua vektor harus menghasilkan skalar dengan perkalian ruang *inner product* karena mereka harus menjadi bagian dari ruang *inner product* yang sama (Ariantini dkk., 2016). Untuk mengetahui seberapa mirip dua vektor, sinus cosinus dari sudut antara mereka digunakan.

$$\cos a = \frac{A \cdot B}{|A| \cdot |B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (2)$$

Keterangan :

A = Vektor A, yang akan dibandingkan kemiripan

B = Vektor B, yang akan dibandingkan kemiripan

A.B = Dot product antara vektor A dan vektor B

|A| = Panjang vektor A

|B| = Panjang vektor B

|A|. |B| = Cross product antara |A| dan |B|

### 2.2.10. Evaluasi Model

Evaluasi pada sebuah model setelah dilakukan pemodelan merupakan cara untuk mengukur performa sebuah model machine learning pada data tertentu. Evaluasi dilakukan agar mengetahui seberapa baik model dalam melakukan tugasnya.



*Average Precision @K* merupakan sebuah metrik evaluasi yang penting dalam sistem rekomendasi. *Average Precision @K* ini akan digunakan untuk menilai kualitas rekomendasi yang akan diberikan, tidak hanya tingkat akurasi tetapi juga juga mengetahui seberapa efektif rekomendasi yang diberikan (Afifaturahman dkk., 2021). Berikut merupakan rumus dari *Average Precision @K* :

$$AP@K = \frac{1}{|R|} \sum_{k=1}^K P(k) \cdot rel(k) \quad (3)$$

Keterangan :

$|R|$  = jumlah total item relevan

$P(k)$  = precision pada posisi  $k$

$Rel(k)$  = indikator apakah item pada posisi  $k$  relevan atau tidak (1 jika relevan, 0 jika tidak relevan )

Untuk menghitung jumlah dari presicion dan juga recall maka dapat menggunakan rumus :

$$\text{Precision} = \frac{\text{Jumlah dokumen relevan yang ditemukan}}{\text{Jumlah total dokumen yang ditemukan}} \quad (4)$$

$$\text{Recall} = \frac{\text{Jumlah dokumen relevan yang ditemukan}}{\text{Jumlah total dokumen relevan dalam dataset}} \quad (5)$$

Ground Truth merupakan sebuah standart evaluasi permodelan machine learning. Ground Truth ini bisa di dapatkan dari proses manual dari pelabelan data. Ground Truth ini memiliki peran penting yang akan menilai seberapa efektif model ini untuk sistem rekomendasi pekerjaan sehingga dapat menghasilkan rekomendasi yang relevan.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Metode Penelitian

Penelitian sistem layanan rekomendasi pekerjaan ini menggunakan algoritma IndoBERT. Dalam penelitian ini IndoBERT berfungsi untuk melakukan tokenisasi kalimat, perubahan token menjadi representasi numerik (*embedding*), pemrosesan dengan arsitektur *Transformer* sehingga dapat menghasilkan sistem rekomendasi lowongan pekerjaan yang mencakup *web development*, *mobile development* dan *data analyst*.

##### 3.1.1. Studi Literatur

Penulis meninjau dan mempelajari teori mengenai Indobert, *Preprocessing* dan juga *web scraping* melalui berbagai sumber seperti website, jurnal, e-book, artikel yang ada di Internet.

##### 3.1.2. Pengumpulan data

Langkah awal yang akan dilakukan yaitu pengumpulan data. Data yang dikumpulkan yaitu berupa data lowongan pekerjaan dan juga CV user. Pengambilan data lowongan pekerjaan tersebut didapatkan dengan menggunakan teknik *web scraping* pada *website* glint. Data lowongan pekerjaan yang diambil hanya yang terkategori sebagai pekerjaan *Computer Science* dan hanya yang terkategori *web development*, *mobile development* dan *data analyst*. Dari setiap kategori yang diambil hanya berjumlah 10 dengan rincian nama pekerjaan, keahlian, nama perusahaan, dan juga gaji.

Untuk pengambilan data kedua yaitu data dari CV user. CV akan di upload dengan format PDF dengan format tidak dalam bentuk gambar. Pada CV tersebut setidaknya terdapat data deskripsi diri, skill, dan riwayat karir. Pada CV tersebut akan dilakukan *preprocessing* untuk dapat dicocokkan dengan data lowongan pekerjaan hasil *web scraping* dari website glints.

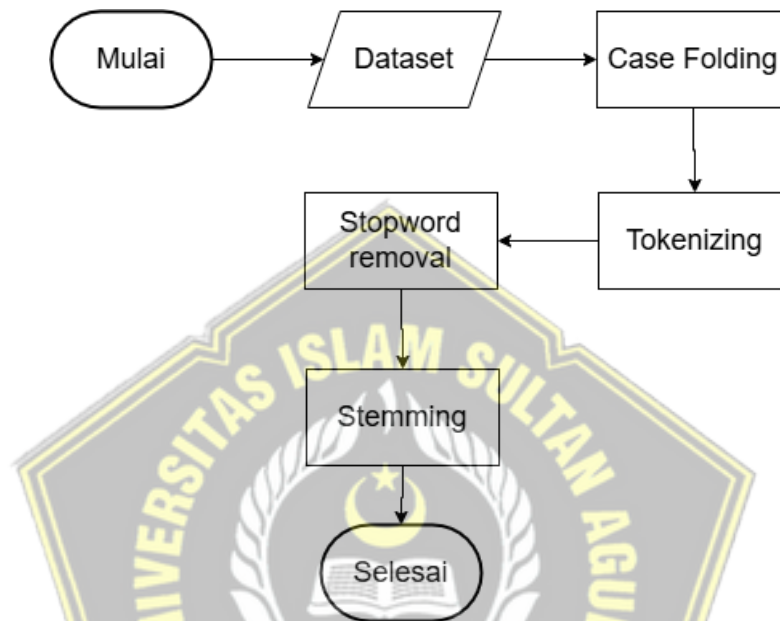
2

---

<sup>2</sup> <https://glints.com/id>

### 3.1.3. Alur Perancangan Sistem

Alur peneliti dimulai dengan mengumpulkan data lowongan pekerjaan menggunakan *web scraping* pada website glints dan mengumpulkan data form dan CV melalui hasil inputan user pada website sistem layanan. Setelah data didapatkan maka akan melalui tahap *preprocessing* seperti pada gambar 3.1.



Gambar 3. 1 Preprocessing

Pada tahapan ini akan melalui beberapa proses yaitu

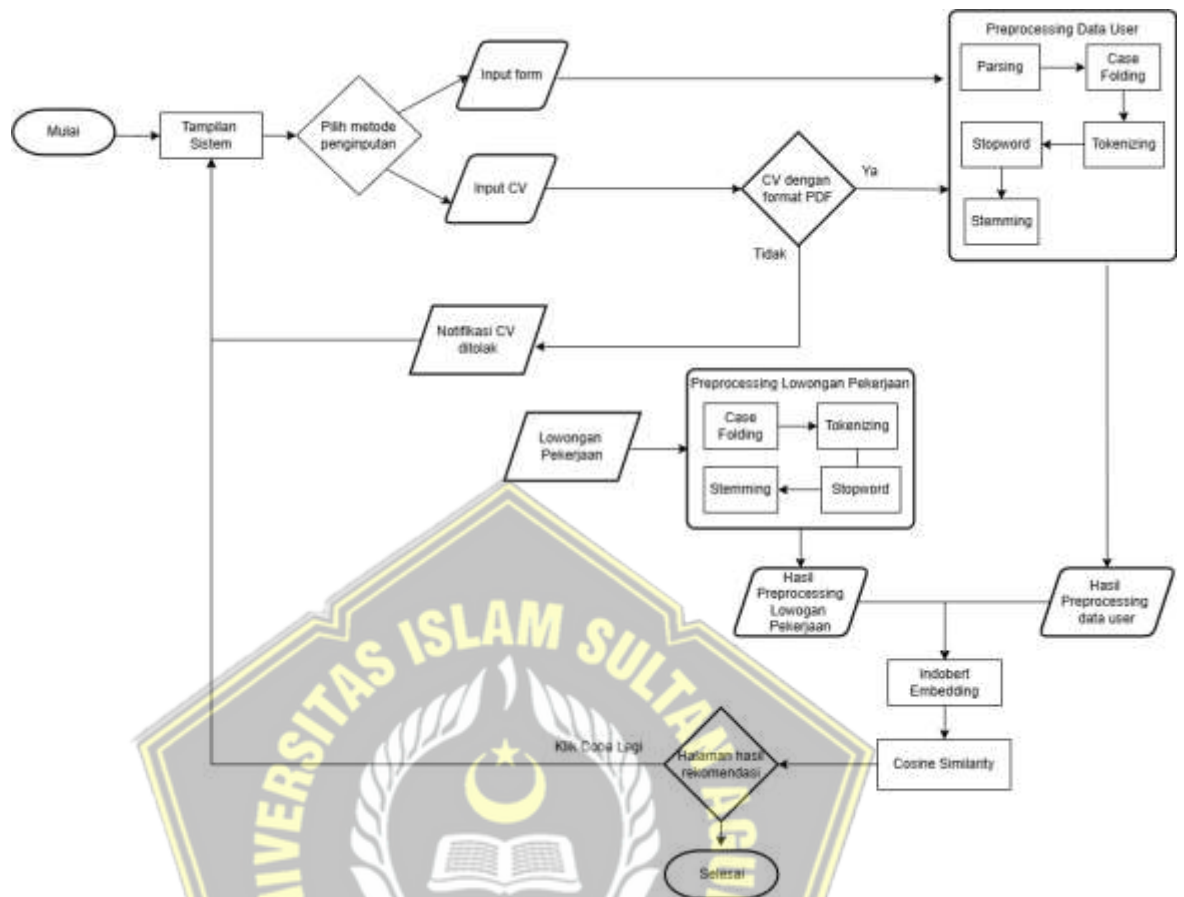
1. *Case Folding* : Tahapan ini merupakan proses untuk mengubah semua huruf kapital menjadi huruf kecil untuk menyamakan representasi kata. Contohnya seperti “Lowongan Pekerjaan” diproses pada tahap *case folding* menjadi “lowongan pekerjaan”.
2. *Tokenizing* : Tahap ini merupakan proses untuk memecah kalimat menjadi per kata sehingga lebih mudah untuk diolah oleh IndoBERT. Contoh dari proses tokenizing yaitu seperti “ Saya ingin mencari pekerjaan sebagai data analyst” menjadi [“saya”, “ingin”, “mencari”, “pekerjaan”, “sebagai”, “data”, “analyst”]
3. *Stopword Removal* : Pada tahap keempat ini yaitu tahapan *stopword removal* untuk menghapus kata-kata yang umum dan tidak mempunyai makna informasi. Contohnya “saya”, “ini”, “yang”, “dan”.

4. *Stemming* : tahapan ini mengubah kata menjadi bentuk dasar untuk mengurangi kata unik dalam data. Contohnya seperti “mencari”, “mencarikan”, “Pencarian” menjadi “cari”.

Setelah melalui tahapan preprocessing maka data CV dan data lowongan pekerjaan akan masuk pada tahapan IndoBERT yang akan mengubah dataset menjadi sebuah representasi vektor numerik yang disebut *embedding*. IndoBERT akan mengekstrak informasi penting dari data CV dan data lowongan pekerjaan seperti keahlian, pengalaman, pendidikan terakhir dan informasi terkait yang relevan. Pada IndoBERT terdapat algoritma *cosine similarity* untuk menghitung tingkat kesamaan antara embedding data CV dan lowongan pekerjaan. Proses penting dalam sistem rekomendasi pekerjaan adalah peninjauan. Ini dilakukan untuk memastikan bahwa model menghasilkan rekomendasi yang berkualitas dan bermanfaat bagi pengguna. Dengan melakukan peninjauan secara berkala dan melakukan perbaikan berdasarkan hasil peninjauan, sistem rekomendasi pekerjaan dapat terus berkembang dan meningkatkan kinerjanya dalam membantu pencari kerja menemukan pekerjaan yang sesuai dengan pengalaman dan kualifikasi mereka.

#### 3.1.4. Gambaran Sistem

Pengembangan website yang penulis lakukan ini bertujuan untuk dapat merekomendasikan sebuah lowongan pekerjaan yang cocok dengan data form dan CV yang akan diupload oleh user.



Gambar 3. 2 Flowchart alur sistem

Pada gambar 3.2 merupakan alur yang akan diimplementasikan pada website ini yaitu user akan mengirimkan sebuah data form dan dokumen CV dengan format pdf. Data dan dokumen pdf itu akan melalui proses parsing untuk nantinya bisa dilakukan tahapan *preprocessing* yaitu *cleansing*, *case folding*, *tokenizing*, *stopwords* dan *stemming*. Tahapan *preprocessing* dilakukan agar data user memiliki tingkat kualitas yang tinggi. Pada data lowongan pekerjaan juga akan melalui tahap *preprocessing* untuk menghasilkan data yang berkualitas tinggi. Setelah kedua dataset tersebut melalui proses *preprocessing* maka akan dilakukan proses *embedding* dengan menggunakan metode IndoBERT supaya dapat dilakukan proses uji kemiripan menggunakan metode *Cosine Similarity*. Ketika uji kemiripan dilakukan dan hasilnya memiliki kecocokan antara data user dan data lowongan pekerjaan maka sistem *website* akan menampilkan lowongan pekerjaan yang relevan kepada user. Kemudian pada tabel hasil rekomendasi



terdapat *button* aksi sehingga user dapat mengetahui secara detail mengenai informasi lowongan pekerjaan yang telah direkomendasikan dan juga terdapat tombol coba lagi yang jika diklik oleh user akan kembali ke halaman awal pengisian form dan upload CV.

### 3.1.5. Perancangan *User Interface*

Dalam pengembangan website yang penulis lakukan pastinya penulis membuat sebuah rancangan yang akan diimplementasikan pada penelitian ini, berikut merupakan design website dari penelitian ini :

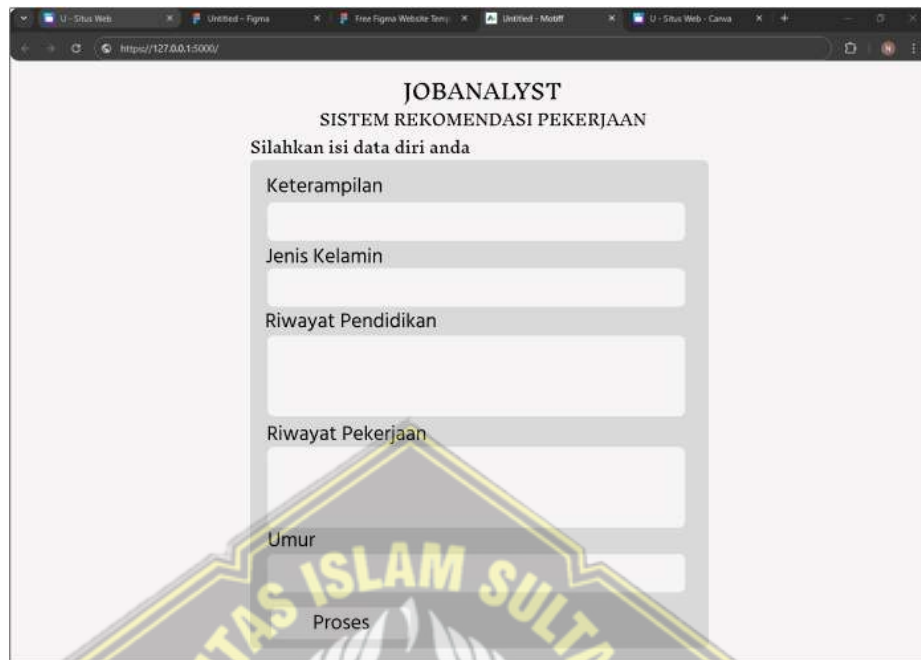
#### 1. Beranda



Gambar 3. 3 Gambaran beranda pada sistem

Pada gambar 3.3 merupakan gambaran awal sistem yang akan ditampilkan kepada user. Pada halaman tersebut terdapat pilihan untuk user dalam pengisian data diri, pada halaman tersebut terdapat pilihan apakah user ingin melakukan input data diri menggunakan form atau ingin melakukan penginputan CV.

## 2. Halaman pengisian data diri menggunakan form



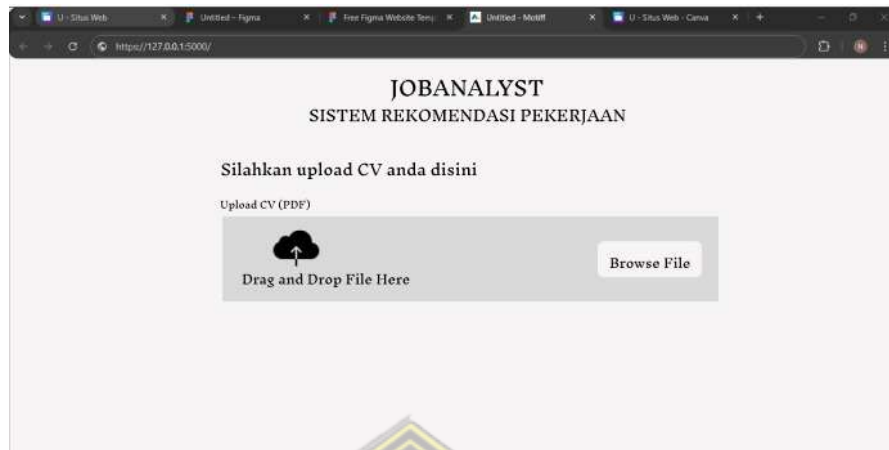
The screenshot shows a web browser window with the URL <https://127.0.0.1:5000/>. The page title is "JOBANALYST" and the subtitle is "SISTEM REKOMENDASI PEKERJAAN". Below the title, it says "Silahkan isi data diri anda". The form contains the following fields and a button:

- Keterampilan
- Jenis Kelamin
- Riwayat Pendidikan
- Riwayat Pekerjaan
- Umur
- Proses

Gambar 3. 4 Interface pengisian form pelamar

Pada gambar 3.4 merupakan tampilan dari interface jika user memilih melakukan pengisian form. Pada halaman tersebut pengisian form yang harus diisi yaitu berupa kemampuan diri, riwayat pekerjaan, riwayat pendidikan, dan umur dari pelamar kerja yang kemudian juga terdapat tombol proses untuk memproses data diri pelamar.

### 3. Halaman pilihan upload CV



Gambar 3. 5 Interface upload CV

Pada gambar 3.5 merupakan tampilan dari interface yang dipilih user untuk melakukan upload CV. Dengan format dokumen yang pelamar upload harus berupa PDF dan tidak dalam format gambar. Setelah CV terupload maka sistem akan memproses CV dan kemudian akan menampilkan halaman hasil rekomendasi seperti pada gambar 3.5.

### 4. Halaman hasil rekomendasi



Gambar 3. 6 Halaman hasil rekomendasi sistem

Pada gambar 3.6 merupakan gambaran dari sistem yang berhasil merekomendasikan sebuah lowongan pekerjaan kepada user yang sudah

melakukan pengisian form dan upload CV pada halaman awal. Pada halaman tersebut akan menampilkan data mengenai lowongan pekerjaan dalam bentuk tabel seperti nama pekerjaan, nama perusahaan, alamat, gaji dan juga tersedia kolom detail untuk menampilkan data lowongan pekerjaan secara detail. Pada tabel tersebut menampilkan 5 data lowongan pekerjaan yang paling relevan dengan data yang sudah user inputkan pada halaman awal.

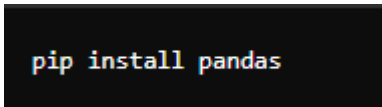
### 3.1.6. Kebutuhan Perangkat Lunak

Untuk menunjang penelitian ini penulis menganalisis kebutuhan perangkat lunak yang dibutuhkan untuk dapat menghasilkan sebuah sistem layanan yang memiliki proses input dan menghasilkan output yang sesuai dengan penulis harapkan yaitu merekomendasikan lowongan pekerjaan yang relevan dengan user. Adapun kebutuhan perangkat lunak tersebut seperti :

#### 1. Visual Studio Code

Visual Studio code merupakan kode editor yang dikembangkan oleh *Microsoft*. Visual studio code digunakan untuk berbagai pengembangan sebuah perangkat lunak seperti pengembangan sebuah web development, pengembangan aplikasi game, pengembangan data *science* dan beberapa pengembangan yang lainnya. Visual studio code ini memiliki performa yang cukup tinggi sehingga dalam pengembangan *web development* akan sangat cepat dan responsif, fitur fitur pada visual studio code juga terus dikembangkan oleh *Microsoft* sehingga terdapat peningkatan performa secara berkala.

#### 2. *Library Pandas*



```
pip install pandas
```

Gambar 3. 7 Instalasi Library Pandas

*Pandas* merupakan salah satu *library* Python yang memiliki fungsi untuk dapat membaca data dari berbagai format file seperti, CSV, Excel dan lainnya. Pada gambar 3. 7 merupakan cara untuk menginstalasi library *pandas*, data yang diproses akan kembali ke dalam format file yang sama. *Pandas* ini juga dapat menganalisis data untuk menemukan pola, anomali atau informasi penting lainnya.

### 3. *Library Fitz*

```
!pip install PyMuPDF==1.23.1
```

Gambar 3. 8 Instalasi *library Fitz*

*Library Fitz* merupakan *library* yang digunakan untuk mengotomatiskan tugas-tugas yang berkaitan dengan dokumen PDF untuk menganalisis data lebih dalam. Pada gambar 3.8 merupakan source code untuk menginstalasi *library fitz* pada program. *Library fitz* ini digunakan ketika ingin mengekstrak teks, gambar atau metadata dari sebuah dokumen dengan format PDF. *Library* ini memiliki performa yang dikenal cepat dalam memproses sebuah dokumen besar.

### 4. *Library Re (Regular Expression)*

*Library Re* ini merupakan *library* yang digunakan untuk dapat mendefinisikan pola pencarian. *Library* ini digunakan dalam penelitian ini untuk *preprocessing cleaning* yang berguna untuk menghapus angka pada dataset. *Library Re* dapat mengubah dataset agar memiliki kualitas yang baik

### 5. *Library String*

*Library string* dalam python ini digunakan untuk mendefinisikan sebuah pola pencarian seperti mencari pola yang terdiri dari sebuah kombinasi karakter, angka, dan karakter khusus. *Library string* digunakan pada proses *case folding* untuk mengubah huruf kapital menjadi huruf kecil.

### 6. *Library NLTK (Natural Language Tool Kit)*

*Library NLTK* merupakan *library* yang memiliki kemampuan untuk membangun program yang dapat bekerja dengan data bahasa manusia. NLTK akan melakukan berbagai tugas pemrosesan menggunakan bahasa alami seperti pada penelitian ini. Pada penelitian ini penulis menggunakan *library NLTK* untuk melakukan *preprocessing* seperti, *tokenizing* dan *stemming*.

### 7. *Library Sastrawi*

```
!pip install Sastrawi==1.0.1
```

Gambar 3. 9 Instalasi *Library Sastrawi*



*Library Sastrawi* ini merupakan library yang berfokus pada pengolahan bahasa indonesia yang pada saat preprocessing data stopwords remover library ini digunakan untuk menghapus kata kata umum dan yang tidak memiliki arti penting sedangkan pada stemmer digunakan untuk mereduksi kata ke bentuk awal. Pada gambar 3.9 merupakan kode dalam penginstalasian *library sastrawi*.

#### 8. *Library Torch*

*Library torch* merupakan library yang digunakan untuk memudahkan dalam pembelajaran *machine learning*. Pada penelitian ini *library torch* digunakan untuk membangun arsitektur dari model IndoBERT untuk mengklasifikasikan suatu teks.

#### 9. *Library Numpy*

*Library numpy* merupakan *library* yang digunakan untuk melakukan operasi perhitungan dan analisis data dengan lebih efisien. Pada analisis data *library numpy* dapat membaca dan membersihkan data numerik. *Library numpy* ini juga menyediakan banyak fungsi matematika untuk mengoperasikan *array*.

#### 10. *Framework Streamlit*

*Streamlit* merupakan sebuah *framework* yang sederhana dan cukup fleksibel dalam membangun *interface* yang digunakan untuk model *machine learning*. *Framework Streamlit* merupakan *framework Python*, *Streamlit* ini memiliki komponen *UI* yang mudah untuk digunakan karena *streamlit* menyediakan komponen yang siap pakai. *Streamlit* dalam mengintegrasikan *library python* sangat mudah dilakukan.

## BAB IV

### HASIL DAN ANALISIS PENELITIAN

#### 4.1 Hasil Perancangan Model

##### 4.1.1. Deskripsi Dataset


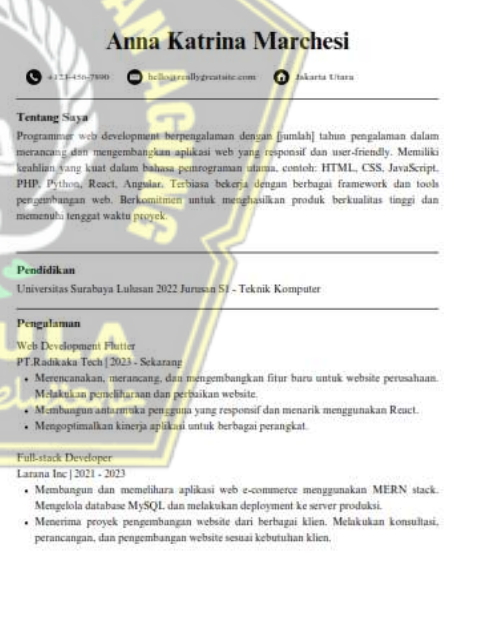
Pada penelitian ini dataset yang digunakan yaitu dataset lowongan pekerjaan dan juga dataset pelamar. Dataset lowongan pekerjaan di dapat dari hasil *web scraping* pada *website glints*. Dataset yang di *scraping* berupa nama pekerjaan, nama perusahaan, lokasi perusahaan, gaji, deskripsi pekerjaan, kualifikasi pekerjaan, jenis kelamin dan umur.

Index	Job Title	Company	Location	Fee	Job Description	Requirement	JenisKelamin	Umur
0	Mobile Manager	Dynami Systems	Hersung	47000000	Membuat API yang efisien dan scalable. Mengorganisir tim/teknolog seperti Java dan Spring Boot.	Kemampuan menulis kode yang baik	Pria/Wanita	35-38
1	Web Analyst	Visiary Apps	Makassar	20000000	Membuat prototype baru baru untuk aplikasi. Mengorganisir tim/teknolog seperti Go/PHP dan Kubernetes untuk Cloud.	Diutamakan dalam pengembangan aplikasi mobile dan web	Pria	35-38
2	Backend Architect	Indovision Apps	Yogyakarta	50000000	Mengorganisir tim/teknolog seperti Go/PHP dan Kubernetes untuk Cloud.	Mahir dalam manajemen database dan cloud computing	Pria/Wanita	30-35
3	Data Consultant	Tech Hub	Mekkah	100000000	Membuat API yang efisien dan scalable. Mengorganisir tim/teknolog seperti Java dan Kubernetes untuk Cloud.	Kemampuan menulis kode yang baik	Pria	35-40
4	Frontend Engineer	Digital Hub	Mekkah	10000000	Membuat prototype baru baru untuk aplikasi. Mengorganisir tim/teknolog seperti Python/Java dan AWS/Azure/Google Cloud.	Fitur dengan mobile app development	Pria	30-35
5	Data Architect	Tech	Bali	30000000	Membuat API yang efisien dan scalable. Mengorganisir tim/teknolog seperti Java dan Kubernetes untuk Cloud.	Membuat prototipe yang sesuai dengan UI/UX	Pria	35-38
6	System Developer	Digital Apps	Bali	30000000	Membuat API yang efisien dan scalable. Mengorganisir tim/teknolog seperti SQL dan NoSQL untuk manajemen database.	Detail System Req, Komputer atau Teknik elektronika	Pria/Wanita	30-35
7	Cloud Engineer	Dynamic Systems	Beraring	30000000	Mengorganisir tim/teknolog seperti Java dan Kubernetes untuk Cloud.	Minimal 2 tahun pengalaman di bidang cloud	Pria/Wanita	35-38
8	Frontend Administrator	Indovision	Beraring	30000000	Mengorganisir tim/teknolog seperti SQL dan NoSQL untuk manajemen database.	Minimal 2 tahun pengalaman di bidang pengembangan & testing	Pria/Wanita	35-40

Gambar 4. 1 Dataset hasil web scraping lowongan pekerjaan

Pada gambar 4.1 Dataset yang di *scraping* hanya yang terkategori sebagai *web development*, *mobile development* dan juga *data analyst*. Jumlah keseluruhan data yang di *web scraping* berjumlah 150 data.

Tabel 4. 1 Tabel CV Pelamar

Format CV	Isi
CV dengan format PDF dengan jumlah 1 halaman.	 <p><b>ALVINO RACHISE</b> WEB DEVELOPER</p> <p><b>BINGKASAN</b> Seorang web developer dengan pengalaman web lebih dari 2 tahun kerja. Berpengalaman dalam mengembangkan website untuk brand dari berbagai industri, mulai dari media hingga e-commerce. Terbukti sukses dalam mengembangkan website user friendly yang berhasil meningkatkan click sebesar 25% dan pembelian hingga 13%.</p> <p><b>PENGALAMAN</b> <b>E-COMMERCE JAYA – Jakarta, Indonesia</b> Web Developer (Juni 2020 – saat ini)</p> <ul style="list-style-type: none"> <li>Membuat antarmuka yang responsif dan fungsional</li> <li>Merancang dan mengembangkan website yang ramah pengguna dan meningkatkan peningkatan klik hingga 25% dan pembelian 13%</li> <li>Membuat mock design dan wireframes untuk landing page 4 konsep produk</li> </ul> <p><b>PENDIDIKAN</b> <b>UNIVERSITAS BANGSA (2015-2019)</b> SI Teknik Informatika – IPK 3,7</p> <p><b>WEB DEVELOPER ACADEMY (Mei – Agustus 2018)</b> Bootcamp web developer</p> <p><b>KEMAMPUAN</b> <b>Hard Skill:</b> WordPress, wireframe, debugging, Photoshop</p> <p><b>Soft Skill:</b> Komunikatif, kritis, team work, teliti</p> <p><b>Bahasa Pemrograman:</b> JavaScript, HTML, CSS, C++, Python</p>
CV dengan format PDF dengan jumlah 2 halaman.	 <p><b>Anna Katrina Marchesi</b></p> <p>☎ +62 812 456 7890    ✉ hellogratia@gmail.com    📍 Jakarta Utara</p> <p><b>Tentang Saya</b> Programmer web development berpengalaman dengan [jumlah] tahun pengalaman dalam merancang dan mengembangkan aplikasi web yang responsif dan user-friendly. Memiliki keahlian yang kuat dalam bahasa pemrograman utama, contoh: HTML, CSS, JavaScript, PHP, Python, React, Angular. Terbiasa bekerja dengan berbagai framework dan tools pengembangan web. Berkomitmen untuk menghasilkan produk berkualitas tinggi dan memenuhi tenggat waktu proyek.</p> <p><b>Pendidikan</b> Universitas Surabaya Lulusan 2022 Jurusan SI - Teknik Komputer</p> <p><b>Pengalaman</b> Web Development Flutter PT.Radikata Tech   2023 - Sekarang</p> <ul style="list-style-type: none"> <li>Merencanakan, merancang, dan mengembangkan fitur baru untuk website perusahaan.</li> <li>Melakukan pemeliharaan dan perbaikan website.</li> <li>Membangun antarmuka pengguna yang responsif dan menarik menggunakan React.</li> <li>Mengoptimalkan kinerja aplikasi untuk berbagai perangkat.</li> </ul> <p>Full-stack Developer Larana Inc   2021 - 2023</p> <ul style="list-style-type: none"> <li>Membangun dan memelihara aplikasi web e-commerce menggunakan MERN stack.</li> <li>Mengelola database MySQL dan melakukan deployment ke server produksi.</li> <li>Menerima proyek pengembangan website dari berbagai klien. Melakukan konsultasi, perancangan, dan pengembangan website sesuai kebutuhan klien.</li> </ul>

	<p><b>Proyek Yang Pernah Dikerjakan</b></p> <ul style="list-style-type: none"> <li>• Pengembangan Website Toko Online (React, Node.js, MongoDB): Membangun toko online lengkap dengan fitur keranjang belanja, pembayaran, dan manajemen produk. Berhasil meningkatkan konversi penjualan sebesar 15% melalui optimasi UX/UI.</li> <li>• Aplikasi Mobile Hybrid: Aplikasi mobile yang dibangun menggunakan teknologi web (React Native, Flutter) dan dapat berjalan di berbagai platform (iOS, Android) dengan menggunakan HTML, CSS, JavaScript, React Native, Flutter, Backend (Node.js) sehingga tercapainya pengalaman pengguna yang native, kinerja yang baik, dan distribusi melalui app store.</li> <li>• Aplikasi web sederhana untuk membuat dan mengelola daftar tugas menggunakan HTML, CSS, JavaScript, Local Storage hingga tercapainya Fitur tambah, hapus, edit tugas, dan tampilan yang user-friendly.</li> </ul> <p><b>Kursus - Kursus</b></p> <ul style="list-style-type: none"> <li>• Data Analyst - 2022</li> <li>• Mobile Programming - 2021</li> <li>• Web Programming - 2022</li> <li>• UI/UX - 2021</li> </ul> <p><b>Skill</b></p> <ul style="list-style-type: none"> <li>• Bahasa Pemrograman: JavaScript (ES6+), TypeScript</li> <li>• Framework: React, Redux</li> <li>• Tools: Webpack, Babel, npm, Git</li> <li>• Soft Skills: Komunikasi, kolaborasi, problem-solving, perhatian terhadap detail</li> </ul>
--	---

Pada tabel 4.1 merupakan contoh CV yang digunakan untuk data penelitian, CV tersebut memiliki rincian yang berbeda beda yang akan dijadikan bahan perbandingan.

ID	Kategori	Skill	Riwayat Pekerjaan	Riwayat Pendidikan	Umur	Jenis Kelamin	
0	B01	Web Development	HTML, CSS, JavaScript, React	Frontend Developer di PT. Teknologi	S1 Teknik Informatika	25	Pria
1	B02	Web Development	PHP, Laravel, MySQL	Backend Developer di PT. Digital	S1 Ilmu Komputer	30	Pria
2	B03	Web Development	HTML, Bootstrap, Node.js	Web Developer di PT. Kreatif	S1 Sistem Informasi	27	Pria
3	B04	Web Development	JavaScript, Vue.js, MongoDB	Fullstack Developer di Startup ABC	S1 Teknik Komputer	25	Wanita
4	B05	Web Development	Angular, TypeScript, SQL	Frontend Developer di PT. Solusi Digital	S1 Teknologi Informasi	29	Pria
5	B06	Web Development	WordPress, PHP, MySQL	Web Administrator di PT. Digital Nusantara	B3 Manajemen Informatika	27	Wanita
6	B07	Web Development	HTML, CSS, JavaScript	Web Designer di Studio Kreatif	S1 Desain Komunikasi Visual	23	Pria
7	B08	Web Development	Django, Python, PostgreSQL	Web Developer di PT. Teknologi Masa Depan	S1 Teknik Informatika	31	Pria
8	B09	Web Development	Ruby on Rails, SQLite, JavaScript	Backend Developer di PT. Startup Kita	S1 Sistem Informasi	26	Pria
9	B10	Web Development	PHP, CodeIgniter, jQuery	Web Developer di CV. Solusi Web	D3 Teknik Informatika	28	Wanita
10	B11	Web Development	JavaScript, React, Redux	Frontend Developer di PT. Digital Solutions	S1 Teknik Informatika	25	Wanita
11	B12	Web Development	Node.js, Express.js, MongoDB	Full-stack Developer di Startup Tech	S1 Teknik Komputer	25	Pria
12	B13	Web Development	Vue.js, Nuxt.js, Firebase	Frontend Developer di PA. Kreatif	S1 Sistem Informasi	25	Wanita
13	B14	Web Development	HTML, CSS, JavaScript	Web Designer di Agency Digital	S1 Desain Komunikasi Visual	29	Pria
14	B15	Web Development	Python, Flask, PostgreSQL	Backend Developer di PPT. Aplikasi Modern	S1 Teknik Informatika	28	Pria
15	B16	Web Development	Ruby, Sinatra, SQLite	Backend Developer di Startup Zeta	S1 Teknologi Informasi	29	Pria
16	B17	Web Development	PHP, Symfony, MySQL	Backend Developer di PT. Inovasi Digital	S1 Ilmu Komputer	28	Wanita
17	B18	Web Development	JavaScript, Angular, SQL	Frontend Developer di PT. Teknologi Baru	S1 Sistem Informasi	28	Pria
18	B19	Web Development	HTML, CSS, WordPress	Web Designer di Freelance	D3 Teknik Informatika	31	Pria
19	B20	Web Development	Laravel, Vue.js, MySQL	Backend Developer di PT. Solusi Tech	S1 Teknik Informatika	29	Wanita
20	B21	Mobile Development	Kotlin, Android Studio, SQLite	Android Developer di PT. Aplikasi Kita	S1 Teknik Informatika	30	Wanita

Gambar 4. 2 Tampilan data kandidat

Pada gambar 4.2 merupakan tampilan data dari kandidat pencari kerja yang sudah dilabel manual dengan ID B01,B02,B03....B50. Data kandidat tersebut berisikan kolom ID, kemampuan, riwayat pendidikan, riwayat pekerjaan, umur dan jenis kelamin.

## 4.1.2. Data Preprocessing

### 4.1.1.1. Case Folding

Pada proses *case folding* ini merupakan proses *preprocessing* yang memiliki tujuan untuk mengubah semua huruf kapital menjadi huruf kecil. Tujuan dilakukannya proses *case folding* yaitu agar mempercepat pemrosesan dan dengan tujuan agar data memiliki kualitas yang cukup baik.

```
Kolom: JobDeskripsi
Sebelum Case Folding:
0 Kami mencari Software Engineer yang berpengalaman...
1 Sebagai Data Analyst yang dapat menentukan bes...
2 Kami mencari Mobile App Developer yang berpeng...
3 Mengembangkan dan memelihara web responsif men...
4 Kami mencari web developer untuk pengembangan ...
```

Gambar 4. 3 Data sebelum dilakukan *case folding*

Pada gambar 4.3 merupakan data sebelum dilakukannya proses *case folding*, dapat dilihat data masih terdapat huruf *uppercase* dan *lowercase*. Untuk hasil dari proses *case folding* dapat dilihat pada gambar 4.2

```
Sesudah Case Folding:
0 kami mencari software engineer yang berpengalaman...
1 sebagai data analyst yang dapat menentukan bes...
2 kami mencari mobile app developer yang berpeng...
3 mengembangkan dan memelihara web responsif men...
4 kami mencari web developer untuk pengembangan ...
```

Gambar 4. 4 Data setelah dilakukan *case folding*

Pada gambar 4.4 merupakan hasil setelah dilakukan proses *case folding* sehingga terlihat bahwa data hanya terdapat huruf *lowercase*. Dapat diartikan proses *case folding* berhasil dilakukan dan dilanjutkan pada proses *preprocessing* selanjutnya yaitu *tokenizing*.



#### 4.1.1.2. *Tokenizing*

```

JobTitle \
0 web development
1 data analyst
2 mobile app developer
3 it web developer
4 web developer / designer

JobDeskripsi \
0 kami mencari software engineer yang berpengalaman...
1 sebagai data analyst yang dapat menentukan bes...
2 kami mencari mobile app developer yang berpeng...
3 mengembangkan dan memelihara web responsif men...
4 kami mencari web developer untuk pengembangan ...

Requirement \
0 gelar sarjana ilmu komputer, minimal 3 tahun p...
1 sarjana teknik, memiliki pengalaman kerja di b...
2 gelar sarjana teknik informatika, pengalaman da...
3 gelar sarjana teknik informatika, memiliki pen...
4 gelar sarjana, minimal 2 tahun pengalaman di b...

tokenized_text
0 [[web, development], [kami, mencari, software,...
1 [[data, analyst], [sebagai, data, analyst, yan...
2 [[mobile, app, developer], [kami, mencari, mob...
3 [[it, web, developer], [mengembangkan, dan, me...
4 [[web, developer, /, designer], [kami, mencari...

```

Gambar 4. 5 Data setelah dan sebelum proses *tokenizing*

Pada gambar 4.5 merupakan gambar saat data melalui proses *tokenizing*, proses ini digunakan untuk memecah kalimat menjadi perkata sehingga menghasilkan data yang lebih mudah untuk dapat diolah oleh model. Dapat dilihat pada gambar 4.3 tersebut bahwa data berhasil melalui proses *tokenizing*.

#### 4.1.1.3. *Stopword Removal*

Proses setelah dilakukannya *tokenizing* yaitu proses *stopword removal* dimana proses ini merupakan proses untuk menghilangkan kata yang tidak memiliki makna pada data. Tujuan dilakukan proses *stopword removal* ini yaitu untuk meningkatkan kinerja model. Pada gambar 4.6 merupakan gambar sebelum data dilakukan proses *stopword removal*.

```

JobTitle \
0      web development
1      data analyst
2      mobile app developer
3      it web developer
4      web developer / designer

JobDeskripsi \
0      kami mencari software engineer yang berpengalaman...
1      sebagai data analyst yang dapat menentukan bes...
2      kami mencari mobile app developer yang berpeng...
3      mengembangkan dan memelihara web responsif men...
4      kami mencari web developer untuk pengembangan ...

Requirement \
0      gelar sarjana ilmu komputer, minimal 3 tahun p...
1      sarjana teknik, memiliki pengalaman kerja di b...
2      gelar sarjana teknik informatika, pengalaman da...
3      gelar sarjana teknik informatika, memiliki pen...
4      gelar sarjana, minimal 2 tahun pengalaman di b...

```

Gambar 4. 6 Data sebelum dilakukan proses *stopword removal*

```

JobTitle_removed_stopwords \
0      web development
1      data analyst
2      mobile app developer
3      it web developer
4      web developer / designer

JobDeskripsi_removed_stopwords \
0      mencari software engineer berpengalaman pengem...
1      data analyst menentukan best practies proces e...
2      mencari mobile app developer berpengalaman men...
3      mengembangkan memelihara web responsif kerangk...
4      mencari web developer pengembangan website men...

Requirement_removed_stopwords \
0      gelar sarjana ilmu komputer , minimal 3 pengal...
1      sarjana teknik , memiliki pengalaman kerja bid...
2      gelar sarjana teknik informatika , pengalaman ...
3      gelar sarjana teknik informatika , memiliki pe...
4      gelar sarjana , minimal 2 pengalaman bidang we...

```

Gambar 4. 7 Data setelah dilakukan proses *stopword removal*

Pada gambar 4.7 merupakan gambar hasil setelah dilakukan proses *stopword removal*, sehingga data sudah bersih dari kata yang tidak memiliki makna, kemudian setelah dilakukannya proses *stopword removal* data akan melalui proses *stemming*.

#### 4.1.1.4. *Stemming*

*Stemming* merupakan proses yang bertujuan untuk menyederhanakan sebuah kata sehingga menjadi kata dasar. Data dilakukan proses *stemming* diharapkan akan meningkatkan kualitas data sehingga pada proses pemodelan meningkatkan akurasi.

```

JobTitle \
0 web development
1 data analyst
2 mobile app developer
3 it web developer
4 web developer / designer

JobDeskripsi \
0 kami mencari software engineer yang berpengalaman...
1 sebagai data analyst yang dapat menentukan bes...
2 kami mencari mobile app developer yang berpeng...
3 mengembangkan dan memelihara web responsif men...
4 kami mencari web developer untuk pengembangan ...

Requirement JobTitle_stemmed \
0 gelar sarjana ilmu komputer, minimal 3 tahun p... web development
1 sarjana teknik, memiliki pengalaman kerja di b... data analyst
2 gelar sarjana teknik informatika, pengalaman da... mobile app developer
3 gelar sarjana teknik informatika, memiliki pen... it web developer
4 gelar sarjana, minimal 2 tahun pengalaman di b... web developer designer

JobDeskripsi_stemmed \
0 kami cari software engineer yang alan dalam ke...
1 bagai data analyst yang dapat tentu best pract...
2 kami cari mobile app developer yang alan untuk...
3 kembang dan pelihara web responsif guna kerang...
4 kami cari web developer untuk kembang website ...

Requirement_stemmed
0 gelar sarjana ilmu komputer minimal 3 tahun al...
1 sarjana teknik milik alan kerja di bidang kait...
2 gelar sarjana teknik informatika alan dalam ke...
3 gelar sarjana teknik informatika milik alan ke...
4 gelar sarjana minimal 2 tahun alan di bidang w...

```

Gambar 4. 8 Data setelah dan sebelum dilakukan proses *stemming*

Pada gambar 4.8 merupakan data saat dilakukannya proses *stemming*, dapat dilihat bahwa proses berhasil dilakukan sehingga pada data hanya terdapat kata dasar yang tidak memiliki imbuhan, sehingga kualitas data cukup baik.

#### 4.1.3. Hasil Perhitungan Pencocokan

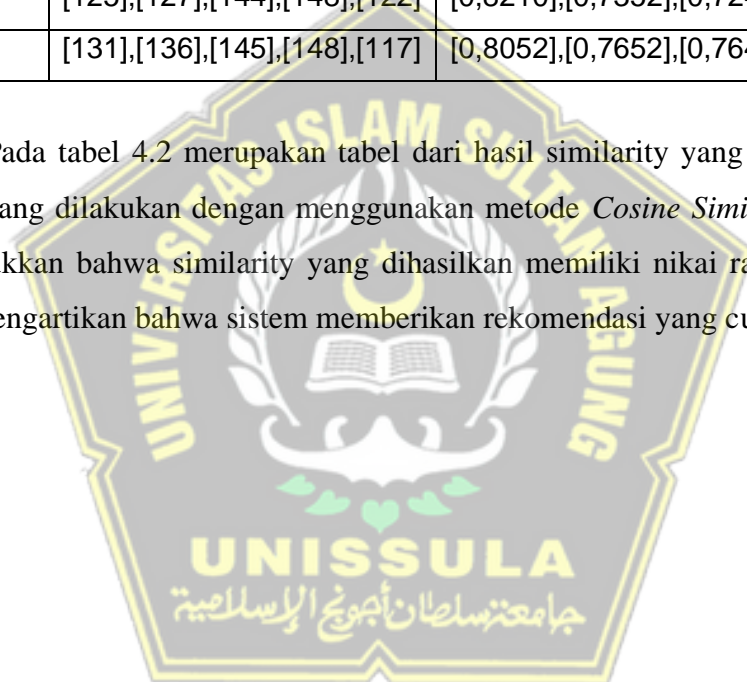
Pada proses pengolahan data terdapat perhitungan untuk mendapatkan hasil kecocokan antara data lowongan pekerjaan dengan data user yang perhitungan tersebut dilakukan dengan metode *cosine similarity*. Hasil dari perhitungan tersebut akan dijadikan sebagai bahan acuan untuk bisa merekomendasikan lowongan pekerjaan yang cocok untuk user.

Tabel 4. 2 Tabel hasil Similarity

ID Kandidat	Rekomendasi sistem	Score Similarity
B01	[22],[26],[27],[23],[43]	[0,8052],[0,7652],[0,7648],[0,7543],[0,7423]
B02	[3],[10],[8],[20],[55]	[0,7811],[0,7623],[0,7192],[0,7325],[0,7331]
B03	[33],[3],[24],[21],[43]	[0,8231],[0,7741],[0,7310],[0,7303],[0,6329]
B04	[23],[33],[12],[8],[9]	[0,6125],[0,6012],[0,5521],[0,5519],[0,5291]

B05	[19],[23],[13],[4],[10]	[0,7219],[0,6852],[0,6781],[0,6601],[0,6510]
B21	[69],[88],[72],[98],[89]	[0,7901],[0,7852],[0,6643],[0,6613],[0,6215]
B22	[66],[87],[66],[95],[73]	[0,7921],[0,7907],[0,7624],[0,6921],[0,6801]
B23	[61],[93],[82],[68],[79]	[0,8421],[0,7990],[0,7871],[0,7751],[0,7463]
B24	[77],[79],[81],[78],[100]	[0,6521],[0,6410],[0,5901],[0,5771],[0,5579]
B25	[68],[86],[62],[93],[91]	[0,8210],[0,7121],[0,7110],[0,6214],[0,6523]
B41	[140],[106],[139],[118],[132]	[0,7891],[0,7877],[0,7801],[0,7781],[0,7753]
B42	[123],[113],[136],[142],[128]	[0,6651],[0,6519],[0,6219],[0,6102],[0,6013]
B43	[148],[119],[142],[138],[129]	[0,7425],[0,7341],[0,7218],[0,6926],[0,6691]
B44	[125],[127],[144],[148],[122]	[0,8210],[0,7352],[0,7247],[0,7219],[0,7423]
B45	[131],[136],[145],[148],[117]	[0,8052],[0,7652],[0,7648],[0,7543],[0,7423]

Pada tabel 4.2 merupakan tabel dari hasil similarity yang dilakukan oleh sistem yang dilakukan dengan menggunakan metode *Cosine Similarity*. Hasil ini menunjukkan bahwa similarity yang dihasilkan memiliki nilai rata-rata 0,8991. Yang mengartikan bahwa sistem memberikan rekomendasi yang cukup baik.



#### 4.1.4. Hasil Pelatihan Model

Tabel 4. 3 Hasil evaluasi AP@K

ID	SKILL	RIWAYAT PEKERJAAN	RIWAYAT PENDIDIKAN	UMUR	JENIS KELAMIN	AP @3
B01	HTML, CSS, JavaScript, React	Frontend Developer di PT. Teknologi	S1 Teknik Informatika	25	Pria	0,8333
B02	PHP, Laravel, MySQL	Backend Developer di PT. Digital	S1 Ilmu Komputer	30	Pria	0,6552
B03	HTML, Bootstrap, Node.js	Web Developer di PT. Kreatif	S1 Sistem Informasi	27	Pria	0,8233
B04	JavaScript, Vue.js, MongoDB	Full-stack Developer di Startup ABC	S1 Teknik Komputer	25	Wanita	0,7210
B05	Angular, TypeScript, SQL	Frontend Developer di PT. Solusi Digital	S1 Teknologi Informasi	29	Pria	0,82177
B21	Kotlin, Android Studio, SQLite	Android Developer di PT. Aplikasi Kita	S1 Teknik Informatika	30	Wanita	0,8214
B22	Swift, Xcode, Core Data	iOS Developer di Startup Mobile	S1 Teknologi Informasi	35	Wanita	0,7218
B23	React Native, JavaScript, Firebase	Mobile Developer di PT. Digital Apps	S1 Sistem Informasi	32	Pria	0,8623
B24	Flutter, Dart, SQLite	Mobile Developer di Startup Tech	S1 Teknik Komputer	27	Pria	0,7371
B25	Swift, Objective-C, Core Data	iOS Developer di PT. Aplikasi Modern	S1 Teknik Informatika	28	Wanita	0,8241

Pada tabel 4.3 merupakan tabel hasil evaluasi menggunakan AP @K dengan nilai  $k=3$ , Metode similaritas memiliki efektivitas sedang Performa rekomendasi paling baik di top 2-4 item Nilai AP@3 menghasilkan score diatas 0,6 yang menandakan bahwa sistem bekerja dengan baik.

#### 4.1.5. Hasil Evaluasi

Untuk mengevaluasi hasil pemodelan maka dilakukan proses evaluasi dengan menggunakan *ground truth*. Pada evaluasi ini akan mengukur *precision* dan *recall*.



Tabel 4. 4 Tabel hasil *Precision* dan *Recall*

Pelamar	Jumlah Rekomendasi	Relevan	Total Relevan	Precision	Recall
B05	10	7	50	0.7	0.014
B32	15	12	48	0.8	0.016
B44	18	14	40	0,7	0,017

Pada tabel 4.4 merupakan tabel dari hasil precision dan recall pada evaluasi AP @K. Hasil nya menunjukkan bahwa padarata rata precision yaitu 0,733 dan rata rata recallnya yaitu 0,0156. Hasil tersebut menunjukkan bahwa pemodelan cukup baik dalam merekomendasikan lowongan pekerjaan kepada pelamar.

Tabel 4. 5 Pelabelan dataset untuk evaluasi *ground truth*

ID	SKILL	RIWAYAT PEKERJAAN	RIWAYAT PENDIDIKAN	UMUR	JENIS KELAMIN	Ground Truth	Rekomendasi sistem	Revelansi
B01	HTML, CSS, JavaScript, React	Frontend Developer di PT. Teknologi	S1 Teknik Informatika	25	Pria	[22]	[22],[26],[27],[23],[43]	1
B02	PHP, Laravel, MySQL	Backend Developer di PT. Digital	S1 Ilmu Komputer	30	Pria	[10]	[3],[10],[8],[20],[55]	0
B03	HTML, Bootstrap, Node.js	Web Developer di PT. Kreatif	S1 Sistem Informasi	27	Pria	[33]	[33],[3],[24],[21],[43]	1
B04	JavaScript, Vue.js, MongoDB	Full-stack Developer di Startup ABC	S1 Teknik Komputer	25	Wanita	[23]	[23],[33],[12],[8],[9]	1
B05	Angular, TypeScript, SQL	Frontend Developer di PT. Solusi Digital	S1 Teknologi Informasi	29	Pria	[19]	[19],[23],[13],[4],[10]	1
B21	Kotlin, Android	Android Developer di	S1 Teknik Informatika	30	Wanita	[69]	[69],[88],[72],[98],[89]	1

	Studio, SQLite	PT. Aplikasi Kita						
B22	Swift, Xcode, Core Data	iOS Developer di Startup Mobile	S1 Teknologi Informasi	35	Wanita	[66]	[66],[87],[66],[95],[73]	1
B23	React Native, JavaScript , Firebase	Mobile Developer di PT. Digital Apps	S1 Sistem Informasi	32	Pria	[61]	[61],[93],[82],[68],[79]	1
B24	Flutter, Dart, SQLite	Mobile Developer di Startup Tech	S1 Teknik Komputer	27	Pria	[79]	[77],[79],[81],[78],[100]	0
B25	Swift, Objective- C, Core Data	iOS Developer di PT. Aplikasi Modern	S1 Teknik Informatika	28	Wanita	[68]	[68],[86],[62],[93],[91]	1

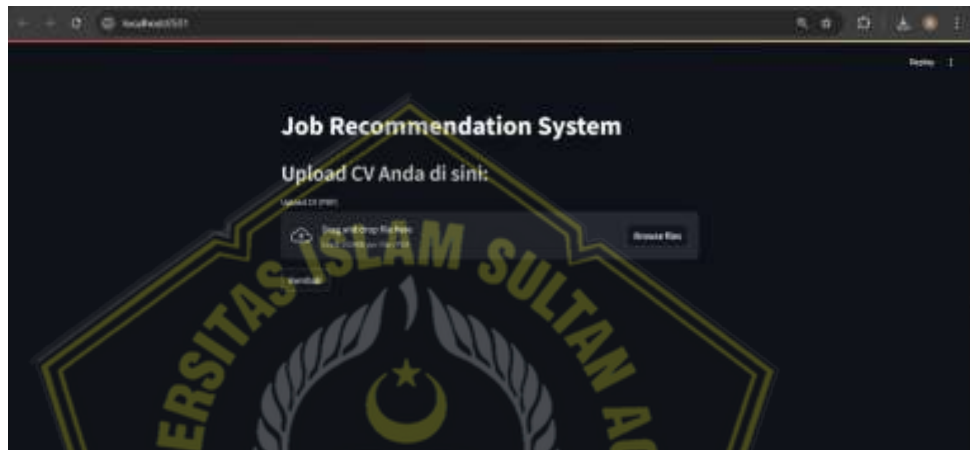
Tabel 4.5 merupakan tabel setelah melalui evaluasi AP @K yang kemudian akan dilakukan evaluasi apakah hasil rekomendasi relevan dengan data asli (Ground Truth) lowongan pekerjaan. dapat dilihat bahwa bisa dikatakan sistem memberikan tingkat akurasi yang cukup baik. Dapat dilihat bahwa dari 10 data yang ditampilkan hanya terdapat 2 yang rekomendasinya berbeda dengan ground truth.

#### 4.2 Hasil Implementasi Sistem



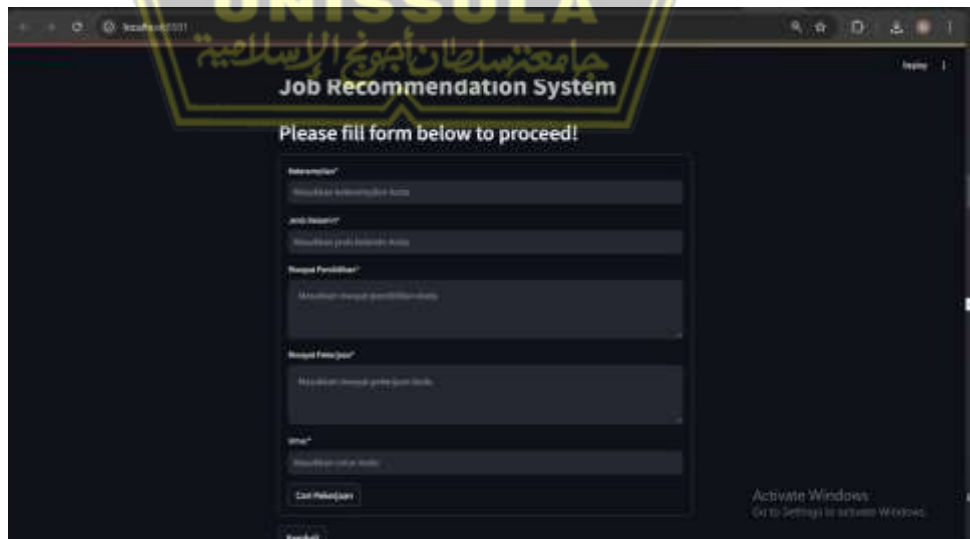
Gambar 4. 9 Halaman utama

Pada gambar 4. 10 ini merupakan sebuah tampilan awal website rekomendasi pekerjaan. Sebelumnya model IndoBERT dan *Cosine Similarity* telah dilatih dan kemudian diimplementasikan menggunakan *framework* Streamlit. Pada halaman awal ini terdapat pilihan dalam melakukan input data pelamar, pelamar dapat memilih pengisian form atau *upload CV* untuk melakukan input data diri pelamar.



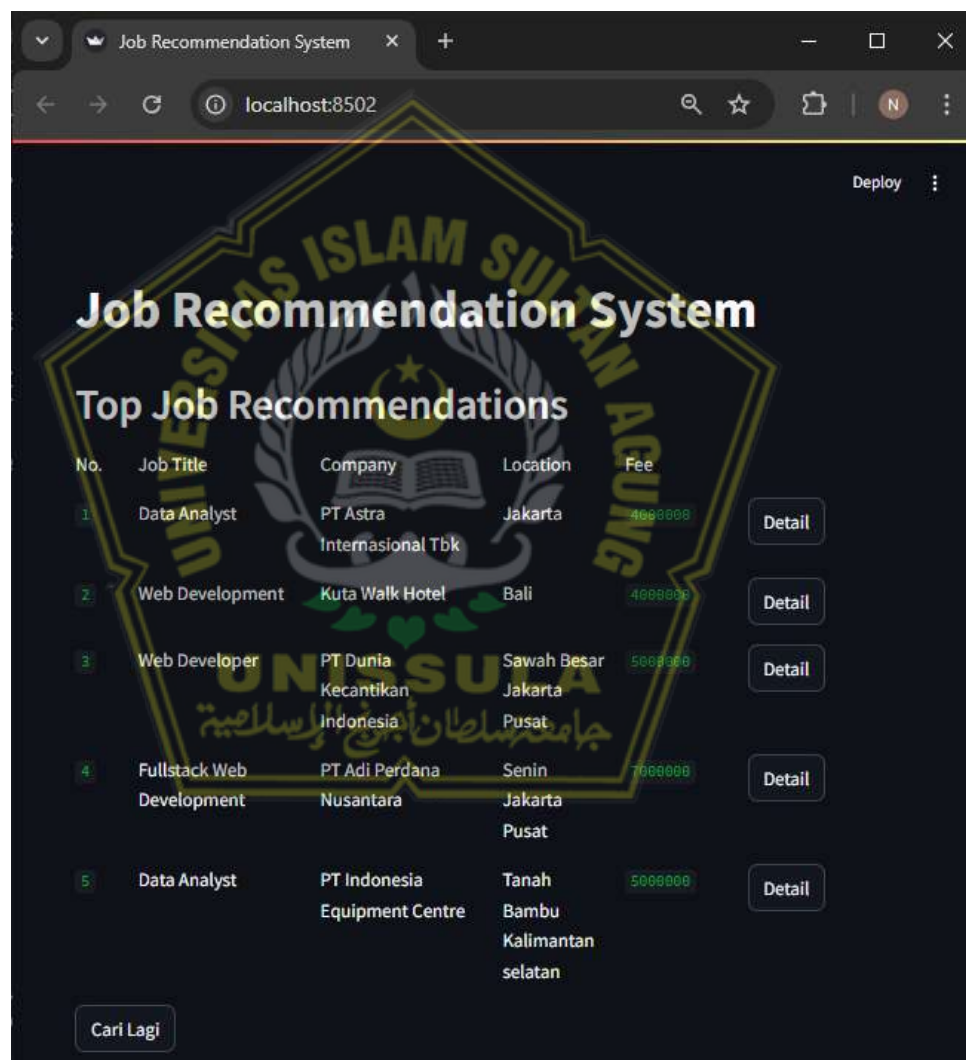
Gambar 4. 10 Tampilan implementasi halaman untuk upload CV user

Pada gambar 4.11 merupakan tampilan dari interface yang digunakan untuk upload CV dengan format PDF oleh user. Tampilan ini akan muncul jika pada halaman awal user memilih option untuk upload CV.



Gambar 4. 11 Tampilan implementasi halaman pengisian form data diri

Pada gambar 4.12 merupakan tampilan jika pada halaman awal user memilih untuk pengisian form. Pada halaman ini user diwajibkan untuk mengisi semua *field* yang ada pada halaman ini. Pada halaman ini terdapat form keterampilan, jenis kelamin, riwayat Pendidikan, riwayat pekerjaan dan juga umur. Yang kemudian setelah pengisian user bisa menekan tombol cari pekerjaan untuk memproses data sehingga akan tampil halaman hasil rekomendasi pekerjaan seperti pada gambar 4.12.



Gambar 4. 12 Tampilan hasil rekomendasi pekerjaan

Pada gambar 4.13 merupakan tampilan website yang menampilkan data hasil rekomendasi kepada user, setelah user melakukan input data dihalaman awal maka setelah diproses akan menuju pada tampilan. Ini data yang ditampilkan mulai yang dari mirip dengan data user. Data yang tampil pada halaman ini yaitu

nama pekerjaan, nama perusahaan, alamat, gaji dan juga tombol detail. Pada halaman hasil rekomendasi ini terdapat 2 aksi yang dapat dilakukan oleh user yaitu dengan menekan tombol detail untuk melihat detail dari lowongan pekerjaan dan juga tombol coba lagi yang akan kembali pada halaman pengisian form.

### 4.3 Hasil Pengujian Sistem

Dalam pengujian sistem terdapat beberapa pengujian seperti pengujian performa sistem dan pengujian *Black Box*. Pada tabel 4.1 merupakan hasil dari pengujian performa sistem dan tabel 4.2 merupakan hasil dari pengujian *Black Box*.

#### 4.3.1. Pengujian Performa Sistem

Pengujian performa sistem ini dilakukan agar mengetahui apakah terjadi kegagalan seperti response time yang lama sehingga dapat menyebabkan gangguan pada user, sehingga sistem dapat dilakukan pengecekan kembali jika terjadi gangguan. Pada tabel 4.6 merupakan hasil dari pengujian sistem job recommendation yang penulis lakukan.

Tabel 4. 6 Tabel Hasil Pengujian Performa Sistem

<i>INPUT</i>	<i>OUTPUT</i>	<i>Time</i>
Data form dengan CV yang memiliki 189 kata	Menampilkan halaman hasil rekomendasi pekerjaan	03:01
Data form dengan CV yang memiliki 325 kata	Menampilkan halaman hasil rekomendasi pekerjaan	03:22
Data form dengan CV yang memiliki 457 kata	Menampilkan halaman hasil rekomendasi pekerjaan	03:51

Hasil pengujian menunjukkan bahwa sistem berjalan dengan normal tergantung dengan kecepatan jaringan yang user gunakan. Pada pengujian yang penulis lakukan menggunakan kecepatan internet 100 sampai 150 *Kbps*. Performa dari sistem juga tergantung pada data form dan CV yang user inputkan. Semakin



banyak jumlah kata maka akan mempengaruhi kecepatan sistem dalam menampilkan hasil rekomendasi.

#### 4.3.2. Pengujian *Black Box*

Pengujian *black box* dilakukan untuk dapat mengetahui fungsi dari sistem apakah sesuai dengan harapan dari peneliti, pengujian *black box* dilakukan apakah pengisian form dan upload CV dapat digunakan dengan baik dan sehingga menghasilkan rekomendasi pekerjaan sesuai dengan data user dan harapan peneliti.

Tabel 4. 7 Hasil pengujian *black box* pada sistem

<i>INPUT</i>	Hasil Harapan Penulis	<i>OUTPUT</i>	Keterangan
Pengisian data form user	Menampilkan halaman hasil rekomendasi pekerjaan	Menampilkan halaman hasil rekomendasi pekerjaan	Berhasil
Upload CV user	Menampilkan halaman hasil rekomendasi pekerjaan	Menampilkan halaman hasil rekomendasi pekerjaan	Berhasil
Pengisian data form dan Upload CV user	Menampilkan halaman hasil rekomendasi pekerjaan	Menampilkan halaman hasil rekomendasi pekerjaan	Berhasil

Pada tabel 4.7 merupakan hasil dari pengujian *black box* yang penulis lakukan, hasil menunjukkan bahwa pengisian form dan upload CV dapat berfungsi dengan baik dan menghasilkan rekomendasi sesuai. Hasil dari evaluasi sesuai dengan yang diharapkan, saat pelamar mengisi data form maka sistem berhasil menampilkan hasil rekomendasi pekerjaan 5 data yang paling relevan. Kemudian dengan upload CV juga menunjukkan sistem berhasil merekomendasikan 5 data paling relevan lowongan pekerjaan yang sesuai dengan data CV.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Setelah melakukan penelitian ini dapat disimpulkan bahwa penggunaan metode IndoBERT dan Cosine Similarity bekerja dengan baik dalam merekomendasikan dengan hasil yang akurat pada saat evaluasi metode. Dari hasil evaluasi menunjukkan model melakukan kinerja yang baik dalam merekomendasikan sebuah lowongan pekerjaan pada sistem dan sesuai dengan data user yang di inputkan oleh user. IndoBERT berhasil mengubah text menjadi sebuah vektor dan kemudian akan diproses oleh Cosine Similarity untuk menguji kemiripan antara data user dengan lowongan pekerjaan. Hasil evaluasi menggunakan AP @K menunjukkan bahwa rata-rata precision yaitu 0,733 dan rata-rata recallnya yaitu 0,0156. Sistem layanan Job Recommendation ini menggunakan Streamlit dalam pengembangan sistemnya. Namun pada sistem ini masih memiliki kekurangan dikarenakan hanya sampai pada menampilkan data hasil rekomendasi. Pada sistem ini belum terdapat fitur *apply* yang bisa digunakan pelamar untuk melamar langsung ke perusahaan atau ke lowongan yang sudah direkomendasikan.

#### 5.2 Saran

Dengan penelitian yang sudah dilakukan, penulis mengharapkan ada pengembangan sistem yang lebih baik dikemudian hari. Dengan itu penulis memberi saran agar sistem layanan dapat dikembangkan supaya user atau pengguna dapat melakukan aksi *apply* pekerjaan langsung kepada perusahaan yang sedang melakukan recruitment. Dan juga penambahan ruang lingkup lowongan pekerjaan sehingga sistem dapat merekomendasikan lebih banyak lagi lowongan pekerjaan yang ditampilkan. Dengan kekurangan sistem yang tidak adanya fitur *apply* dan informasi kontak perusahaan, diharapkan adanya pengembangan fitur dan juga peningkatan tingkat akurasi pada sistem rekomendasi.

## DAFTAR PUSTAKA

- Dharmawan, S., Viny, Mawardi, C., Novario, & Perdana, J. (2022). *Jurnal Ilmu Komputer dan Sistem Informasi Klasifikasi Ujaran Kebencian Menggunakan Metode FeedForward Neural Network (IndoBERT)*.
- Elfirdaus Ivana, Wahyuni, & Eka Dyar. (2023). *Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2023 Surabaya*.  
[https://www.imdb.com/?ref\\_=nv\\_home](https://www.imdb.com/?ref_=nv_home)
- Flores, V. A., Permatasari, P. A., & Jasa, L. (2020). Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword. *Majalah Ilmiah Teknologi Elektro*, 19(2), 157.  
<https://doi.org/10.24843/mite.2020.v19i02.p06>
- Habibi, Roni, Dzihan Albanna, & Muhammad. (2022). ANALISIS SISTEM REKOMENDASI PADA JOB RECOMMENDATION BERDASARKAN PROFIL LINKEDIN MENGGUNAKAN COSINE SIMILARITY. Dalam *Jurnal Teknik Informatika* (Vol. 14, Nomor 3).
- Koloman, C., Maulana, R., Dwi, R., Putri, Z., & Harahap, W. A. (2023). Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering. *Journal of Creative Student Research (JCSR)*, 1(6), 78–88.  
<https://doi.org/10.55606/jcsrpolitama.v1i6.2992>
- Muarif, Ahmad Samsul, Winarno, & Edy. (2022). Sistem Rekomendasi Tempat Parkir di Kota Lama Semarang Menggunakan Collaborative Filtering. *Jurnal Ilmiah Universitas Batanghari Jambi*, 22(2), 906.  
<https://doi.org/10.33087/jiubj.v22i2.2066>
- Nastiti, P. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *Teknika*, 8(1), 1–10.  
<https://doi.org/10.34148/teknika.v8i1.139>
- Saadah, S., Kaenova Mahendra Auditama, Ananda Affan Fattahila, Fendi Irfan Amorokhman, Annisa Aditsania, & Aniq Atiqi Rohmawati. (2022). Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(4), 648–655. <https://doi.org/10.29207/resti.v6i4.4215>

- Vaswani, A. dkk. (2017) "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017-Decem(Nips), hal. 5999–6009.
- Alsaif, S. A., Sassi Hidri, M., Eleraky, H. A., Ferjani, I., & Amami, R. (2022). Learning-Based Matched Representation System for Job Recommendation. *Computers*, 11(11). <https://doi.org/10.3390/computers11110161>
- Baharuddin, Fikri Naufal, & Mohammad Farid. (2023). Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom's Taxonomy. *Journal of Information Systems Engineering and Business Intelligence*, 9(2), 253–263. <https://doi.org/10.20473/jisebi.9.2.253-263>
- Parwita, W. G. S. (2019). Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 14(1), 27. <https://doi.org/10.30872/jim.v14i1.1272>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-Decem(Nips), 5999–6009.
- Wahyuni, R. T., Prastiyanto, D., & Suprpto, D. E. (t.t.). *Penerapan Algoritma Cosine Similarity pada Sistem Klasifikasi Dokumen Skripsi*.
- Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2022). *Are Transformers Effective for Time Series Forecasting?* <http://arxiv.org/abs/2205.13504>