

# Implementasi Semantic Matching Menggunakan Sentence BERT dan KNN Pada Sistem Rekomendasi Lowongan Kerja

Michael Roda Robert<sup>1</sup>, Ery Hartati<sup>2</sup>

<sup>1,2</sup>Program Studi Informatika, Universitas Multi Data Palembang, JL. Rajawali No.14, Palembang dan Indonesia  
ggwpg32@mhs.mdp.ac.id<sup>1</sup>, ery\_hartati@mdp.ac.id<sup>2</sup>

**Abstrak.** Peningkatan jumlah lowongan kerja pada platform digital menyebabkan pencari kerja sering mengalami kesulitan dalam menemukan pekerjaan yang sesuai dengan keterampilan dan pengalaman yang dimiliki. Permasalahan tersebut dapat menimbulkan terjadinya job mismatch akibat banyaknya informasi lowongan yang tersedia dan kompleksitas deskripsi pekerjaan. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi lowongan kerja berbasis semantic matching menggunakan Sentence-BERT (SBERT) dan K-Nearest Neighbors (KNN) untuk menghasilkan rekomendasi pekerjaan yang lebih relevan dan kontekstual. Dataset yang digunakan berasal dari LinkedIn Job Postings Dataset 2023–2024 yang diperoleh dari Kaggle dengan sekitar 55.000 data lowongan pekerjaan. Tahapan penelitian meliputi preprocessing data, pembentukan embedding menggunakan SBERT, serta proses pencarian dan pemeringkatan lowongan kerja menggunakan KNN berdasarkan cosine similarity. Evaluasi sistem dilakukan menggunakan beberapa query pengguna dan ground truth relevansi yang dibentuk secara semi-manual. Hasil pengujian menunjukkan bahwa sistem mampu menghasilkan rekomendasi lowongan kerja yang relevan dengan nilai average similarity berkisar antara 72,4% hingga 77,1%. Hasil tersebut menunjukkan bahwa pendekatan semantic matching menggunakan SBERT dan KNN cukup efektif dalam memahami hubungan semantik antara kebutuhan pengguna dan deskripsi pekerjaan sehingga dapat membantu meningkatkan relevansi rekomendasi lowongan kerja.

**Kata Kunci :** semantic matching, Sentence-BERT, K-Nearest Neighbors, sistem rekomendasi, lowongan kerja.

**Abstract.** The increasing number of job vacancies on digital recruitment platforms has made it difficult for job seekers to find positions that match their skills and experience. This issue can lead to job mismatch due to the large amount of available job information and the complexity of job descriptions. This study aims to develop a job recommendation system based on semantic matching using Sentence-BERT (SBERT) and K-Nearest Neighbors (KNN) to generate more relevant and contextual job recommendations. The dataset used in this research was obtained from the LinkedIn Job Postings Dataset 2023–2024 available on Kaggle, consisting of approximately 55,000 job vacancy records. The research process includes data preprocessing, embedding generation using SBERT, and job ranking using KNN based on cosine similarity. System evaluation was conducted using several user queries and semi-manually constructed relevance ground truth. The experimental results show that the system was able to generate relevant job recommendations with average similarity values ranging from 72.4% to 77.1%. These results indicate that the semantic matching approach using SBERT and KNN is effective in understanding the semantic relationship between user preferences and job descriptions, thereby improving the relevance of job recommendations.

**Keyword :** semantic matching, Sentence-BERT, K-Nearest Neighbors, recommendation system, job vacancies.

## PENDAHULUAN

Pencarian kerja merupakan proses yang penting bagi setiap individu, namun hingga saat ini masih banyak permasalahan yang muncul dalam menemukan pekerjaan yang sesuai dengan kompetensi dan pengalaman pencari kerja. Salah satu permasalahan utama adalah terjadinya ketidaksesuaian antara keterampilan tenaga kerja dengan kebutuhan industri [1]. Di Indonesia, hal ini menjadi isu serius, terbukti dari data Badan Pusat Statistik (BPS) yang menunjukkan bahwa 59,12% tenaga kerja bekerja tidak sesuai dengan bidang keahliannya [2]. Kondisi ini tidak hanya menghambat perkembangan karier individu, tetapi juga menurunkan produktivitas dan efisiensi pasar tenaga kerja nasional. Selain itu, setiap tahun tercatat lebih dari 3 juta perekrutan baru melalui linkedIn membuat proses pencarian lowongan yang sesuai menjadi semakin menantang [3].



Platform rekrutmen digital seperti LinkedIn melaporkan adanya lebih dari 65 juta pencarian pekerjaan setiap minggu dan sekitar 8 orang direkrut setiap menit pada tahun 2025 [4]. Meningkatnya jumlah lowongan kerja yang diunggah setiap tahun, yang dapat mencapai jutaan posting, mengakibatkan pencari kerja sering mengalami *information overload*. Deskripsi pekerjaan yang panjang, kompleks, dan bervariasi menyulitkan pencari kerja untuk menilai relevansi suatu lowongan secara cepat dan tepat [5]. Hal tersebut mempertegas perlunya sistem yang mampu menyaring dan merekomendasikan lowongan kerja yang relevan secara otomatis dan personal.

Sejumlah penelitian terdahulu menunjukkan bahwa sistem rekomendasi pekerjaan masih memiliki berbagai keterbatasan. Penelitian oleh Arshkon menggunakan model BERT pada dataset *LinkedIn Job Postings* dan menghasilkan precision sebesar 52,67% dan recall 76,67% untuk rekomendasi Top-5. Namun, proses perankingan hanya mengandalkan cosine similarity tanpa integrasi algoritma pencarian seperti *K-Nearest Neighbor* (KNN) [6], sehingga akurasi rekomendasi masih dapat ditingkatkan [7]. Pada domain yang berbeda, penelitian oleh A'la yang menerapkan IndoBERT untuk klasifikasi sentimen ulasan game berbahasa Indonesia menunjukkan peningkatan performa model setelah optimasi menggunakan teknik penanganan ketidakseimbangan data, yang membuktikan bahwa model berbasis *Transformer* sangat efektif dalam memahami konteks Bahasa Indonesia [8]. Temuan serupa juga dilaporkan oleh Wau yang mengimplementasikan IndoBERT pada tugas klasifikasi sentimen ulasan produk lokal dan mendapatkan hasil akurasi yang kompetitif [9]. Namun, kedua penelitian tersebut tidak diterapkan pada domain rekomendasi lowongan pekerjaan maupun dikombinasikan dengan algoritma ranking seperti KNN, sehingga belum mampu menjawab tantangan *mismatch* pekerjaan yang masih umum terjadi pada sistem rekomendasi pekerjaan saat ini.

Penelitian oleh Arshkon (2024) menggunakan model BERT dan cosine similarity dalam sistem rekomendasi lowongan kerja berbasis *semantic matching*. Penelitian tersebut menghasilkan precision sebesar 52,67% dan recall sebesar 76,67% pada rekomendasi Top-5. Namun, proses pencarian rekomendasi masih terbatas pada perhitungan *cosine similarity* tanpa menggunakan metode *nearest neighbor* sehingga relevansi rekomendasi masih dapat ditingkatkan.

Penelitian lain oleh A'la menggunakan IndoBERT untuk klasifikasi sentimen ulasan game berbahasa Indonesia dan menunjukkan peningkatan performa model setelah dilakukan optimasi penanganan data *imbalance*. Sementara itu, penelitian oleh Wau juga menunjukkan bahwa model IndoBERT mampu menghasilkan performa klasifikasi sentimen yang kompetitif pada ulasan produk lokal [10]. Akan tetapi, kedua penelitian tersebut tidak diterapkan pada domain rekomendasi lowongan kerja dan tidak menggunakan pendekatan *semantic matching* maupun proses pemeringkatan berbasis *nearest neighbor*.

Berbeda dengan penelitian sebelumnya, penelitian ini menggabungkan *Sentence-BERT* (SBERT) dan algoritma *K-Nearest Neighbors* (KNN) untuk membangun sistem rekomendasi lowongan kerja berbasis *semantic matching*. SBERT digunakan untuk menghasilkan representasi embedding teks yang mampu memahami hubungan semantik antara profil pengguna dan deskripsi pekerjaan, sedangkan KNN digunakan untuk proses pencarian dan pemeringkatan lowongan kerja berdasarkan tingkat kemiripan embedding. Dengan pendekatan tersebut, sistem diharapkan mampu menghasilkan rekomendasi yang lebih relevan, kontekstual, dan sesuai dengan kebutuhan pengguna.

Sebagai solusi, perkembangan teknologi *Natural Language Processing* (NLP) berbasis arsitektur Transformer memungkinkan penerapan pendekatan *semantic matching*. Model seperti *Sentence-BERT* (SBERT) mampu menghasilkan representasi semantik kalimat dalam bentuk embedding, sehingga kesesuaian antar teks dapat diukur secara lebih akurat menggunakan *cosine similarity*. Selanjutnya, algoritma *K-Nearest Neighbors* (KNN) dapat digunakan untuk melakukan proses pencarian dan perankingan lowongan kerja berdasarkan kedekatan vektor embedding, sehingga menghasilkan rekomendasi *Top-K* yang lebih relevan.

Rencana pemecahan masalah dalam penelitian ini adalah dengan mengembangkan sistem rekomendasi lowongan kerja berbasis *semantic matching* yang mengintegrasikan SBERT sebagai

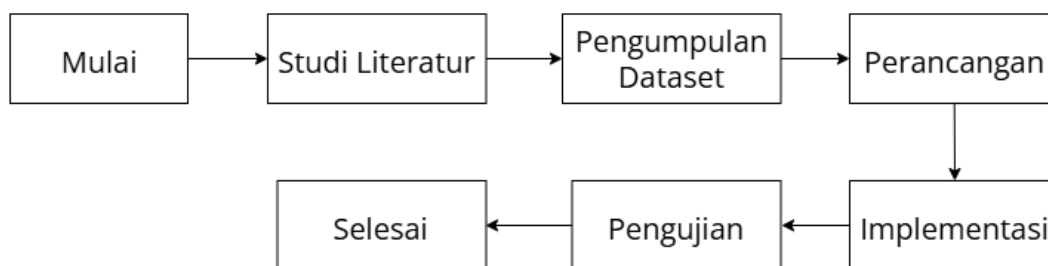


model pembentuk representasi semantik dan KNN sebagai metode pencarian serta perankingan. Sistem ini dirancang untuk mengatasi keterbatasan pendekatan konvensional, meningkatkan relevansi rekomendasi, serta mengurangi tingkat ketidaksesuaian pekerjaan.

Tujuan dari penelitian ini adalah untuk menghasilkan sistem rekomendasi lowongan kerja yang lebih akurat, kontekstual, dan sesuai dengan kebutuhan pengguna. Selain itu, penelitian ini bertujuan untuk meningkatkan kualitas pencocokan antara pencari kerja dan lowongan pekerjaan, sehingga dapat membantu mengurangi masalah *job mismatch* serta meningkatkan efisiensi dalam proses pencarian kerja di era digital.

### METODOLOGI PENELITIAN

Untuk meminimalisir kesalahan selama proses penelitian, perancangan metode diterapkan terlebih dahulu.



Gambar 1. Metodologi Penelitian

1. Studi Literatur

Tahapan ini dimulai dengan studi literatur dari jurnal ilmiah dan buku referensi yang relevan dengan sistem rekomendasi lowongan kerja berbasis pemahaman semantik teks. Kajian mencakup konsep sistem rekomendasi berbasis konten, penggunaan model BERT dan SBERT untuk representasi teks, serta algoritma KNN untuk pencarian dan pemeringkatan. Literatur yang digunakan berasal dari publikasi tahun 2020–2025 agar sesuai dengan perkembangan terbaru di bidang NLP, *information retrieval*, dan sistem rekomendasi.

2. Pengumpulan Dataset

Pada tahap ini dilakukan proses pengumpulan dataset lowongan pekerjaan yang bersifat terbuka (*open dataset*). Dataset yang digunakan adalah *LinkedIn Job Postings Dataset (2023–2024)* yang tersedia pada platform Kaggle. Dataset tersebut berisi informasi lowongan pekerjaan seperti *title*, *company*, *description*, *skills\_desc*, *formatted\_experience*, *work\_type*, *location*, dan *salary range*.

Dataset kemudian melalui proses seleksi atribut untuk memastikan hanya kolom yang relevan yang digunakan dalam pembentukan profil pekerjaan. Atribut yang digunakan meliputi *title*, *description*, *skills\_desc*, *formatted\_experience*, *work\_type* dan *location*. Data yang memiliki entri kosong dan tidak relevan dibersihkan (*data cleaning*) sehingga dapat diproses sebagai input model.

Tabel 1. Data Kaggle yang digunakan

Nama File	Total Lowongan	Total Kolom Asli	Total Kolom Yang Digunakan
LinkedIn Job Posting.csv	±55.000 Lowongan	±20 Kolom	7 Kolom

Tabel 2. Dataset LinkedIn Job Posting (2023-2024) yang digunakan

No	Atribut	Deskripsi Singkat
1	<i>Title</i>	Judul Pekerjaan
2	<i>Company name</i>	Perusahaan yang membuka lowongan



3	<i>Location</i>	Lokasi Pekerjaan
4	<i>Skills desc</i>	Keterampilan yang dibutuhkan
5	<i>Formatted experience level</i>	Level Pengalaman
6	<i>Description</i>	Deskripsi Pekerjaan Lengkap
7	<i>Job type</i>	Jenis Pekerjaan

### 3. Perancangan

Perancangan sistem bertujuan membangun rekomendasi lowongan kerja berbasis pemahaman semantik menggunakan BERT/SBERT untuk representasi teks dan KNN untuk pencarian serta pemeringkatan[11]. Proses dimulai dengan menggabungkan berbagai atribut teks lowongan menjadi satu dokumen agar informasi tersaji dalam satu konteks semantik.

Pra-pemrosesan dilakukan melalui normalisasi sederhana seperti lowercase, penghapusan karakter non-alfabet, dan penyeragaman frasa, tanpa *stemming* atau *stopword removal* karena BERT mampu memahami konteks secara otomatis. Proses ini dapat dinyatakan secara matematis sebagai fungsi transformasi teks, dimana setiap teks asli  $T$  dipetakan menjadi teks terstandarisasi  $T'$  melalui fungsi *preprocessing*  $f(\cdot)$  dengan demikian, hubungan tersebut dapat ditulis kedalam rumus seperti dibawah ini :

$$T' = f(T) \tag{1}$$

Dimana  $T$  merupakan teks awal seperti judul atau deskripsi lowongan pekerjaan, sedangkan  $T'$  merupakan hasil teks yang telah dinormalisasi dan siap untuk diproses lebih lanjut oleh model. Fungsi  $f(\cdot)$  mencakup operasi transformasi seperti konversi huruf kecil, penghapusan karakter khusus, serta penyederhanaan struktur frasa agar input menjadi lebih konsisten sebelum dimasukkan ke tahap embedding menggunakan SBERT[12].

Selanjutnya, BERT/SBERT digunakan untuk menghasilkan *embedding* dari data lowongan pekerjaan dan input pengguna ke dalam ruang vektor berdimensi tinggi yang sama, sehingga kedua representasi tersebut dapat dibandingkan secara semantik berdasarkan maknanya, bukan sekadar kesamaan kata. SBERT memetakan setiap teks yang telah diproses menjadi representasi vektor numerik yang menangkap konteks dan hubungan antar kata dalam kalimat. Proses pemetaan ini dapat dinyatakan secara matematis sebagai berikut:

$$E = SBERT(T') \tag{2}$$

Dimana  $E$  merupakan vektor *embedding* hasil dari model SBERT, sedangkan  $T'$  adalah teks hasil pra-pemrosesan. Dengan demikian, setiap data lowongan pekerjaan dan input pengguna akan direpresentasikan sebagai vektor  $E_{job}$  dan  $E_{user}$  dalam ruang vektor yang sama, sehingga memungkinkan perhitungan tingkat kemiripan antar keduanya pada tahap selanjutnya menggunakan ukuran kesamaan berbasis jarak atau sudut antar vektor[13].

Proses rekomendasi dilakukan dengan pendekatan *K-Nearest Neighbors* (KNN) menggunakan cosine distance untuk memperoleh top-K lowongan pekerjaan dengan tingkat kemiripan tertinggi terhadap representasi vektor pengguna. Dalam metode ini, setiap vektor embedding pengguna akan dibandingkan dengan seluruh vektor *embedding* lowongan pekerjaan di dalam dataset, kemudian dipilih sejumlah K data dengan jarak paling dekat atau nilai kemiripan paling tinggi. *Cosine distance* yang digunakan dalam proses ini merupakan bentuk turunan dari *cosine similarity*, yang secara matematis dapat dinyatakan sebagai berikut:

$$Cosine\ Distance(A, B) = 1 - \frac{A \cdot B}{\|A\| \|B\|} \tag{3}$$

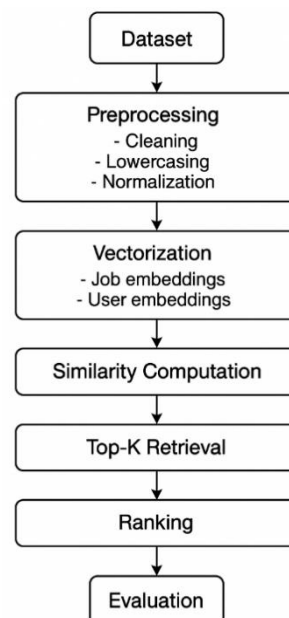
Dimana  $A$  adalah vektor embedding pengguna,  $B$  adalah vektor embedding lowongan pekerjaan, sedangkan  $\|A\|$  dan  $\|B\|$  merupakan norma dari masing-masing vektor[14]. Semakin kecil nilai *cosine distance*, maka semakin tinggi tingkat kemiripan antara kedua vektor, sehingga data tersebut akan diprioritaskan dalam hasil rekomendasi. Dengan

demikian, KNN akan mengurutkan seluruh kandidat lowongan berdasarkan nilai jarak tersebut dan memilih  $K$  data teratas sebagai hasil rekomendasi yang paling relevan.

Hasil rekomendasi kemudian dievaluasi menggunakan metrik *Precision* dengan membandingkan hasil rekomendasi yang dihasilkan sistem terhadap data ground truth yang telah ditentukan sebelumnya. Evaluasi ini bertujuan untuk mengukur tingkat ketepatan sistem dalam menampilkan lowongan yang relevan terhadap *query* pengguna. Nilai *precision* dihitung berdasarkan perbandingan antara jumlah rekomendasi yang relevan dengan total rekomendasi yang diberikan sistem[15]. Semakin tinggi nilai *precision*, semakin baik kemampuan sistem dalam menghasilkan rekomendasi yang sesuai dengan kebutuhan pengguna menggunakan rumus seperti dibawah ini:

$$Precision = \frac{Jumlah\ Rekomendasi\ Relevan}{Total\ Rekomendasi\ yang\ dihasilkan} \quad (4)$$

Skema perancangan sistem rekomendasi penelitian ini dapat dilihat dibawah ini.



Gambar 2. Skema Perancangan Sistem

#### 4. Implementasi

Pada tahapan ini dilakukan implementasi dari sistem yang telah dirancang sebelumnya agar sistem dapat memberikan rekomendasi lowongan pekerjaan yang sesuai dengan preferensi yang pengguna input kedalam sistem.

#### 5. Pengujian

Tahapan pengujian dilakukan untuk mengevaluasi kinerja sistem rekomendasi lowongan kerja berbasis embedding BERT/SBERT dan KNN dalam menghasilkan rekomendasi yang relevan berdasarkan kesesuaian semantik antara profil pengguna dan deskripsi pekerjaan. Evaluasi menggunakan sejumlah *user queries* yang telah dipersiapkan, di mana setiap *query* memiliki *ground truth* relevansi. Hasil rekomendasi sistem kemudian dibandingkan dengan *ground truth* untuk mengukur performa secara objektif.

Ground truth digunakan sebagai acuan utama dalam penilaian karena dataset *LinkedIn Job Postings* tidak menyediakan label relevansi eksplisit. Oleh karena itu, *ground truth* dibentuk menggunakan pendekatan semi-manual (*hybrid*), yaitu melalui penyusunan *query* pengguna simulatif yang mencakup *job title*, keterampilan, dan tingkat pengalaman,

kemudian dilakukan penyaringan awal berbasis aturan (kecocokan judul, skill, dan level pengalaman) untuk memperoleh kandidat lowongan relevan. Kandidat tersebut selanjutnya diverifikasi secara manual guna memastikan kesesuaian semantik. *Ground truth* ini digunakan untuk membandingkan hasil Top-K rekomendasi sistem.

Evaluasi kinerja dilakukan menggunakan metrik *Precision* yang umum digunakan dalam information retrieval. *Precision* mengukur ketepatan rekomendasi. Perhitungan metrik dilakukan berdasarkan perbandingan antara hasil rekomendasi sistem dan *ground truth* menggunakan persamaan yang telah ditentukan.

### HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset lowongan kerja yang diperoleh dari platform *Kaggle*, yang berisi informasi terkait berbagai posisi pekerjaan, perusahaan, lokasi, serta deskripsi pekerjaan. Dataset yang digunakan mencakup beberapa atribut utama seperti *job title, company name, location, job description, dan formatted experience level*. Data deskripsi pekerjaan kemudian dimanfaatkan sebagai representasi utama dalam proses pembentukan fitur teks untuk sistem rekomendasi.

Sebelum digunakan dalam proses pemodelan, data terlebih dahulu melalui tahapan preprocessing yang meliputi pembersihan data (*data cleaning*), penanganan nilai kosong (*missing values*), serta normalisasi teks. Selain itu, dilakukan pula tahapan *text preprocessing* berupa *case folding* untuk menyeragamkan huruf, serta penghapusan karakter yang tidak diperlukan. Tahapan ini bertujuan untuk meningkatkan kualitas data sehingga dapat menghasilkan representasi yang lebih baik dalam proses embedding.

Sistem rekomendasi yang dikembangkan pada penelitian ini menggunakan pendekatan berbasis *content-based filtering* dengan memanfaatkan representasi semantik teks menggunakan model embedding berbasis *Sentence-BERT* (SBERT). Model ini digunakan untuk mengubah teks menjadi vektor numerik sehingga dapat dihitung tingkat kemiripannya. Selanjutnya, algoritma *K-Nearest Neighbors* (KNN) digunakan untuk mencari sejumlah data pekerjaan yang memiliki kemiripan paling tinggi dengan input pengguna berdasarkan nilai cosine similarity.

Pengujian dilakukan dengan memberikan beberapa variasi input berupa keterampilan (*skills*), lokasi, tingkat pengalaman, dan jenis pekerjaan. Evaluasi sistem dilakukan menggunakan metrik *precision* untuk mengukur tingkat relevansi hasil rekomendasi yang dihasilkan. Selain itu, nilai similarity juga dianalisis untuk melihat tingkat kedekatan antara query pengguna dengan hasil rekomendasi yang diberikan oleh sistem.

**Tabel 3.** Percobaan Menggunakan Data

Pengujian ke-	Skill yang diuji	Hasil Rekomendasi	Average Similarity
---------------	------------------	-------------------	--------------------

1	<i>Python Developer</i>	<p><b>Rank: 1</b>                  python developer                  Company: Talent Groups                  Location: New York, United States                  Experience: Mid-Senior level                  Work Type: FULL_TIME                  Similarity: 86.14%</p> <p><b>Rank: 2</b>                  python developer                  Company: Talent Groups                  Location: New York, United States                  Experience: Mid-Senior level                  Work Type: FULL_TIME                  Similarity: 84.40%</p> <p><b>Rank: 3</b>                  python developer                  Company: Tech Mahindra                  Location: United States                  Experience: Director                  Work Type: FULL_TIME                  Similarity: 74.13%</p> <p><b>Rank: 4</b>                  big data developer                  Company: The Judge Group                  Location: Jersey City, NJ                  Experience: Executive                  Work Type: FULL_TIME                  Similarity: 72.35%</p> <p><b>Rank: 5</b>                  python developer                  Company: The Judge Group                  Location: Jersey City, NJ                  Experience: Executive                  Work Type: FULL_TIME                  Similarity: 72.19%</p>	77,84%
2	<i>Designer</i>	<p><b>Rank: 1</b>                  technical designer                  Company: Fourth Floor                  Location: New York City Metropolitan Area                  Experience: Mid-Senior level                  Work Type: FULL_TIME                  Similarity: 74.35%</p> <p><b>Rank: 2</b>                  design engineer                  Company: Apex Systems                  Location: Atlanta, GA                  Experience: Entry level                  Work Type: FULL_TIME                  Similarity: 74.21%</p>	72,4%



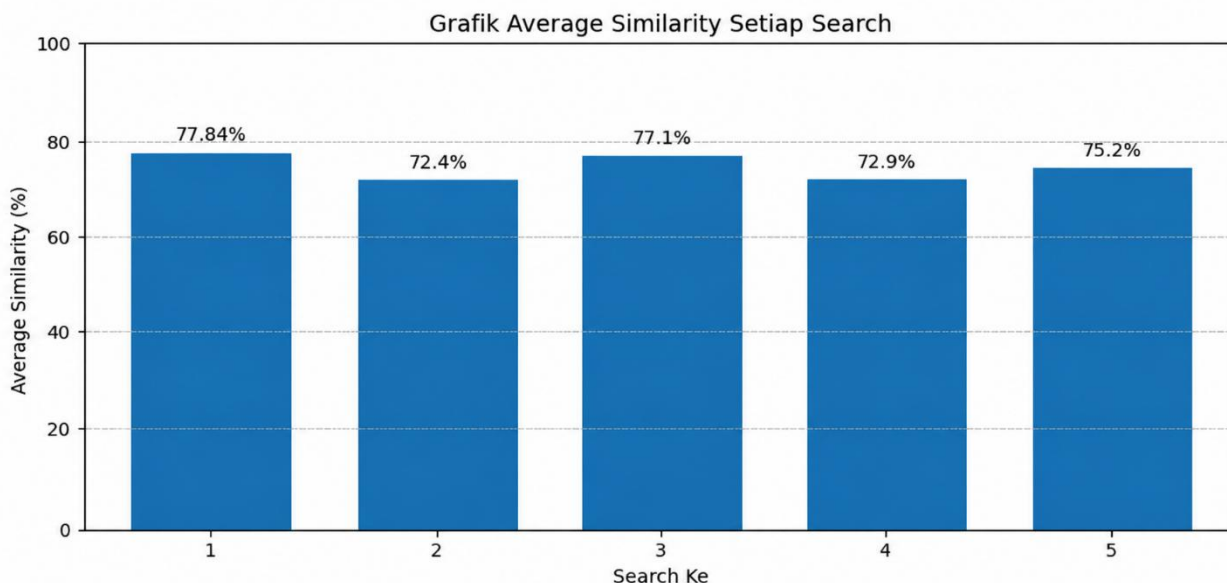
		<p><b>Rank: 3</b>  <b>design technician</b>  <b>Company:</b> FastTek Global  <b>Location:</b> Jackson, MI  <b>Experience:</b> Entry level  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 73.48%</p> <p><b>Rank: 4</b>  <b>technical writer/instructional designer(100% remote)</b>  <b>Company:</b> Prudent Technologies and Consulting, Inc.  <b>Location:</b> United States  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> CONTRACT  <b>Similarity:</b> 72.22%</p> <p><b>Rank: 5</b>  <b>technical designer</b>  <b>Company:</b> Torrid  <b>Location:</b> City of Industry, CA  <b>Experience:</b> Entry level  <b>Work Type:</b> INTERNSHIP  <b>Similarity:</b> 71.50%</p>	
3	<i>Registered Nurse</i>	<p><b>Rank: 1</b>  <b>registered nurse, radiology</b>  <b>Company:</b> UF Health  <b>Location:</b> Gainesville Metropolitan Area  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 78.26%</p> <p><b>Rank: 2</b>  <b>registered nurse progressive care unit full time days</b>  <b>Company:</b> Tenet Healthcare  <b>Location:</b> El Paso, TX  <b>Experience:</b> Associate  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 77.45%</p>	77,1%

		<p><b>Rank: 3</b>                  registered nurse - ft - resource team  <b>Company:</b> Methodist Health System  <b>Location:</b> Dallas, TX  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 76.79%</p> <p><b>Rank: 4</b>                  registered nurse (rn) supervisor agency free facility  <b>Company:</b> Life Care Centers of America  <b>Location:</b> North Charleston, SC  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> PART_TIME  <b>Similarity:</b> 76.55%</p> <p><b>Rank: 5</b>                  registered nurse (rn) i - med/surg  <b>Company:</b> Houston Methodist  <b>Location:</b> Houston, TX  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> CONTRACT  <b>Similarity:</b> 76.39%</p>	
4	Sales	<p><b>Rank: 1</b>                  sales engineer  <b>Company:</b> Adecco  <b>Location:</b> Waco, TX  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> INTERNSHIP  <b>Similarity:</b> 73.39%</p> <p><b>Rank: 2</b>                  sales team leader  <b>Company:</b> Millennium Recruiting, Inc.  <b>Location:</b> Houston, TX  <b>Experience:</b> Entry level  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 73.35%</p> <p><b>Rank: 3</b>                  sales specialist  <b>Company:</b> White Cap  <b>Location:</b> Parsippany, NJ  <b>Experience:</b> Associate  <b>Work Type:</b> INTERNSHIP  <b>Similarity:</b> 73.21%</p>	72,9%



		<p><b>Rank: 4</b>  <b>technical sales engineer</b>  <b>Company:</b> ASMPT SEMI Solutions  <b>Location:</b> Beaverton, OR  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> INTERNSHIP  <b>Similarity:</b> 72.25%</p> <p><b>Rank: 5</b>  <b>sales operations specialist</b>  <b>Company:</b> Beacon Hill Staffing Group  <b>Location:</b> Arlington, VA  <b>Experience:</b> Entry level  <b>Work Type:</b> INTERNSHIP  <b>Similarity:</b> 72.18%</p>	
5	<i>business intelligence</i>	<p><b>Rank: 1</b>  <b>is business intelligence developer</b>  <b>Company:</b> Ochsner Health  <b>Location:</b> New Orleans, LA  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 77.40%</p> <p><b>Rank: 2</b>  <b>business intelligence developer</b>  <b>Company:</b> CNA Search  <b>Location:</b> Atlanta, GA  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> INTERNSHIP  <b>Similarity:</b> 75.00%</p> <p><b>Rank: 3</b>  <b>financial business intelligence developer</b>  <b>Company:</b> POWER Engineers  <b>Location:</b> Boise, ID  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 74.89%</p> <p><b>Rank: 4</b>  <b>business intelligence specialist</b>  <b>Company:</b> Allied OneSource  <b>Location:</b> Chattanooga, TN  <b>Experience:</b> Entry level  <b>Work Type:</b> CONTRACT  <b>Similarity:</b> 74.49%</p> <p><b>Rank: 5</b>  <b>business intelligence developer</b>  <b>Company:</b> eNGINE  <b>Location:</b> Mount Pleasant, PA  <b>Experience:</b> Mid-Senior level  <b>Work Type:</b> FULL_TIME  <b>Similarity:</b> 74.33%</p>	75,2%





**Gambar 3.** Grafik Average Similarity 5 search

Berdasarkan hasil pengujian pada query "Python Developer", sistem rekomendasi berhasil menghasilkan lima lowongan pekerjaan dengan tingkat kemiripan (similarity) yang relatif tinggi terhadap kebutuhan pengguna. Nilai similarity masing-masing rekomendasi adalah 86,14%, 84,40%, 74,13%, 72,35%, dan 72,19%. Dari kelima hasil tersebut diperoleh Average Similarity sebesar 77,84%, yang menunjukkan bahwa secara rata-rata rekomendasi yang diberikan sistem memiliki tingkat kesesuaian semantik yang tinggi dengan query yang dimasukkan pengguna dapat dilihat pada Figure 1.

Nilai *Average Similarity* yang digunakan untuk 5 percobaan di atas dihitung dengan menjumlahkan seluruh nilai similarity dari hasil rekomendasi, kemudian membaginya dengan jumlah rekomendasi yang ditampilkan. Berikut adalah salah satu contoh perhitungannya :

$$Average\ Similarity = \frac{86.14 + 84.40 + 74.13 + 72.35 + 72.19}{5}$$

$$Average\ Similarity = 77.84\%$$

Hasil pengujian menunjukkan bahwa kombinasi metode *Sentence-BERT* (SBERT) sebagai pembentuk representasi semantik dan algoritma *K-Nearest Neighbors* (KNN) sebagai mekanisme pencarian mampu menghasilkan rekomendasi lowongan pekerjaan yang relevan berdasarkan kesamaan makna teks, bukan semata-mata berdasarkan kecocokan kata kunci. Untuk mengevaluasi tingkat relevansi rekomendasi yang dihasilkan, disusun ground truth secara semi-manual dengan mempertimbangkan kesesuaian antara query pengguna dan atribut lowongan pekerjaan, meliputi *job title*, *skills*, serta *experience level*. Lowongan pekerjaan dinyatakan relevan apabila memiliki hubungan semantik yang sesuai dengan kebutuhan pengguna. Kriteria dan hasil penentuan ground truth tersebut disajikan pada tabel berikut.

**Tabel 4.** Evaluasi Ground Truth dan Hasil Sistem

No	Query Pengguna	Kriteria Relevansi	Contoh Lowongan Relevan
1	Phyton Developer	Memiliki skill phyton, backend development, API, atau framework Phyton	Python Developer, Backend Engineer, Software Engineer

2	Designer	Memiliki keterampilan desain grafis, UI/UX, atau multimedia	UI/UX Designer, Graphic Designer
3	Registered Nurse	Berkaitan dengan pelayanan kesehatan dan keperawatan	Registered Nurse, Clinical Nurse
4	Sales	Berkaitan dengan penjualan, pemasaran, dan negosiasi	Sales Executive, Sales Representative
5	Bussiness Intelligence	Berkaitan dengan analisis data dan business reporting	BI Analyst, Data Analyst, Business Analyst

Tabel 5. Evaluasi Ground Truth

No	Query Pengguna	Jumlah Top-K Rekomendasi	Jumlah Lowongan relevan	Average Similarity
1	Phyton Developer	5	4	74,9%
2	Designer	5	4	72,4%
3	Registered Nurse	5	5	77,1%
4	Sales	5	4	72,9%
5	Bussiness Intelligence	5	4	75,2%

### KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem rekomendasi lowongan kerja berbasis semantic matching menggunakan *Sentence-BERT* (SBERT) dan algoritma *K-Nearest Neighbors* (KNN). Sistem dibangun dengan memanfaatkan representasi embedding semantik untuk memahami hubungan antara profil pengguna dan deskripsi pekerjaan secara kontekstual sehingga mampu menghasilkan rekomendasi lowongan kerja yang lebih relevan.

Berdasarkan hasil pengujian menggunakan beberapa skenario query pengguna, sistem mampu menghasilkan rekomendasi dengan nilai *average similarity* berkisar antara 72,4% hingga 77,1%. Hasil tersebut menunjukkan bahwa pendekatan SBERT dan KNN cukup efektif dalam memahami kesesuaian semantik antara kebutuhan pengguna dan informasi lowongan pekerjaan. Selain itu, penggunaan *semantic matching* mampu membantu meningkatkan relevansi rekomendasi dibandingkan pendekatan pencocokan kata kunci secara konvensional.

*Ground truth* yang dibentuk secara semi-manual juga menunjukkan bahwa sebagian besar hasil rekomendasi yang diberikan sistem termasuk ke dalam kategori relevan sesuai dengan kebutuhan pengguna. Dengan demikian, sistem yang dikembangkan dapat membantu proses pencarian kerja menjadi lebih efektif dan personal serta berpotensi mengurangi terjadinya job mismatch pada platform rekrutmen digital.

### DAFTAR PUSTAKA

- [1] Crismastiana Koloman, Raihan Maulana, Raisya Dwi Zahra Putri, dan Wahyu Abadi Harahap, "Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering," *Journal of Creative Student Research*, vol. 1, no. 6, hlm. 78–88, Des 2023, doi: 10.55606/jcsrpolitama.v1i6.2992.
- [2] Ali Said, "Tingkat Pengangguran Terbuka (TPT) sebesar 4,85 persen. Rata-rata upah buruh sebesar 3,33 juta rupiah.," Badan Pusat Statistik.
- [3] R. Ningsih, Y. Azhar, dan Y. Munarko, "Rekomendasi Lowongan Pekerjaan dari Portal Bursa Lowongan Kerja Memanfaatkan Cosine Similarity dan Simple Additive Weighting," *Jurnal Repositor*, vol. 2, no. 5, Jan 2024, doi: 10.22219/repositor.v2i5.30635.
- [4] Sagar Hedau, "LinkedIn's Impact on Job Hunting: Facts & Figures (2025)," LinkedIn.



- [5] Alya Fitria, Syahiduz Zaman, dan Muhammad Ainul Yaqin, "Sistem Rekomendasi Lowongan Pekerjaan Menggunakan Content-based filtering," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 10, no. 3, hlm. 1–7, Des 2024.
- [6] Z.-T. Yap, S.-C. Haw, dan N. E. Binti Ruslan, "Hybrid-based food recommender system utilizing KNN and SVD approaches," *Cogent Eng.*, vol. 11, no. 1, hlm. 2436125, Des 2024, doi: 10.1080/23311916.2024.2436125.
- [7] M. R. A. Zayyad, "SISTEM REKOMENDASI BUKU MENGGUNAKAN METODE CONTENT BASED FILTERING," 2021.
- [8] F. Y. A'la, "Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia: IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, hlm. 256–265, Apr 2025, doi: 10.29408/edumatic.v9i1.29666.
- [9] K. Wau, "Application of Fine-Tuned IndoBERT for Sentiment Classification Local Product Reviews on Tokopedia Marketplace with Limited Dataset," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 5, no. 1, hlm. 1377–1381, Okt 2025, doi: 10.59934/jaiea.v5i1.1629.
- [10] F. T. Sabilillah, S. Winarno, dan R. B. Abiyyi, "Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, hlm. 585–594, Des 2024, doi: 10.29408/edumatic.v8i2.27791.
- [11] S. Zhang dan J. Li, "KNN Classification With One-Step Computation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, hlm. 2711–2723, 2023, doi: 10.1109/TKDE.2021.3119140.
- [12] A. D. M. Putri, N. Sulistianingsih, dan R. Rismayati, "Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF terhadap Akurasi Klasifikasi Sentimen Teks Multi-Domain," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 7, no. 4, hlm. 675–688, Okt 2025, doi: 10.35746/jtim.v7i4.756.
- [13] N. Reimers dan I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," dalam *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, hlm. 3980–3990. doi: 10.18653/v1/D19-1410.
- [14] R. M. Holis, P. E. P. Utomo, dan B. F. Hutabarat, "Semantic FAQ Chatbot Using SBERT (Sentence-BERT) and Cosine Similarity for Academic Services," *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, hlm. 915–922, Okt 2025, doi: 10.47709/brilliance.v5i2.7027.
- [15] A. P. Putra, D. P. Singgih Putri, dan AA. Kt. A. C. Wiranatha, "Scientific Paper Recommendation System: Application of Sentence Transformers and Cosine Similarity Using arXiv Data," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, hlm. 1374–1382, Agu 2025, doi: 10.30871/jaic.v9i4.9766.

