



Analisis Kesesuaian Peta Okupasi Skema Pengembang Web dengan Kebutuhan Kompetensi Lowongan Kerja Online Menggunakan Cosine Similarity

Silvi Nurcahyaningih^{1*}, Mustafa²

¹⁻² Teknik Informatika, Universitas Islam Sultan Agung, Indonesia

*Penulis Korespondensi : silviinurcahyani22@gmail.com

Abstract. This study analyzes the compatibility between competencies in the National Occupation Map (SKKNI) of the ICT sector under the Web Developer scheme and competency requirements found in online job vacancies in Indonesia. The study employs Natural Language Processing (NLP) with Multilingual BERT (mBERT) to generate semantic representations of competency-related texts and Cosine Similarity to measure their semantic similarity. The dataset consists of 18 competency units, 58 competency elements, and 157 Performance Criteria (Kriteria Unjuk Kerja/KUK) from the National Occupation Map, along with 255 online job vacancies collected from Jobstreet Indonesia and LinkedIn. In the similarity analysis, all Performance Criteria and Competency Elements were aggregated into separate text representations and compared with the text representation of each job vacancy. The results show a mean similarity of 0.792932 for the aggregated KUK and 0.800890 for the aggregated Competency Elements, categorized as Suitable and Highly Suitable, respectively. The similarity values ranged from 0.646282 to 0.940792 for the aggregated KUK and from 0.639456 to 0.947454 for the aggregated Competency Elements. Word occurrence analysis further identified C, API, and Java as the three most frequently required competencies, appearing in 65.10%, 48.24%, and 45.10% of the analyzed vacancies, respectively. These results indicate that the competencies represented in the National Occupation Map generally show a high level of semantic compatibility with competency requirements in online job vacancies.

Keywords: Competency GAP; Cosine Similarity; Multilingual BERT; National Occupation Map; Online Job Vacancy.

Abstrak. Penelitian ini menganalisis tingkat kesesuaian kompetensi pada Peta Okupasi Nasional SKKNI sektor Teknologi Informasi dan Komunikasi (TIK) skema Pengembang Web dengan kebutuhan kompetensi yang terdapat pada lowongan kerja online di Indonesia. Penelitian menggunakan pendekatan Natural Language Processing (NLP) dengan Multilingual BERT (mBERT) sebagai representasi teks dan Cosine Similarity untuk mengukur kesamaan semantik. Dataset penelitian terdiri atas 18 unit kompetensi, 58 elemen kompetensi, dan 157 Kriteria Unjuk Kerja (KUK) pada Peta Okupasi serta 255 lowongan kerja online yang dikumpulkan dari Jobstreet Indonesia dan LinkedIn. Pada analisis similarity, seluruh KUK dan seluruh Elemen Kompetensi masing-masing digabungkan menjadi representasi teks, kemudian dibandingkan dengan setiap lowongan pekerjaan. Hasil penelitian menunjukkan nilai mean similarity sebesar 0,792932 untuk Gabungan KUK dan 0,800890 untuk Gabungan Elemen Kompetensi, yang masing-masing termasuk kategori Sesuai dan Sangat Sesuai. Statistik deskriptif menunjukkan bahwa nilai similarity relatif homogen dengan standar deviasi masing-masing sebesar 0,067331 dan 0,071692. Analisis *word occurrence* menunjukkan bahwa C, API, dan Java merupakan tiga kompetensi yang paling sering dibutuhkan, masing-masing muncul pada 65,10%, 48,24%, dan 45,10% dari keseluruhan lowongan yang dianalisis. Hasil tersebut menunjukkan bahwa kompetensi pada Peta Okupasi Nasional SKKNI secara umum telah selaras dengan kebutuhan industri, meskipun masih terdapat lowongan dengan tingkat kesesuaian yang lebih rendah yang dapat menjadi bahan evaluasi.

Kata kunci: Cosine Similarity; Kesenjangan Kompetensi; Lowongan Kerja Online; Multilingual BERT; Peta Okupasi Nasional.

1. LATAR BELAKANG

Kesenjangan antara kompetensi lulusan perguruan tinggi dan kebutuhan industri masih menjadi persoalan, terutama pada bidang teknologi informasi yang berkembang pesat. Perkembangan teknologi mengubah jenis pekerjaan dan keterampilan yang dibutuhkan industri (Barrera dkk., 2022), sementara pendidikan di Indonesia masih cenderung berfokus pada aspek

teori sehingga lulusan belum sepenuhnya siap menghadapi dunia kerja (Ardhana dkk., 2025). Pemerintah melalui Kominfo dan Komdigi telah menyusun Peta Okupasi Nasional Sektor TIK sebagai acuan kompetensi, termasuk skema Pengembang Web. Namun, kebutuhan kompetensi industri terus berubah sehingga perlu dianalisis kesesuaiannya dengan kebutuhan yang tercermin pada lowongan kerja online.

Data lowongan kerja telah digunakan dalam penelitian untuk mengidentifikasi kebutuhan kompetensi industri (Ula & Faisal, 2025) (Yolanda Reza dkk., 2025). Dari sisi metode, mBERT dapat merepresentasikan teks multibahasa (Kenton dkk., 2019) (Chanif dkk., 2026), sedangkan Cosine Similarity digunakan untuk mengukur kemiripan antar teks (Hasibuan dkk., 2023) (No & Ferris Tita Sabilillah, Sri Winarto, 2024). Pendekatan tersebut dapat digunakan untuk membandingkan kompetensi pada Peta Okupasi dengan kebutuhan kompetensi dalam lowongan kerja.

Penelitian yang secara khusus mengukur kesesuaian Peta Okupasi Nasional SKKNI sektor TIK skema Pengembang Web dengan lowongan kerja online di Indonesia masih terbatas. Penelitian ini mengisi celah tersebut dengan menggabungkan mBERT, Cosine Similarity, statistik deskriptif, dan *Word Occurrence*. Penelitian bertujuan menganalisis tingkat kesesuaian kompetensi Peta Okupasi dengan kebutuhan industri serta mengidentifikasi kompetensi yang paling sering muncul pada lowongan kerja.

2. KAJIAN TEORITIS

Ketimpangan antara kompetensi lulusan perguruan tinggi dan kebutuhan dunia kerja (skill gap) masih menjadi persoalan, khususnya pada bidang teknologi informasi yang berkembang cepat. Perubahan teknologi turut mengubah jenis pekerjaan dan keterampilan yang dibutuhkan industri (Barrera dkk., 2022). Di Indonesia, lulusan juga masih menghadapi kesulitan beradaptasi karena kompetensi yang diperoleh belum sepenuhnya sesuai dengan kebutuhan praktis industri (Ardhana dkk., 2025) (Ayofe & Ajetola, 2021). Kondisi tersebut menunjukkan pentingnya mengukur kesesuaian kompetensi pendidikan dengan kebutuhan dunia kerja.

Pemerintah melalui Kominfo/Komdigi menyusun Peta Okupasi Nasional Sektor TIK berdasarkan Standar Kompetensi Kerja Nasional Indonesia (SKKNI) sebagai acuan kompetensi tenaga kerja, termasuk skema Pengembang Web (Karim, 2024). Namun, kebutuhan industri bersifat dinamis dan dapat berubah mengikuti perkembangan teknologi serta kebutuhan perusahaan (Thornhill-Miller dkk., 2023) (Karakolis dkk., 2022). Oleh karena

itu, kesesuaian antara standar kompetensi dan kebutuhan industri perlu diuji berdasarkan data aktual.

Salah satu sumber yang dapat digunakan adalah data lowongan kerja online karena secara langsung mencerminkan kompetensi yang dicari perusahaan (Ula & Faisal, 2025) (Quintini & Verhagen, 2022). Beberapa penelitian juga telah memanfaatkan data lowongan kerja untuk menganalisis kebutuhan keterampilan dan membandingkannya dengan kompetensi akademik menggunakan semantic similarity (Yolanda Reza dkk., 2025) (Mandende dkk., 2024).

Perbandingan antara Peta Okupasi dan lowongan kerja memerlukan metode yang mampu mengolah teks dan menangkap hubungan semantik. Natural Language Processing (NLP) digunakan untuk mengolah data teks, sedangkan Multilingual BERT (mBERT) dapat menghasilkan representasi teks dengan mempertimbangkan konteks dan mendukung data multibahasa (Keraghel dkk., 2024) (Wilie dkk., 2020). *Embedding* yang dihasilkan mBERT selanjutnya dibandingkan menggunakan Cosine Similarity untuk mengukur kedekatan semantik antar teks. Pendekatan tersebut telah digunakan dalam berbagai penelitian pencocokan teks dan analisis keterampilan (Ahadi dkk., 2022) (Hasibuan dkk., 2023) (Javadian Sabet dkk., 2024).

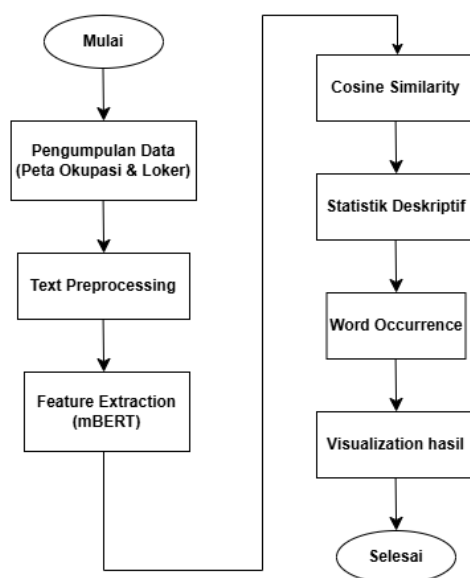
Selain mengukur kesesuaian semantik, penelitian ini menggunakan statistik deskriptif untuk menggambarkan karakteristik nilai similarity melalui mean, median, minimum, maksimum, dan standar deviasi (Ariyanta dkk., 2025). Analisis *Word Occurrence* juga digunakan untuk mengetahui kompetensi yang paling sering muncul pada lowongan kerja berdasarkan keberadaan kata kunci kompetensi pada setiap lowongan.

Meskipun penelitian mengenai skill gap, NLP, BERT/mBERT, dan Cosine Similarity telah banyak dilakukan, penelitian yang secara khusus mengukur kesesuaian Peta Okupasi Nasional SKKNI sektor TIK skema Pengembang Web dengan kebutuhan kompetensi pada lowongan kerja online di Indonesia masih terbatas. Penelitian ini mengisi celah tersebut dengan menggabungkan mBERT, Cosine Similarity, statistik deskriptif, dan *Word Occurrence* untuk menganalisis kesesuaian kompetensi sekaligus mengidentifikasi kompetensi yang paling sering dibutuhkan industri.

3. METODE PENELITIAN

Penelitian ini dilakukan dengan pendekatan text mining dan Natural Language Processing (NLP) untuk membandingkan dua sumber data, yaitu Peta Okupasi Nasional SKKNI sektot TIK skema Pengembang Web dan data lowongan kerja online. Text mining

secara umum mencakup tahapan pengumpulan data, *text preprocessing*, ekstraksi fitur, dan analisis pola, sedangkan NLP dipakai supaya komputer dapat memahami dan mengolah bahasa manusia secara otomatis, mulai dari tahap preprocessing samapai analisis semantik. Tahapan penelitian secara keseluruhan digambarkan pada Gambar 1.



Gambar 1. Alur Penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian ini terdiri dari dua sumber utama. Sumber pertama adalah Peta Okupasi Nasional dan SKKNI sektor TIK untuk skema Pengembang Web, yang terdiri atas 18 unit kompetensi, 58 Elemen Kompetensi, dan 157 Kriteria Unjuk Kerja (KUK). Untuk keperluan analisis, seluruh KUK digabungkan menjadi satu teks yang disebut Gabungan KUK, sedangkan seluruh Elemen Kompetensi digabungkan menjadi Gabungan Elemen Kompetensi. Sumber kedua berupa 255 data lowongan kerja yang diperoleh secara manual dari Jobstreet Indonesia dan LinkedIn. Data tersebut mencakup beberapa atribut, yaitu judul lowongan, lokasi, persyaratan, deskripsi, tautan, tanggal, serta kompetensi yang diekstraksi dari masing-masing lowongan. Dalam proses pengukuran similarity, bagian persyaratan dan deskripsi pada setiap lowongan digabungkan menjadi satu teks representatif. Selanjutnya, seluruh data disimpan dalam format .csv untuk digunakan pada tahap pengolahan dan analisis.

Text Preprocessing

Karena mBERT merupakan model transformer yang telah dilatih pada korpus multibahasa dalam jumlah besar dan memiliki mekanisme tokenisasi internal sendiri, proses pembersihan teks pada penelitian ini sengaja tidak dilakukan secara berlebihan. Preprocessing

difokuskan pada penghapusan URL dan spasi berlebih tanpa mengubah struktur kalimat secara drastis. Stemming, stopword removal, dan tokenisasi manual sengaja tidak diterapkan karena berpotensi menghilangkan informasi kontekstual yang justru dibutuhkan mBERT untuk menghasilkan representasi semantik yang optimal.

Representasi Text Menggunakan Multilingual BERT (mBERT)

Setelah preprocessing, teks kompetensi direpresentasikan menggunakan model bert-base-multilingual-cased dari pustaka Transformers (*HuggingFace*). Pada Peta Okupasi, seluruh KUK digabungkan menjadi satu teks representatif dan seluruh Elemen Kompetensi digabungkan menjadi satu teks representatif. Pada data lowongan, bagian persyaratan dan deskripsi pekerjaan digabungkan menjadi satu teks untuk setiap lowongan. Representasi kalimat diperoleh menggunakan mean pooling terhadap output token dengan mempertimbangkan attention mask. Untuk teks yang melebihi batas maksimum input mBERT, teks dibagi menjadi beberapa chunk berukuran maksimal 510 token, kemudian *embedding* setiap chunk digabungkan menggunakan weighted average berdasarkan jumlah token. Hasil *embedding* kemudian dilakukan L2 normalization untuk menormalkan panjang vektor. *Embedding* yang dihasilkan selanjutnya digunakan dalam perhitungan Cosine Similarity.

Pengukuran Kesesuaian dengan Cosine Similarity

Tingkat kesesuaian antara Peta Okupasi dan lowongan pekerjaan dihitung menggunakan Cosine Similarity. Seluruh KUK dan Elemen Kompetensi masing-masing digabungkan menjadi Gabungan KUK dan Gabungan Elemen Kompetensi, kemudian diubah menjadi *embedding* vector menggunakan mBERT. Kedua representasi tersebut dibandingkan dengan *embedding* dari masing-masing 255 lowongan berdasarkan Cosine Similarity, sehingga diperoleh satu nilai similarity untuk setiap representasi pada setiap lowongan. Nilai yang mendekati 1 menunjukkan kedekatan semantik yang semakin tinggi.

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \times \|B\|}$$

di mana A dan B adalah vektor *embedding* mBERT dari dua dokumen, A.B adalah hasil dot product antar vektor, sedangkan $\|A\|$ dan $\|B\|$ adalah panjang (norma) masing-masing vektor. Nilai similarity berkisar dari 0 (tidak mirip sama sekali) sampai 1 (identik).

Analisis Statistik Deskriptif

Nilai similarity yang diperoleh dari perbandingan Gabungan KUK dan Gabungan Elemen Kompetensi dengan 255 data lowongan pekerjaan selanjutnya dianalisis menggunakan statistik deskriptif. Ukuran statistik yang digunakan meliputi mean, median, minimum, maksimum, dan standar deviasi. Analisis ini bertujuan untuk memberikan gambaran mengenai

kecenderungan, rentang, dan penyebaran nilai similarity secara keseluruhan. Nilai mean digunakan sebagai indikator tingkat kesesuaian rata-rata antara representasi kompetensi pada Peta Okupasi dan kebutuhan kompetensi pada data lowongan pekerjaan.

Nilai similarity kemudian diinterpretasikan secara bertingkat dengan mengadaptasi interval pada pendekatan multi-grade classification yang digunakan oleh (Neogi dkk., 2012). Dalam penelitian ini, interval 0,40, 0,60, dan 0,80 dioperasikan menjadi empat kategori, yaitu Sangat Sesuai ($\geq 0,80$), Sesuai ($0,60 < 0,80$), Cukup Sesuai ($0,40 < 0,60$), dan Kurang Sesuai ($< 0,40$). Pemberian label kategori tersebut merupakan operasionalisasi penelitian yang disesuaikan dengan tujuan analisis kesesuaian kompetensi antara Peta Okupasi dan kebutuhan industri.

Analisis Word Occurrence

Selain mengukur kesesuaian lewat cosine similarity, penelitian ini juga melakukan analisis *word occurrence* pada kolom kompetensi lowongan untuk mengidentifikasi kompetensi yang paling banyak diminta industri. Setiap kompetensi pada satu lowongan hanya dihitung satu kali, sehingga frekuensi yang diperoleh menunjukkan jumlah lowongan yang membutuhkan kompetensi tersebut, bukan jumlah pengulangan kata dalam satu dokumen. Semakin tinggi frekuensinya, semakin tinggi pula tingkat urgensi kompetensi tersebut di dunia kerja.

4. HASIL DAN PEMBAHASAN

Pengumpulan Data

Proses pengambilan data dilakukan dalam dua tahap yang berjalan berdampingan. Data peta Okupasi Nasional SKKNI sektor TIK skema Pengembang Web diambil langsung dari dokumen resmi Peta Okupasi Nasional, sedangkan data lowongan kerja dikumpulkan secara manual dari platform Jobstreet Indonesia dan LinkedIn dengan menyaring lowongan yang relevan dengan bidang pengembang web. Kedua data tersebut kemudian dirapikan dan disimpan dalam format .csv agar siap diproses pada tahap *text preprocessing*, representasi teks, hingga perhitungan kesesuaian.

Dataset yang digunakan dalam penelitian ini terdiri atas dua kelompok data dengan jumlah lowongan yang berbeda, sebagaimana dirangkum pada Tabel 1.

Tabel 1. Ringkasan Dataset Penelitian

No	Dataset	Sumber	Jumlah Data
1	Peta Okupasi Nasional SKKNI	Peta Okupasi Nasional SKKNI Sektor TIK	18 unit, 58 elemen kompetensi, 157 KUK
2	Lowongan Kerja Online	Jobstreet Indonesia dan LinkedIn	255 lowongan kerja

Berdasarkan Tabel 1, Kedua dataset ini saling melengkapi. Peta Okupasi berperan sebagai tolok ukur standar kompetensi, sedangkan data lowongan kerja mencerminkan kompetensi yang benar-benar dicari perusahaan di lapangan. Dataset lowongan kerja sendiri didominasi oleh posisi-posisi di bidang pengembangan perangkat lunak dan pengembangan web, dengan variasi kualifikasi mulai dari fresh graduate sampai posisi yang mensyaratkan pengalaman kerja tertentu.

Hasil Text Preprocessing

Sebelum dibandingkan, kedua dataset melalui tahap *text preprocessing* berupa penghapusan URL, simbol yang tidak relevan, dan spasi berlebih tanpa proses stemming atau stopword removal, karena mBERT sudah memiliki tokenizer otomatis yang justru membutuhkan struktur kalimat asli untuk membentuk representasi semantik yang baik. Hasil preprocessing menunjukkan istilah teknis seperti nama bahasa pemrograman dan tools tetap dipertahankan, sementara elemen yang tidak bermakna seperti tautan dan simbol berhasil dihilangkan. Contoh preprocessing ditampilkan pada Tabel 2.

Tabel 2. Contoh Hasil Preprocessing

No	Tahapan	Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
1	Menghapus URL	https://id.jobstreet.com/id/job/88675853?type=standard	(kosong/URL dihapus)
2	Menghapus spasi berlebih	Laravel MySQL	Laravel MySQL

Berdasarkan Tabel 2, proses preprocessing berhasil menghapus URL dan spasi berlebih, sehingga teks menjadi lebih bersih. Istilah-istilah teknis seperti nama bahasa pemrograman dan framework tetap dipertahankan apa adanya, sebab istilah tersebut justru menjadi informasi utama yang perlu ditangkap mBERT pada tahap *embedding*.

Hasil Representasi Teks mBERT

Setelah proses preprocessing, data kompetensi dari kedua sumber diubah menjadi vektor numerik menggunakan mBERT. Konfigurasi model yang dipakai ditampilkan pada Tabel 3.

Tabel 3. Konfigurasi Model mBERT

Parameter	Nilai
Model	bert-base-multilingual-cased
Metode Representasi	<i>Sentence Embedding</i> (Mean Pooling)
Jenis Model	Multilingual BERT (mBERT)
Bahasa	Multibahasa (Indonesia dan Inggris)
Output	<i>Embedding Vector</i>

```
print("\n=====")
print("=== Contoh Embedding Lowongan ===")
print("Teks :")
print(lowongan["loker"].iloc[0])
print("\nEmbedding (5 dimensi pertama):")
print(embedding_lowongan[0][1:5])

---
--- Contoh Embedding Lowongan ---
Teks :
Persyaratan: a. Lulusan baru dari semua jurusan, jurusan Ilmu Komputer, Teknologi Informasi, Teknik Elekt

Embedding (5 dimensi pertama):
tensor([-0.0085, -0.0374, -0.0197,  0.0068,  0.0392])
```

Gambar 2. Contoh Hasil *Embedding* Teks Peta Okupasi dan Lowongan Kerja

Gambar 2 memperlihatkan cuplikan hasil *embedding* dari salah satu deskripsi lowongan kerja, ditampilkan berupa lima dimensi pertama dari total ratusan dimensi vektor yang dihasilkan mBERT. Setiap dokumen teks, baik dari Peta Okupasi maupun lowongan kerja, diubah menjadi satu vektor utuh yang merepresentasikan makna semantiknya secara keseluruhan, dan vektor inilah yang selanjutnya dipakai pada proses perhitungan cosine similarity.

Hasil Analisis Cosine Similarity

Pada tahap ini dilakukan pengukuran tingkat kesesuaian antara representasi kompetensi pada Peta Okupasi dan data lowongan pekerjaan menggunakan Cosine Similarity. Seluruh KUK dan seluruh Elemen Kompetensi terlebih dahulu digabungkan menjadi masing-masing satu representasi teks. Representasi tersebut kemudian diubah menjadi *embedding* vector menggunakan mBERT dan dibandingkan dengan *embedding* setiap lowongan pekerjaan. Dengan demikian, setiap lowongan memiliki nilai Similarity Gabungan KUK dan Similarity Gabungan Elemen Kompetensi.

Tabel 4. Contoh Hasil Mean Similarity per Kriteria Unjuk Kerja

Judul Lowongan	Lokasi	Similarity Gabungan KUK	Similarity Gabungan Elemen
IT Operational Engginer	Cipondoh, Banten	0.940792	0.947454
Frontend Developer	Surabaya, Jawa Timur	0.927907	0.934228
Full Stack Developer	Surabaya, Jawa Timur	0.926619	0.933458
IT Project Management	Jakarta Selatan, Jakarta Raya	0.923691	0.928954
Full Stack Developer	Jakarta Utara, Jakarta Raya	0.922417	0.927849

Terlihat bahwa nilai mean similarity antar KUK cukup bervariasi meskipun berasal dari unit kompetensi yang sama. Semakin tinggi nilainya, semakin kuat kemiripan makna KUK tersebut dengan kompetensi-kompetensi yang diminta pada lowongan kerja yang dianalisis. Sepuluh KUK dengan nilai similarity tertinggi ditampilkan pada Gambar 3.



Gambar 3. Sepuluh Kriteria Unjuk Kerja dengan Mean Similarity Tertinggi

Berdasarkan hasil tersebut, terlihat bahwa nilai similarity antara Gabungan KUK dengan setiap lowongan pekerjaan memiliki variasi. Lowongan dengan nilai similarity tertinggi menunjukkan tingkat kedekatan semantik yang lebih tinggi terhadap keseluruhan Kriteria Unjuk Kerja (KUK) pada Peta Okupasi. Hal ini menunjukkan bahwa semakin tinggi nilai similarity, semakin dekat representasi semantik teks lowongan pekerjaan dengan Gabungan KUK berdasarkan *embedding* mBERT dan pengukuran Cosine Similarity.

Distribusi Tingkat Kesesuaian Lowongan

Berdasarkan hasil pengukuran terhadap 255 lowongan pekerjaan, nilai similarity dikelompokkan menjadi empat kategori, yaitu Sangat Sesuai ($\geq 0,80$), Sesuai ($0,60 < 0,80$), Cukup Sesuai ($0,40 < 0,60$), dan Kurang Sesuai ($< 0,40$). Hasil pengelompokan ditunjukkan pada Tabel 5.

Tabel 5. Distribusi Tingkat Kesesuaian

Tingkat Kesesuaian	Gabungan KUK	Gabungan Elemen Kompetensi
Sangat Sesuai ($\geq 0,80$)	141	142
Sesuai ($0,60 < 0,80$)	114	113
Cukup Sesuai ($0,40 < 0,60$)	0	0
Kurang Sesuai ($< 0,40$)	0	0

Hasil menunjukkan bahwa 141 lowongan pada Gabungan KUK dan 142 pada Gabungan Elemen Kompetensi termasuk kategori Sangat Sesuai, sedangkan 114 dan 113 lowongan masing-masing termasuk kategori Sesuai. Tidak terdapat lowongan pada kategori Cukup Sesuai maupun Kurang Sesuai, sehingga seluruh lowongan memiliki nilai similarity minimal 0,60.

Perbedaan nilai similarity menunjukkan adanya perbedaan kedekatan semantik antara isi lowongan dan representasi kompetensi Pengembang Web. Junior Backend Engineer memiliki nilai relatif rendah, yaitu 0,646282 pada Gabungan KUK dan 0,663661 pada Gabungan Elemen Kompetensi. Lowongan tersebut lebih berfokus pada Node.js, ExpressJS, basis data, REST API, struktur data dan algoritma, serta pengujian, sehingga cakupannya lebih spesifik dibandingkan representasi keseluruhan kompetensi Pengembang Web.

Sebaliknya, IT Operational Engineer memperoleh nilai tertinggi, yaitu 0,940792 pada Gabungan KUK dan 0,947454 pada Gabungan Elemen Kompetensi. Isi lowongan mencakup berbagai aspek pengembangan perangkat lunak dan web sehingga memiliki kedekatan semantik yang lebih tinggi. Nilai similarity dihitung berdasarkan gabungan teks Persyaratan dan Deskripsi, bukan hanya judul lowongan.

Hasil Statistik Deskriptif

Nilai similarity dari Gabungan KUK dan Gabungan Elemen Kompetensi yang dibandingkan dengan 255 lowongan pekerjaan selanjutnya dianalisis menggunakan statistik deskriptif. Hasil analisis ditampilkan pada Tabel 5.

Tabel 6. Hasil Statistik Deskriptif

Statistik	Gabungan KUK	Gabungan Elemen
Mean	0.792932	0.800890
Median	0.814932	0.823985
Minimum	0.646282	0.639456
Maximum	0.940792	0.947454
Standar Deviasi	0.067331	0.071692



```
display(hasil_statistik)
```

	Statistik	Gabungan KUK	Gabungan Elemen
0	Mean	0.792932	0.800890
1	Median	0.814932	0.823985
2	Minimum	0.646282	0.639456
3	Maximum	0.940792	0.947454
4	Std Deviasi	0.067331	0.071692

Gambar 4. Output Perhitungan Statistik Deskriptif

Nilai mean Gabungan KUK sebesar 0,792932 menunjukkan kedekatan semantik yang sesuai terhadap keseluruhan KUK, sedangkan Gabungan Elemen sebesar 0,800890 menunjukkan kedekatan yang sangat sesuai terhadap keseluruhan Elemen Kompetensi. Nilai median masing-masing 0,814932 dan 0,823985 yang lebih tinggi daripada mean menunjukkan bahwa sebagian besar nilai similarity berada pada tingkat relatif tinggi.

Nilai similarity Gabungan KUK memiliki rentang 0,646282–0,940792, sedangkan Gabungan Elemen 0,639456–0,947454. Standar deviasi masing-masing sebesar 0,067331 dan 0,071692, yang menunjukkan penyebaran nilai relatif dekat dengan rata-rata. Secara umum, hasil tersebut menunjukkan kedekatan semantik yang cukup tinggi antara kebutuhan kompetensi pada lowongan kerja dan kompetensi dalam Peta Okupasi..

```

mean_kuk = hasil_statistik[
    hasil_statistik["Statistik"] == "Mean"
][ "Gabungan KUK" ].iloc[0]

if mean_kuk >= 0.80:
    kategori_kuk = "Sangat Sesuai"
elif mean_kuk >= 0.60:
    kategori_kuk = "Sesuai"
elif mean_kuk >= 0.40:
    kategori_kuk = "Cukup Sesuai"
else:
    kategori_kuk = "Kurang Sesuai"

print("Mean Similarity Gabungan KUK :", mean_kuk)
print("Kategori Gabungan KUK      :", kategori_kuk)

*** Mean Similarity Gabungan KUK : 0.7929316
    Kategori Gabungan KUK      : Sesuai

```

Gambar 5. Proses Penkategorian Tingkat Kesesuaian

Berdasarkan ambang batas kategori yang digunakan dalam penelitian ini, yaitu Sangat Sesuai ($\geq 0,80$), Sesuai ($0,60 < 0,80$), Cukup Sesuai ($0,40 < 0,60$), dan Kurang Sesuai ($< 0,40$), nilai mean similarity Gabungan KUK sebesar 0,792932 termasuk kategori “Sesuai”, sedangkan nilai mean similarity Gabungan Elemen Kompetensi sebesar 0,800890 termasuk kategori “Sangat Sesuai”. Hasil tersebut menunjukkan bahwa secara umum kompetensi pada Peta Okupasi Nasional SKKNI skema Pengembang Web memiliki tingkat kesesuaian semantik yang tinggi terhadap kebutuhan kompetensi pada data lowongan kerja yang dianalisis.

Hasil Analisis Word Occurrence

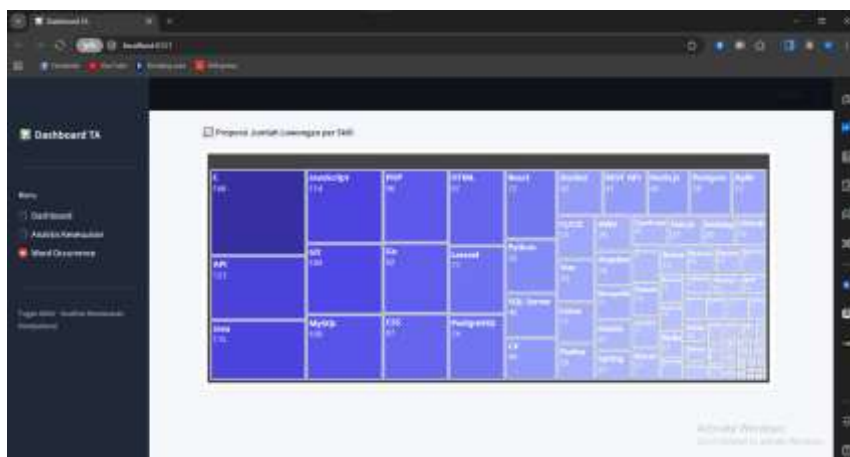
Selain mengukur kesesuaian lewat cosine similarity, penelitian ini juga menghitung frekuensi kemunculan tiap kompetensi pada data lowongan kerja untuk melihat kompetensi mana yang paling banyak dicari industri. Dari 78 kompetensi hasil ekstraksi, sepuluh kompetensi dengan frekuensi tertinggi ditampilkan pada Tabel 6.

Tabel 7. Sepuluh Kompetensi dengan Jumlah Lowongan Tertinggi (*Word Occurrence*)

No	Kompetensi	Jumlah Lowongan	Persentase
1	C	166	65,10%
2	API	123	48,24%
3	Java	115	45,10%
4	JavaScript	114	44,71%
5	Git	109	42,75%
6	MySQL	100	39,22%
7	PHP	96	37,65%
8	Go	92	36,08%
9	HTML	87	34,12%
10	CSS	87	34,12%

Kompetensi C menjadi kompetensi yang paling banyak muncul, ditemukan pada 166 dari 255 lowongan kerja yang dianalisis atau setara 65,10%. Posisi berikutnya ditempati API (48,24%) dan Java (45,10%). Menariknya, kompetensi dengan frekuensi tertinggi pada *word occurrence* ini tidak selalu sama dengan kompetensi yang memiliki nilai cosine similarity tertinggi pada Peta Okupasi, karena kedua analisis ini mengukur hal yang berbeda: cosine

similarity melihat kedekatan makna antara uraian KUK dan deskripsi lowongan secara keseluruhan, sedangkan *word occurrence* murni menghitung seberapa sering istilah kompetensi tertentu disebut secara eksplisit.

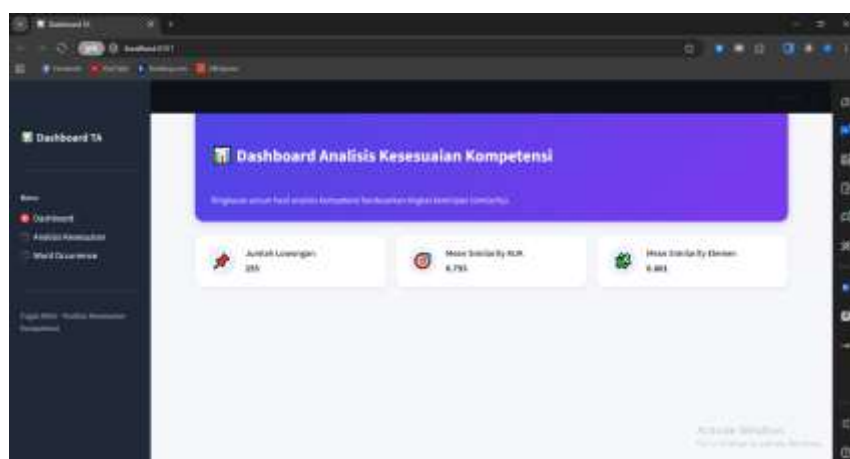


Gambar 6. Proporsi Jumlah Lowongan per Kompetensi dalam Bentuk Treemap

Gambar 6 menampilkan visualisasi *Word Occurrence* dalam bentuk treemap, dengan ukuran dan warna kotak menunjukkan jumlah lowongan yang mencantumkan kompetensi. Kompetensi yang muncul meliputi bahasa pemrograman dan tools seperti C, API, Java, JavaScript, Git, dan MySQL, serta framework, basis data, dan layanan cloud seperti Laravel, React, PostgreSQL, MongoDB, AWS, Azure, dan GCP dengan frekuensi yang lebih rendah.

Implementasi Dashboard

Supaya hasil penelitian ini lebih mudah dipakai oleh pihak-pihak terkait, seluruh hasil analisis divisualisasikan lewat dashboard interaktif berbasis Streamlit yang membaca langsung berkas hasil perhitungan similarity dan *word occurrence*. Dashboard ini terbagi menjadi tiga menu, yaitu Dashboard, Analisis Kesesuaian, dan *Word Occurrence*.



Gambar 7. Tampilan Utama Dashboard Analisis Kesesuaian Kompetensi

Menu Dashboard menampilkan ringkasan umum hasil analisis berupa jumlah lowongan sebanyak 255, nilai mean similarity Gabungan KUK sebesar 0,793, dan nilai mean similarity Gabungan Elemen Kompetensi sebesar 0,801.

Pembahasan

Hasil penelitian menunjukkan bahwa kompetensi pada Peta Okupasi Nasional SKKNI skema Pengembang Web memiliki kedekatan semantik dengan kebutuhan kompetensi pada 255 lowongan kerja yang dianalisis. Nilai mean similarity sebesar 0,792932 pada Gabungan KUK termasuk kategori Sesuai, sedangkan nilai 0,800890 pada Gabungan Elemen Kompetensi termasuk kategori Sangat Sesuai. Distribusi nilai juga menunjukkan bahwa seluruh lowongan memiliki similarity minimal 0,60 pada kedua pengukuran.

Perbedaan nilai similarity antar-lowongan menunjukkan adanya perbedaan kedekatan semantik berdasarkan isi persyaratan dan deskripsi pekerjaan. Lowongan dengan nilai relatif lebih rendah cenderung memiliki cakupan kompetensi yang lebih spesifik atau informasi kompetensi yang lebih terbatas dibandingkan representasi keseluruhan kompetensi Pengembang Web. Sebagai contoh, Junior Backend Engineer memperoleh nilai 0,646282 pada Gabungan KUK dan 0,663661 pada Gabungan Elemen Kompetensi dengan fokus pada Node.js, ExpressJS, basis data, REST API, struktur data dan algoritma, serta pengujian. Sebaliknya, IT Operational Engineer memperoleh nilai tertinggi, yaitu 0,940792 dan 0,947454, dengan persyaratan dan deskripsi yang mencakup berbagai aspek pengembangan perangkat lunak dan web. Nilai similarity dalam penelitian ini dihitung dari gabungan Persyaratan dan Deskripsi, bukan hanya berdasarkan judul lowongan.

Hasil *Word Occurrence* melengkapi pengukuran semantik dengan menunjukkan kompetensi yang paling sering disebutkan dalam lowongan. Kompetensi C, API, dan Java merupakan tiga kompetensi dengan frekuensi tertinggi. Perbedaan antara hasil Cosine Similarity dan *Word Occurrence* menunjukkan bahwa kedua analisis mengukur aspek yang berbeda. Cosine Similarity mengukur kedekatan makna teks secara keseluruhan, sedangkan *Word Occurrence* menghitung frekuensi kemunculan kompetensi secara eksplisit.

Secara keseluruhan, kombinasi kedua pendekatan memberikan gambaran yang lebih lengkap mengenai kesesuaian kompetensi dan kebutuhan kompetensi pada lowongan kerja. Namun, penelitian ini memiliki keterbatasan pada jumlah dan sumber data yang berasal dari 255 lowongan pada dua platform. Selain itu, penggabungan seluruh KUK dan Elemen Kompetensi menjadi satu representasi menyebabkan hasil similarity menggambarkan kesesuaian secara keseluruhan dan belum dapat menunjukkan kompetensi atau unit kompetensi tertentu yang paling sesuai dengan setiap lowongan. Penelitian selanjutnya dapat memperluas

sumber data dan membandingkan mBERT dengan model *embedding* lain seperti IndoBERT atau Sentence-BERT.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian mengenai analisis kesesuaian Peta Okupasi Nasional SKKNI sektor TIK skema Pengembang Web dengan kebutuhan kompetensi lowongan kerja online, dapat disimpulkan beberapa hal berikut.

Pendekatan NLP yang menggabungkan Multilingual BERT (mBERT) sebagai representasi teks dan Cosine Similarity sebagai pengukur kesesuaian berhasil diterapkan untuk mengukur tingkat kesesuaian antara Peta Okupasi Nasional SKKNI skema Pengembang Web dengan kebutuhan kompetensi pada 255 lowongan kerja online. Hasil perhitungan menunjukkan nilai mean similarity sebesar 0,792932 untuk Gabungan KUK yang termasuk kategori Sesuai, sedangkan Gabungan Elemen Kompetensi memperoleh mean similarity sebesar 0,800890 yang termasuk kategori Sangat Sesuai.

Analisis statistik deskriptif menunjukkan nilai median sebesar 0,814932 untuk Gabungan KUK dan 0,823985 untuk Gabungan Elemen Kompetensi. Nilai minimum masing-masing sebesar 0,646282 dan 0,639456, sedangkan nilai maksimum sebesar 0,940792 dan 0,947454. Standar deviasi masing-masing sebesar 0,067331 dan 0,071692, yang menunjukkan bahwa nilai similarity relatif homogen dan sebagian besar berada di sekitar nilai rata-ratanya.

Berdasarkan pengelompokan nilai similarity pada 255 lowongan, Gabungan KUK terdiri atas 141 lowongan (55,29%) dalam kategori Sangat Sesuai dan 114 lowongan (44,71%) dalam kategori Sesuai. Sementara itu, Gabungan Elemen Kompetensi terdiri atas 142 lowongan (55,69%) dalam kategori Sangat Sesuai dan 113 lowongan (44,31%) dalam kategori Sesuai. Tidak terdapat lowongan dalam kategori Cukup Sesuai maupun Kurang Sesuai. Perbedaan nilai similarity antar-lowongan menunjukkan adanya perbedaan kedekatan semantik berdasarkan isi Persyaratan dan Deskripsi masing-masing lowongan terhadap keseluruhan kompetensi Pengembang Web.

Analisis *Word Occurrence* berhasil mengidentifikasi kompetensi yang paling banyak dibutuhkan berdasarkan frekuensi kemunculannya pada data lowongan kerja. Kompetensi C, API, dan Java menjadi tiga kompetensi dengan frekuensi tertinggi, masing-masing muncul pada 65,10%, 48,24%, dan 45,10% dari keseluruhan lowongan yang dianalisis.

Kombinasi analisis Cosine Similarity dan *Word Occurrence* memberikan informasi yang saling melengkapi. Cosine Similarity digunakan untuk mengukur kesesuaian kompetensi secara semantik antara Peta Okupasi dan kebutuhan industri, sedangkan *Word Occurrence*

digunakan untuk mengidentifikasi kompetensi yang paling sering dipersyaratkan oleh perusahaan. Dengan demikian, penelitian dapat memberikan gambaran mengenai tingkat kesesuaian kompetensi sekaligus kompetensi yang memiliki tingkat kebutuhan tinggi di dunia industri.

Untuk penelitian selanjutnya, disarankan agar cakupan data lowongan kerja diperluas dari lebih banyak platform sehingga hasil analisis dapat lebih merepresentasikan kebutuhan industri secara nasional, serta membandingkan performa mBERT dengan model representasi teks lain seperti IndoBERT atau Sentence-BERT untuk melihat metode mana yang memberikan hasil pengukuran kesesuaian paling optimal. Analisis juga dapat dikembangkan lebih lanjut dengan menambahkan teknik seperti clustering atau analisis skill gap yang lebih rinci, serta menyempurnakan dashboard dengan fitur pencarian dan penyaringan agar lebih mudah dipakai oleh pengguna.

DAFTAR REFERENSI

- Ahadi, A., Kitto, K., Rizioiu, M. A., & Musial, K. (2022). Skills Taught vs Skills Sought: Using Skills Analytics to Identify the Gaps between Curriculum and Job Markets. *Proceedings of the 15th International Conference on Educational Data Mining, EDM 2022*. <https://doi.org/10.5281/zenodo.6853121>
- Ardhana, A. Y. A., Syazeedah, H. N. U., Fitriyaningrum, R. I., & Gunawan, A. (2025). Analisis Ketidaksesuaian antara Pendidikan dengan Kebutuhan Dunia Kerja di Indonesia. *Kompeten: Jurnal Ilmiah Ekonomi dan Bisnis*, 3(4), 1020–1026. <https://doi.org/10.57141/kompeten.v3i4.156>
- Ariyanta, N. D., Prasetya, D. D., Ari, I., Zaeni, E., Wicaksono, R., & Hirashima, T. (2025). *Assessing the Semantic Alignment in Multilingual Student-Teacher Concept Maps Using mBERT*. 25(1), 113–126. <https://doi.org/10.30812/matrik.v25i1.5046>
- Ayofe, A. N., & Ajetola, A. R. (2021). *Exploration of the Gap Between Computer Science Curriculum and Industrial I.T Skills Requirements*. 4(1). <http://arxiv.org/abs/0908.4353>
- Barrera, M., Beraja, J., Galí, G., Lorenzoni, H., Rey, R., Schoenle, J., Steinsson, L., Straub, S., & Werning, I. (2022). *For excellent research assistance I thank Tomás Caravello and Pedro Martínez-Bruera. I am grateful for useful conversations and comments from Marc de la*. <http://www.nber.org/papers/w30260>
- Chanif, M. C., Much, I., & Subroto, I. (2026). *PENERAPAN MULTILINGUAL BERT UNTUK KLASIFIKASI BAHASA INDONESIA DAN ATAU BAHASA MALAYSIA PADA TEKS PENDEK*. 3(3), 446–457.
- Hasibuan, F. Z., Simangunsong, J., & Ilmiah, A. (2023). *ANALISA METODE COSINE SIMILARITY DALAM MENDETEKSI*. 3(1), 1–6.
- Javadian Sabet, A., Bana, S. H., Yu, R., & Frank, M. R. (2024). Course-Skill Atlas: A national longitudinal dataset of skills taught in U.S. higher education curricula. *Scientific Data*, 11(1), 1–21. <https://doi.org/10.1038/s41597-024-03931-8>

- Karakolis, E., Kapsalis, P., Skalidakis, S., Kontzinos, C., Kokkinakos, P., Markaki, O., & Askounis, D. (2022). Bridging the Gap between Technological Education and Job Market Requirements through Data Analytics and Decision Support Services. *Applied Sciences (Switzerland)*, 12(14). <https://doi.org/10.3390/app12147139>
- Karim, A. A. (2024). *Bridging the employability gap: aligning higher education curricula with industry-required competencies in the digital era*. 18(2), 93–110. [https://doi.org/10.29270/JBS.18.2\(24\).06](https://doi.org/10.29270/JBS.18.2(24).06)
- Kenton, M. C., Kristina, L., & Devlin, J. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 4171–4186.
- Keraghel, I., Morbieu, S., & Nadif, M. (2024). *Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study*. <http://arxiv.org/abs/2401.10825>
- Mandende, I. ., Sebola, M., Mudau, M. ., & Ndhobela, A. . (2024). Not on my watch: (Re)clamations of my Tshimanda, my identity, my language. *South African Journal of Higher Education*, 38(2), 196–216. <https://doi.org/10.20853/38-2-5424>
- Neogi, S., Pakray, P., Bandyopadhyay, S., & Gelbukh, A. (2012). *JU _ CSE _ NLP : Multi-grade Classification of Semantic Similarity Between Text Pairs*. 571–574.
- No, V., & Ferris Tita Sabilillah, Sri Winarto, R. B. A. (2024). *Edumatic : Jurnal Pendidikan Informatika Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir*. 8(2), 585–594. <https://doi.org/10.29408/edumatic.v8i2.27791>
- Quintini, G., & Verhagen, A. (2022). Skills for Jobs 2022. *Oecd*. https://www.oecdskillsforjobsdatabase.org/data/S4J2022_results.pdf
- Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, 11(3). <https://doi.org/10.3390/jintelligence11030054>
- Ula, K. R., & Faisal, M. (2025). Analisis Kesesuaian Lowongan Pekerjaan dan Kemampuan Pencari Kerja dengan Menggunakan Metode Cosine Similarity. *Techno.Com*, 24(3), 617–631. <https://doi.org/10.62411/tc.v24i3.12852>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU : Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*. 843–857.
- Yolanda Reza, E., Dwi Kartika, S., & Wahyu Widyaningsih, T. (2025). Analisis Relevansi Mata Kuliah Peminatan dengan Kompetensi Universitas di Industri Teknologi Informasi. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 8(2), 99–107. <https://doi.org/10.47970/siskom-kb.v8i1.761>