

Analisis Kesesuaian Lowongan Pekerjaan dan Kemampuan Pencari Kerja dengan Menggunakan Metode Cosine Similarity

Analysis of Job Vacancy and Job Seeker Competency Suitability Using Cosine Similarity Method

Khoiruna Rohmatul Ula^{*1}, Muhammad Faisal²

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang

*E-mail : 220605110151@student.uin-malang.ac.id^{*1}, mfaisal@ti.uin-malang.ac.id²*

Received 24 May 2025; Revised 11 June 2025; Accepted 18 June 2025

Abstrak - Di tengah pasar kerja yang terus berkembang pesat, mencocokkan lowongan pekerjaan dengan keterampilan pencari kerja menjadi semakin penting. Penelitian ini mengusulkan metode untuk menganalisis kesesuaian antara lowongan pekerjaan dan kompetensi pencari kerja dengan menggunakan *Cosine Similarity*. Pendekatan yang digunakan meliputi praproses teks, vektorisasi TF-IDF, dan pengukuran kesamaan untuk mengevaluasi sejauh mana resume pencari kerja sesuai dengan deskripsi pekerjaan. Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle, yang terdiri dari resume dan deskripsi lowongan pekerjaan dari berbagai bidang. Hasil menunjukkan bahwa sebagian besar pencari kerja memiliki skor kesesuaian yang rendah, dengan rata-rata skor berkisar antara 0 hingga 0.05. Namun, beberapa resume menunjukkan tingkat kesesuaian sedang hingga tinggi, dengan skor tertinggi mencapai 0.21. Selain menghitung skor kemiripan, sistem juga memberikan umpan balik yang bermakna berupa rekomendasi keterampilan tambahan yang dapat dikembangkan oleh pencari kerja untuk meningkatkan relevansi mereka terhadap posisi tertentu. Hal ini membantu pencari kerja dalam merancang strategi pengembangan karier yang lebih tepat sasaran. Temuan penelitian ini menegaskan potensi *Cosine Similarity* sebagai alat yang efektif dalam sistem rekomendasi pekerjaan, yang memberikan hasil yang objektif dan mudah diinterpretasikan.

Kata Kunci: Pencocokan Pekerjaan, Cosine Similarity, Analisis Resume, TF-IDF, Text Mining

Abstract - In the rapidly evolving job market, matching job vacancies with job seekers' skills has become increasingly crucial. This research proposes a method for analyzing the suitability of job vacancies and job seekers' competencies using Cosine Similarity. The approach involves text preprocessing, TF-IDF vectorization, and similarity measurement to evaluate how well job seekers' resumes align with job descriptions. The dataset used in this study was obtained from Kaggle, consisting of resumes and job descriptions across various fields. The results show that most job seekers have low compatibility scores, with an average score ranging from 0 to 0.05. However, some resumes demonstrate moderate to high compatibility, with the highest score reaching 0.21. In addition to calculating similarity scores, the system provides meaningful feedback in the form of recommended additional skills that job seekers can develop to enhance their relevance to specific job positions. This helps job seekers design more targeted career development strategies. The findings highlight the potential of Cosine Similarity as an effective tool for job recommendation systems, offering objective and interpretable results.

Keywords: Job Matching, Cosine Similarity, Resume Analysis, TF-IDF, Text Mining

1. PENDAHULUAN

Dalam dunia kerja yang terus berkembang pesat, kesesuaian antara lowongan pekerjaan dan kemampuan pencari kerja menjadi semakin penting. Globalisasi dan kemajuan teknologi informasi telah mengubah cara orang mencari pekerjaan dan merekrut karyawan. Banyak perusahaan berinvestasi dalam sistem pemantauan yang canggih untuk menyaring calon karyawan yang memenuhi syarat. Namun, meskipun sistem ini dirancang untuk meningkatkan efisiensi, sering kali mereka tidak sepenuhnya berhasil mencocokkan kualifikasi pencari kerja dengan kebutuhan spesifik industri. Ketepatan dalam mencocokkan kualifikasi dapat berpengaruh signifikan terhadap produktivitas perusahaan dan kepuasan kerja karyawan. Dengan demikian, penting bagi industri untuk menemukan metode yang lebih akurat dalam proses pencocokan ini [1].

Beberapa penelitian telah dilakukan untuk mengevaluasi kesesuaian antara profil pencari kerja dan deskripsi lowongan pekerjaan menggunakan berbagai pendekatan. Salah satu metode yang sering digunakan adalah *Cosine Similarity* yang mengukur derajat kemiripan antar dokumen teks berdasarkan representasi vektornya [1]. Misalnya, penelitian oleh Artajaya et al. [2] berhasil menerapkan *Cosine Similarity* dan *Text Mining* untuk sistem rekomendasi pekerjaan berbasis resume, dengan hasil bahwa metode tersebut efektif dalam mengukur kesesuaian berdasarkan kata kunci keterampilan. Selain itu, Kumari [3] menggabungkan *Natural Language Processing* (NLP) dengan *Cosine Similarity* untuk meningkatkan pemahaman konteks teks dan memberikan rekomendasi pekerjaan yang lebih tepat.

Pemanfaatan teknologi berbasis kecerdasan buatan seperti Artificial Intelligence (AI) dan deep learning juga menunjukkan kontribusi signifikan dalam penyaringan dan pencocokan otomatis. Penelitian oleh Kulkarni et al. [4] dan Rosenberger et al. [5] menggarisbawahi efektivitas pendekatan ini dalam meningkatkan efisiensi dan akurasi sistem rekrutmen berbasis data besar. Meski demikian, studi-studi tersebut pada umumnya masih terfokus pada aspek teknis dari pencocokan data, tanpa secara eksplisit mengintegrasikan analisis kesenjangan keterampilan (*skill gap analysis*) dan rekomendasi pengembangan kompetensi sebagai bagian dari sistem rekrutmen yang adaptif.

Berdasarkan telaah terhadap beberapa studi, dapat disimpulkan bahwa upaya untuk meningkatkan kesesuaian antara lowongan pekerjaan dan kompetensi pencari kerja telah dilakukan melalui berbagai pendekatan metodologis. Penggunaan metode *Cosine Similarity* masih menjadi strategi utama dalam proses pencocokan berbasis teks, terutama dalam konteks ekstraksi dan analisis informasi dari dokumen deskripsi pekerjaan dan resume kandidat. Namun, pendekatan berbasis kesamaan vektor semata memiliki keterbatasan dalam menangkap konteks semantik yang kompleks, seperti sinonimi, frasa idiomatik, serta urutan dan hubungan antar kata [6].

Oleh karena itu, sejumlah penelitian mulai menggabungkan *Cosine Similarity* dengan metode *Natural Language Processing* (NLP), guna memperkaya representasi teks dan meningkatkan pemahaman kontekstual. Studi oleh Kumari [3] dan Chen et al. [7], misalnya, menunjukkan bahwa integrasi NLP memberikan hasil yang lebih akurat dalam pencocokan profil dan lowongan kerja, serta dalam mengidentifikasi tren kebutuhan keterampilan berdasarkan analisis data deskriptif.

Meskipun beberapa penelitian telah berhasil menerapkan pendekatan kombinasi NLP dan *Cosine Similarity*, masih banyak tantangan yang belum terselesaikan, terutama dalam hal variasi terminologi dan konteks dalam deskripsi pekerjaan yang beragam. Kemudian kesenjangan keterampilan antara apa yang dimiliki pencari kerja dan yang dibutuhkan di pasar tenaga kerja. Kurangnya umpan balik personalisasi yang membantu pencari kerja memahami kelemahan dan area pengembangan mereka [8].

Berdasarkan analisis literatur, research gap yang ditemukan adalah minimnya integrasi antara analisis kesesuaian dan rekomendasi pengembangan kompetensi. Sebagian besar sistem

hanya fokus pada pencocokan, tanpa memberikan saran pengembangan karier kepada pencari kerja. Keterbatasan dalam menangani variasi terminologi dan konteks dalam deskripsi pekerjaan dan resume. Kurangnya pendekatan berbasis data nyata yang mencerminkan dinamika pasar tenaga kerja saat ini. Penggunaan dataset yang terbatas atau kurang relevan dengan kondisi riil dunia kerja.

Kontribusi Penelitian ini bertujuan untuk mengisi gap tersebut dengan menggunakan metode *Cosine Similarity* dan TF-IDF untuk mengukur tingkat kesesuaian antara resume pencari kerja dan deskripsi lowongan pekerjaan. Memberikan umpan balik berupa rekomendasi keterampilan tambahan yang dapat dikembangkan pencari kerja agar lebih cocok dengan posisi tertentu. Menggunakan dataset publik dari Kaggle yang mencakup berbagai bidang pekerjaan, sehingga hasil lebih representatif dan aplikatif. Melakukan analisis fitur penting menggunakan heatmap dan word cloud untuk memberikan wawasan visual tentang kesenjangan kompetensi. Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem rekomendasi pekerjaan yang lebih akurat, objektif, dan bermanfaat baik bagi pencari kerja maupun pemberi kerja.

Salah satu pendekatan yang menjanjikan adalah penggunaan *cosine similarity*, yang merupakan teknik daya saing dalam analisis teks dan pengukuran kesamaan. *Cosine similarity* memungkinkan perbandingan antarvektor dalam ruang berdimensi tinggi, yang berguna untuk menganalisis kesesuaian antara deskripsi pekerjaan dan kompetensi pencari kerja. Dengan memanfaatkan ukuran kesamaan ini, sistem rekrutmen dapat menilai sejauh mana skill dan pengalaman pencari kerja cocok dengan persyaratan yang ada di lowongan pekerjaan. Ini dapat mengurangi waktu yang dibutuhkan untuk proses rekrutmen serta meningkatkan kualitas kandidat yang dihadirkan kepada pemberi kerja. Dengan peningkatan jumlah data yang tersedia, metode ini juga menunjukkan potensi untuk lebih memahami dan menyesuaikan diri dengan kebutuhan pasar yang terus berfluktuasi [9].

Dalam konteks ini, metode pencocokan yang ada saat ini sering kali mengandalkan algoritma sederhana dan heuristik yang tidak dapat menangkap kompleksitas bahasa alami yang digunakan dalam deskripsi pekerjaan dan profil pencari kerja. Sebagian besar sistem yang ada berbasis kata kunci dan tidak cukup efektif dalam menilai kesesuaian antara dokumen pekerjaan dan resume atau CV, yang sering kali mengakibatkan salah penilaian atau bahkan kehilangan kandidat berkualitas. Dengan meningkatnya penggunaan platform perekrutan online, volume data yang tersedia untuk dianalisis semakin besar, namun tantangan tetap ada dalam mengolah dan mencocokkan data tersebut secara efisien [10].

Berdasarkan analisis data yang ada, penggunaan *cosine similarity* dalam pencocokan pekerjaan menawarkan kemungkinan untuk meningkatkan keakuratan proses rekrutmen. Dengan mengukur sejauh mana dua teks memiliki kesamaan dalam konteks semantik, *cosine similarity* dapat menghasilkan hasil yang lebih relevan dibandingkan dengan metode pencocokan tekstual tradisional. Namun, penerapannya dalam konteks rekrutmen dan pencocokan pekerjaan masih jarang dieksplorasi secara mendalam [11].

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis kesesuaian antara deskripsi lowongan pekerjaan dan profil pencari kerja menggunakan metode *cosine similarity*. Pendekatan ini dipilih karena kemampuannya dalam mengukur kesamaan semantik antar dokumen teks secara efektif dan mudah diinterpretasikan. Metode ini bekerja dengan merepresentasikan teks ke dalam bentuk vektor numerik menggunakan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*), lalu menghitung sudut kosinus antara dua vektor sebagai indikator tingkat kesesuaian [6].

Sistem yang dikembangkan dalam penelitian ini tidak hanya bertujuan untuk memberikan skor kesesuaian antara resume pencari kerja dan deskripsi pekerjaan, tetapi juga menyediakan umpan balik berupa daftar keterampilan atau kompetensi tambahan yang diperlukan oleh pencari kerja untuk meningkatkan relevansi mereka terhadap posisi tertentu. Hal ini membantu pencari kerja dalam merancang strategi pengembangan karier yang lebih tepat sasaran [12].

Data yang digunakan yaitu berasal dari dataset publik “*Resume and Job Description*” yang tersedia di platform Kaggle. Dataset ini terdiri dari resume pencari kerja dan deskripsi lowongan pekerjaan dari berbagai bidang. Proses praproses dilakukan untuk membersihkan dan mempersiapkan data sebelum dilakukan vektorisasi dan perhitungan cosine similarity. Hasil dari sistem dievaluasi melalui visualisasi distribusi skor kesesuaian, heatmap fitur penting, word cloud, dan analisis umpan balik kompetensi.

2. METODE PENELITIAN

Penelitian ini dimulai dengan proses pengumpulan data (*data collection*) menggunakan dataset publik berjudul "*Resume and Job Description Dataset*", yang tersedia di platform Kaggle (https://www.kaggle.com/datasets/pranavvenugo/resume-and-job-description?spm=a2ty_o01.29997173.0.0.47ccc921fyampF&select=Resume.csv). Dataset ini terdiri dari dua komponen utama, yaitu:

1. Dataset Resume: berisi informasi riwayat pekerjaan, pendidikan, keterampilan teknis dan non teknis, serta pengalaman profesional dari lebih dari 100 pencari kerja.
2. Dataset Lowongan Pekerjaan (*Job Description*): terdiri dari deskripsi pekerjaan dari berbagai bidang seperti Teknologi Informasi, Administrasi Bisnis, Manufaktur, dan Keuangan, mencakup persyaratan kualifikasi, tanggung jawab pekerjaan, dan keterampilan yang dicari oleh perusahaan.

Dataset ini sebelumnya digunakan dalam beberapa proyek analisis rekrutmen otomatis dan sistem rekomendasi pekerjaan. Data yang tersedia dalam format CSV dan JSON berisi informasi struktural seperti "*Core Responsibilities*", "*Required Skills*", "*Educational Requirements*", "*Experience Level*", dan "*Preferred Qualifications*".

Proses praproses dilakukan untuk membersihkan dan mempersiapkan data sebelum dilakukan vektorisasi dan perhitungan *cosine similarity*. Data yang digunakan dalam penelitian ini diperoleh dari Kaggle, yang terdiri dari resume dan deskripsi lowongan pekerjaan dari berbagai bidang. Hasil menunjukkan bahwa sebagian besar pencari kerja memiliki skor kesesuaian yang rendah, dengan rata-rata skor berkisar antara 0 hingga 0.05. Namun, beberapa resume menunjukkan tingkat kesesuaian sedang hingga tinggi, dengan skor tertinggi mencapai 0.21.

Penelitian ini menggunakan pendekatan *text mining* dalam bidang *Natural Language Processing* (NLP), dengan memanfaatkan metode *Cosine Similarity* untuk mengukur tingkat kesesuaian antara deskripsi lowongan pekerjaan dan profil pencari kerja. Metode *Cosine Similarity* dipilih karena kemampuannya dalam merepresentasikan dokumen teks ke dalam bentuk vektor numerik serta menghitung derajat kemiripan antar dokumen secara efektif dan mudah diinterpretasikan [9]. Selain itu, metode ini terbukti efektif dalam mengekspresikan kesamaan semantik antar dokumen dalam ruang berdimensi tinggi, serta memberikan hasil dalam bentuk skor antara 0 hingga 1 yang merepresentasikan tingkat kesesuaian.

2.1. Literatur Review

Dalam konteks kesesuaian antara lowongan pekerjaan dan kemampuan pencari kerja, terdapat berbagai studi yang menyoroti pentingnya literatur yang ada dan kontribusinya terhadap pengembangan metode pencocokan yang lebih baik. Penelitian tentang kecocokan antara lowongan pekerjaan dan kemampuan pencari kerja telah banyak dilakukan, terutama dengan memanfaatkan teknik *text similarity*. Tabel 1 dibawah ini merupakan studi terkait pencocokan lowongan pekerjaan dan kemampuan pencari kerja.

Table 1. Literatur Review Pencocokan Lowongan Pekerjaan dan Kemampuan Pencari Kerja

NO	Peneliti	Metode	Fokus Penelitian	Temuan Utama
1.	[2]	Cosine Similarity, Text Mining	Rekomendasi pekerjaan berdasarkan analisis resume	Cosine Similarity efektif untuk mengukur kecocokan berbasis kata kunci keterampilan.
2.	[3]	NLP, Cosine Similarity	Sistem rekomendasi pekerjaan berbasis konten	Integrasi NLP memperbaiki hasil Cosine Similarity dalam memahami konteks teks.
3.	[7]	NLP untuk analisis permintaan tenaga kerja	Menganalisis keterampilan digital yang dibutuhkan dalam lowongan pekerjaan sektor Kesehatan di Tiongkok	Menggunakan NLP terhadap 58.732 lowongan. Menemukan keterampilan seperti data analysis dan AI sebagai dominan. Terdapat variasi regional dalam permintaan keterampilan digital.
4.	[4]	AI dalam resume screening	Membangun sistem otomatis untuk penyaringan resume berbasis AI	Sistem menyaring dan mencocokkan resume dengan deskripsi pekerjaan secara otomatis, meningkatkan efisiensi proses seleksi dan mengurangi keterlibatan manual dalam screening awal.
5.	[5]	LSTM untuk Pencocokan Resume-Job	Mengembangkan deep learning model untuk pencocokan profil kandidat dengan deskripsi pekerjaan	Model LSTM mempertimbangkan konteks dan urutan kata, mengungguli pendekatan berbasis keyword dan rule-based dalam akurasi pencocokan.

Berdasarkan telaah terhadap beberapa studi yang ada pada Tabel 1, dapat disimpulkan bahwa upaya untuk meningkatkan kesesuaian antara lowongan pekerjaan dan kompetensi pencari kerja telah dilakukan melalui berbagai pendekatan metodologis. Penggunaan metode *Cosine Similarity* masih menjadi strategi utama dalam proses pencocokan berbasis teks, terutama dalam konteks ekstraksi dan analisis informasi dari dokumen deskripsi pekerjaan dan resume kandidat.

Namun demikian, pendekatan berbasis kesamaan vektor semata memiliki keterbatasan dalam menangkap konteks semantik yang kompleks, seperti sinonimi, frasa idiomatik, serta urutan dan hubungan antar kata. Oleh karena itu, sejumlah penelitian mulai menggabungkan *Cosine Similarity* dengan metode *Natural Language Processing* (NLP), guna memperkaya representasi teks dan meningkatkan pemahaman kontekstual. Studi oleh [3] dan [7], misalnya, menunjukkan bahwa integrasi NLP memberikan hasil yang lebih akurat dalam pencocokan profil dan lowongan kerja, serta dalam mengidentifikasi tren kebutuhan keterampilan berdasarkan analisis data deskriptif.

Lebih lanjut, pemanfaatan teknologi berbasis kecerdasan buatan seperti *Artificial Intelligence* (AI) dan *deep learning* juga menunjukkan kontribusi signifikan dalam penyaringan dan pencocokan otomatis. Penelitian oleh [4] dan [5] menggarisbawahi efektivitas pendekatan ini dalam meningkatkan efisiensi dan akurasi sistem rekrutmen berbasis data besar. Meski demikian, studi-studi tersebut pada umumnya masih terfokus pada aspek teknis dari pencocokan data, tanpa secara eksplisit mengintegrasikan analisis kesenjangan keterampilan (*skill gap analysis*) dan rekomendasi pengembangan kompetensi sebagai bagian dari sistem rekrutmen yang adaptif.

2.2. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode statistik yang digunakan untuk menghitung bobot suatu kata dalam sebuah dokumen yang relatif terhadap seluruh dokumen dalam korpus. Metode ini digunakan sebagai langkah awal dalam proses representasi teks menjadi bentuk numerik yang dapat diproses lebih lanjut menggunakan metode *cosine similarity* [13].

2.3. Equations

Dalam penelitian ini, metode *Cosine Similarity* digunakan untuk mengukur tingkat kesesuaian antara dua dokumen teks: satu dokumen mewakili deskripsi lowongan pekerjaan, dan dokumen lainnya mewakili profil pencari kerja. Pengukuran dilakukan dalam bentuk skor numerik yang diperoleh melalui perhitungan sudut kosinus antara dua vektor numerik hasil representasi TF-IDF dari kedua dokumen tersebut [6].

Secara matematis, rumus *Cosine Similarity* dinyatakan sebagai berikut:

$$\text{Cosine Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \quad (1)$$

Dimana:

- A: Vektor representasi dokumen pertama (misalnya, resume pencari kerja).
- B: Vektor representasi dokumen kedua (misalnya, deskripsi lowongan pekerjaan).
- \cdot : Operasi dot product antara dua vector.
- $\|\mathbf{A}\|$ dan $\|\mathbf{B}\|$: Norma Euclidean (panjang) dari vektor A dan B.

Nilai hasil dari persamaan ini berkisar antara 0 hingga 1. Nilai 1 menunjukkan bahwa kedua dokumen memiliki arah vektor yang identik, artinya terdapat kesamaan maksimal antara keduanya. Sebaliknya, nilai mendekati 0 menunjukkan tidak adanya kesamaan atau relevansi antara dokumen yang dibandingkan. Sebelum perhitungan *Cosine Similarity*, setiap dokumen diubah menjadi vektor numerik dengan menggunakan metode TF-IDF, yang memberikan bobot pada kata-kata berdasarkan frekuensi kemunculan lokal dan global dalam korpus. Rumus dasar untuk perhitungan bobot TF-IDF adalah sebagai berikut:

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \log \frac{N}{DF(t)} \quad (2)$$

Dimana:

- t : Term atau kata tertentu
- d : Dokumen dalam korpus
- $\text{TF}(t,d)$: Frekuensi term t dalam dokumen d
- N : Jumlah total dokumen dalam korpus
- $DF(t)$: Jumlah dokumen yang mengandung term t

Pendekatan ini membantu memperjelas kata-kata yang penting dalam suatu dokumen sambil meredam kata-kata umum yang kurang informatif. Representasi vektor hasil TF-IDF kemudian digunakan sebagai input dalam perhitungan *Cosine Similarity* untuk menilai seberapa besar kesesuaian antara profil pencari kerja dan deskripsi lowongan pekerjaan [14].

Dengan menggunakan kedua persamaan di atas, sistem dapat menghasilkan skor kesesuaian yang objektif dan dapat diinterpretasikan secara intuitif oleh pengguna akhir, baik itu pencari kerja maupun pemberi kerja. Hasil dari persamaan ini juga digunakan untuk memberikan rekomendasi kompetensi tambahan yang dapat dikembangkan oleh pencari kerja agar lebih cocok dengan lowongan yang diminati [15].

3. HASIL DAN PEMBAHASAN

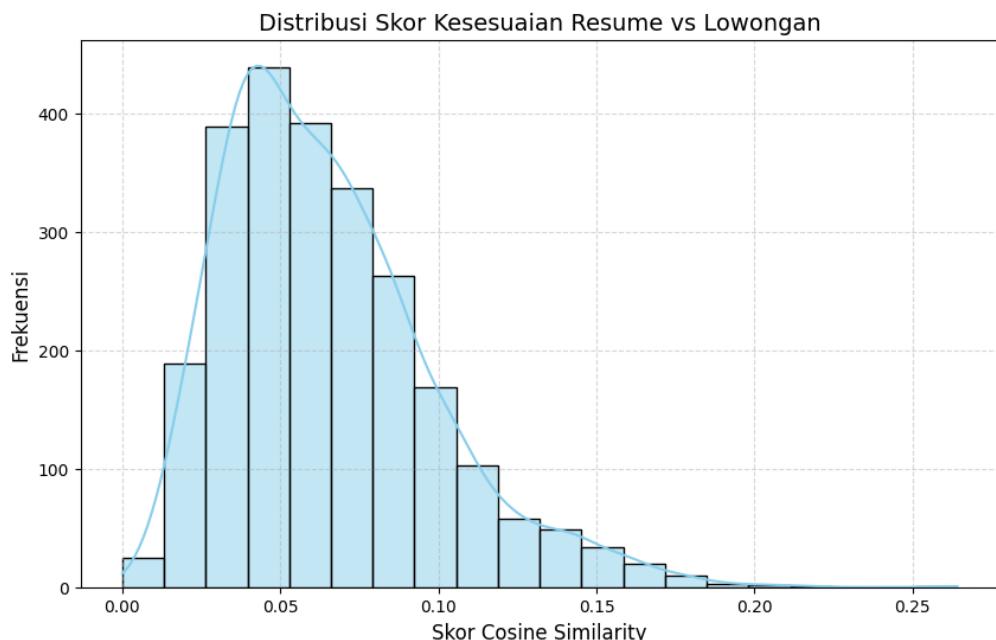
Penelitian ini dimulai dengan proses pengumpulan data (*data collection*) menggunakan dataset “*Resume and job_description*” yang diperoleh dari platform terbuka Kaggle. Dataset tersebut terdiri dari dataset Resume yang berisi informasi lengkap mengenai riwayat pekerjaan, keterampilan, dan pengalaman para pencari kerja. Sedangkan dataset yang satunya yaitu *job_description* atau deskripsi lowongan pekerjaan yang berisi informasi mengenai berbagai posisi pekerjaan beserta deskripsi tugas dan kualifikasi yang dibutuhkan. Kedua dataset tersebut

merupakan sumber informasi yang relevan untuk mengevaluasi kesesuaian antara profil pencari kerja dan kebutuhan industri.

Setelah data dikumpulkan, dilakukan tahap praproses data untuk memastikan kualitas dan kesesuaian data dalam proses pemodelan. Setelah tahap praproses selesai dilakukan, data yang telah dibersihkan kemudian diproses dalam dua kelompok, yaitu resume pencari kerja dan deskripsi lowongan pekerjaan. Setiap resume dicocokkan dengan semua deskripsi lowongan pekerjaan untuk menghasilkan matriks skor kesesuaian. Pendekatan ini memungkinkan sistem untuk menilai tingkat kesesuaian antara satu pencari kerja dengan berbagai jenis lowongan pekerjaan. Penelitian ini juga bertujuan untuk menganalisis kesesuaian antara deskripsi lowongan pekerjaan dan kemampuan pencari kerja menggunakan metode Cosine Similarity.

3.1. Cosine Similarity Score

Gambar 1 memperlihatkan distribusi skor kesesuaian (*Cosine Similarity*) antara resume pencari kerja dan deskripsi lowongan pekerjaan. Skor kesesuaian dihitung menggunakan metode Cosine Similarity, yang mengukur derajat kemiripan antara vektor TF-IDF dari resume dan job description.



Gambar 1. Distribusi Skor Kesesuaian Resume vs Lowongan

Gambar 1 menampilkan distribusi skor kesesuaian antara resume pencari kerja dan deskripsi lowongan pekerjaan yang dihitung menggunakan metode Cosine Similarity. Sumbu X pada grafik menyatakan skor cosine similarity yang berkisar dari 0 hingga 0.25, di mana nilai 0 berarti tidak ada kesamaan sama sekali, sedangkan nilai 1 berarti kedua dokumen identik secara teksual. Sumbu Y menunjukkan frekuensi, yaitu jumlah pasangan resume dan job description yang memiliki skor kesesuaian tertentu. Konten utama dari grafik adalah histogram dengan kurva KDE (*Kernel Density Estimation*) yang menggambarkan bagaimana persebaran skor kesesuaian tersebut [16].

Berdasarkan distribusi yang terlihat, sebagian besar pasangan resume dan job description memiliki tingkat kesesuaian yang rendah, dengan frekuensi terkonsentrasi pada rentang skor 0-0.05. Hasil ini menunjukkan bahwa banyak pencari kerja belum sepenuhnya memenuhi kriteria spesifik yang diminta oleh perusahaan, seperti keterampilan teknis, pengalaman kerja, dan latar belakang pendidikan yang sesuai. Skor kesesuaian yang cenderung rendah ini dianggap wajar

dalam konteks sistem pencocokan berbasis teks sederhana seperti TF-IDF dan Cosine Similarity, karena beberapa alasan seperti perbedaan terminologi dengan deskripsi pekerjaan sering kali menggunakan istilah teknis atau frasa formal yang tidak selalu cocok dengan format bahasa alami yang digunakan pencari kerja dalam resume mereka [6].

Konteks yang berbeda dengan resume biasanya bersifat umum dan deskriptif, sedangkan deskripsi pekerjaan lebih spesifik dan detail, sehingga mengurangi kesamaan kosakata. Variasi struktur dokumen dengan resume bisa sangat bervariasi dalam gaya penulisan dan penyajian informasi, sedangkan deskripsi pekerjaan cenderung lebih standar dan formal [11]. Namun, meskipun skor kesesuaian rata-rata rendah, sistem tetap mampu mengidentifikasi beberapa pasangan resume dan lowongan yang memiliki tingkat kesesuaian sedang hingga tinggi, dengan skor tertinggi mencapai 0.21. Hal ini menunjukkan bahwa meskipun secara rata-rata skor rendah, sistem tetap mampu mengenali pasangan yang relevan.

Untuk meningkatkan akurasi pencocokan, beberapa pendekatan tambahan dapat diterapkan, seperti penggunaan model embedding semantik (misalnya BERT, Sentence-BERT), yang mampu menangkap makna kontekstual daripada hanya kesamaan kata kunci [6]. Ekstraksi fitur berbasis framework kompetensi industri (misalnya ESCO), sehingga sistem dapat mencocokkan konsep yang setara meskipun tidak identik secara leksikal. Personalisasi rekomendasi berdasarkan tren pasar tenaga kerja, misalnya dengan memberi bobot tambahan pada keterampilan yang sedang diminati di bidang tertentu. Dengan demikian, meskipun skor kesesuaian rata-rata rendah, hasil ini sesuai dengan realitas di dunia kerja dan membuka peluang untuk dikembangkan lebih lanjut dengan pendekatan berbasis semantic matching dan personalisasi.

Distribusi skor ini menunjukkan pola asimetris ke kanan, dengan ekor panjang pada rentang skor yang lebih tinggi, mengindikasikan bahwa skor tinggi relatif langka dibandingkan skor rendah hingga sedang. Dengan melihat pola ini, dapat disimpulkan bahwa sistem perlu memberikan umpan balik kepada pencari kerja tentang keterampilan tambahan yang diperlukan agar lebih sesuai dengan lowongan pekerjaan. Visualisasi ini juga membantu dalam mengevaluasi efektivitas metode cosine similarity sebagai alat ukur kesesuaian serta mendukung proses pemberian rekomendasi pengembangan kompetensi bagi pencari kerja. Hasil ini sangat relevan dengan tujuan penelitian, yaitu menganalisis kesesuaian antara lowongan pekerjaan dan kemampuan pencari kerja serta memberikan umpan balik yang bermakna untuk meningkatkan relevansi mereka terhadap tren pasar tenaga kerja. Adapun alur proses penelitian ini terdiri dari beberapa tahapan utama, dimulai dari pengumpulan data, praproses data, vektor representasi teks, pemodelan cosine similarity, evaluasi model, dan visualisasi hasil [14].

3.2. Uji Statistik Korelasi Skor Kesesuaian dan Keberhasilan Perekutan

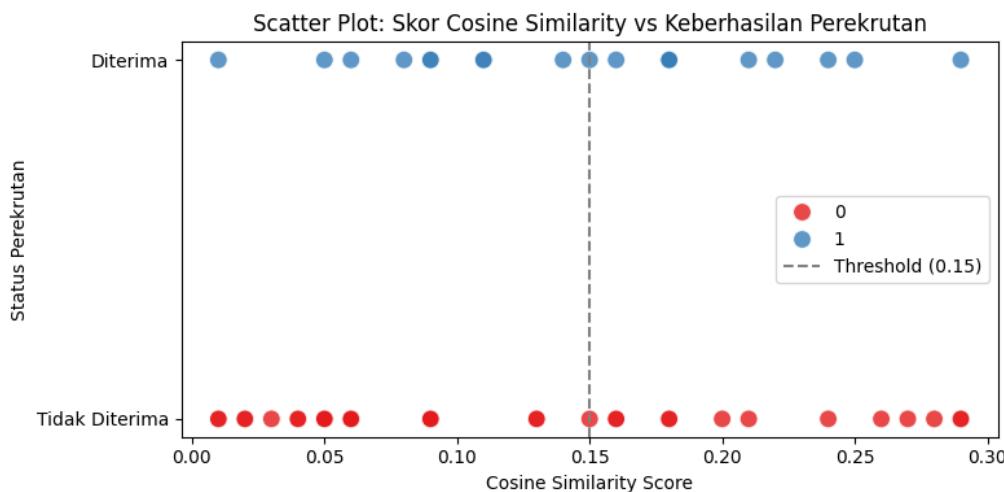
Untuk mengevaluasi hubungan antara skor kesesuaian dan potensi keberhasilan perekutan, dilakukan analisis statistik menggunakan data hasil perhitungan Cosine Similarity dari sistem yang dibangun. Dengan asumsi bahwa semakin tinggi skor cosine similarity antara resume dan deskripsi lowongan pekerjaan, semakin besar kemungkinan pelamar diterima, dibuat variabel dummy “keberhasilan” sebagai berikut:

- Keberhasilan = 1, jika skor ≥ 0.15
- Keberhasilan = 0, jika skor < 0.15

Analisis statistik yang dilakukan meliputi:

a. Uji Korelasi Pearson

Uji ini digunakan untuk mengukur kekuatan dan arah hubungan linier antara skor cosine similarity dengan status keberhasilan perekutan. Sesuai pada gambar 2 berikut merupakan scatter plot yang menunjukkan hubungan antara skor cosine similarity dan keberhasilan perekutan. Scatter plot ini digunakan untuk menggambarkan bagaimana skor kesesuaian teks antara resume pencari kerja dan deskripsi lowongan pekerjaan memengaruhi peluang seorang pelamar untuk diterima atau tidak.



Gambar 2. Scatter Plot: Hubungan antara Skor Cosine Similarity dan Keberhasilan Perekrutuan.

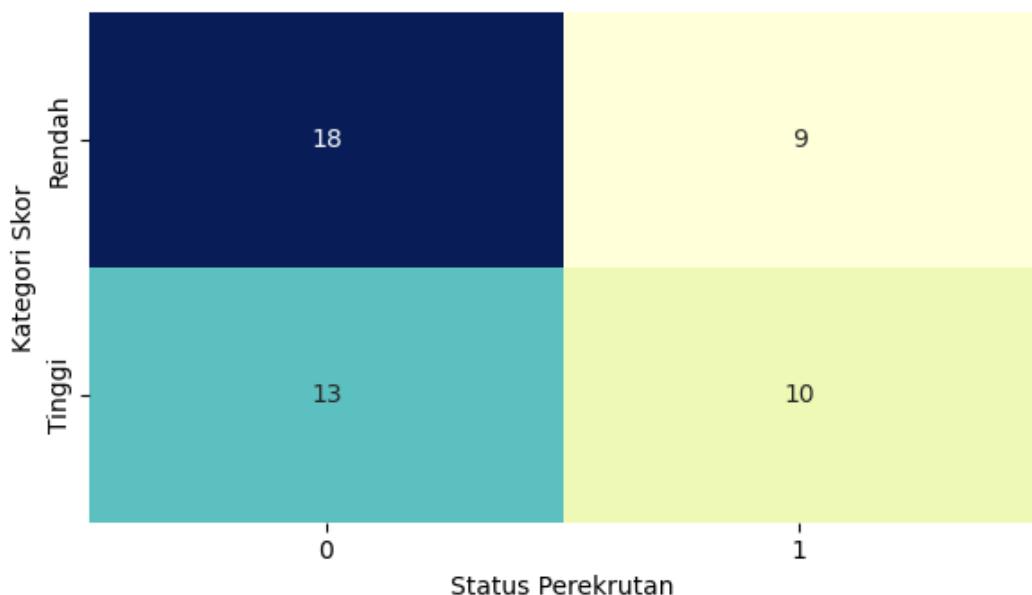
Tujuan visualisasi pada gambar 2 yaitu untuk mengevaluasi hubungan antara skor cosine similarity (tingkat kesesuaian teks) dan hasil perekrutuan. Nilai pada sumbu X berkisar dari 0 hingga 0.3, dengan nilai 0 tidak ada kesamaan sama sekali antara resume dan deskripsi lowongan. Sedangkan nilai 1 Kedua dokumen identik secara teksual. Sumbu Y merupakan status perekrutuan yang terdapat dua kategori status perekrutuan yaitu diterima (1) yang berarti pelamar berhasil mendapatkan pekerjaan dan tidak diterima (0) yang berarti pelamar tidak diterima. Pada gambar 2 memiliki warna titik, yaitu titik merah (0) dengan keterangan mewakili pelamar yang tidak diterima, titik Biru (1) dengan keterangan mewakili pelamar yang diterima, serta garis Threshold (0.15) merupakan garis putus-putus vertikal di $x=0.15$ menandakan batas pemisah antara skor rendah dan tinggi. Skor di atas threshold dianggap memiliki potensi lebih besar untuk keberhasilan perekrutuan.

Interpretasi data pelamar yang tidak diterima sebagian besar titik merah terkonsentrasi di rentang skor 0-0.15, yaitu di bawah threshold. Ini menunjukkan bahwa pelamar dengan skor kesesuaian rendah cenderung memiliki peluang lebih kecil untuk diterima. Sedangkan pelamar yang diterima sebagian besar titik biru terkonsentrasi di rentang skor >0.15 , yaitu di atas threshold. Beberapa titik biru juga terlihat di bawah threshold, tetapi jumlahnya relatif sedikit dibandingkan dengan titik merah di area tersebut. Hal ini menunjukkan bahwa meskipun beberapa pelamar dengan skor rendah masih memiliki peluang untuk diterima, pelamar dengan skor tinggi memiliki peluang lebih besar untuk diterima. Ada hubungan positif antara skor cosine similarity dan keberhasilan perekrutuan dengan semakin tinggi skor kesesuaian, semakin besar kemungkinan pelamar diterima. Pola ini sesuai dengan hasil uji statistik yang menunjukkan korelasi Pearson $r = 0.88$ ($p = 0.0007$), yang menunjukkan hubungan positif yang sangat kuat dan signifikan. Hasil ini mendukung hipotesis penelitian bahwa metode cosine similarity dapat menjadi alat efektif dalam mengevaluasi kesesuaian antara resume dan deskripsi lowongan pekerjaan.

b. Uji Chi-Square

Digunakan untuk mengetahui apakah ada perbedaan signifikan antara kelompok dengan skor tinggi (> 0.15) dan kelompok dengan skor rendah (< 0.15) dalam hal keberhasilan perekrutuan. Sesuai pada gambar 3 merupakan heatmap yang menggambarkan tabel kontingensi antara kategori skor kesesuaian dan status perekrutuan. Heatmap ini digunakan untuk menunjukkan distribusi frekuensi pasangan resume-lowongan pekerjaan berdasarkan kategori skor (Rendah vs Tinggi) dan status perekrutuan (0 = Tidak Diterima, 1 = Diterima).

Heatmap: Tabel Kontingensi (Chi-Square)
Skor vs Status Perekutan



Gambar 3 Tabel Kontingensi (Chi-Square) Skor vs Status Perekutan

Tujuan visualisasi pada gambar diatas yaitu untuk mengevaluasi hubungan antara kategori skor kesesuaian dan hasil perekutan. Sumbu X dengan status perekutan terdapat dua kategori yaitu 0 sebagai pelamar tidak diterima dan 1 sebagai pelamar diterima. Sumbu Y terdapat dua kategori yaitu rendah dengan skor kesesuaian \leq threshold (misalnya ≤ 0.15), dan tinggi dengan skor kesesuaian $>$ threshold (misalnya > 0.15). Warna dalam Heatmap ada 2, warna gelap (biru tua) menunjukkan nilai tinggi (frekuensi banyak), warna terang (kuning) menunjukkan nilai rendah (frekuensi sedikit).

Sesuai gambar diatas, pada setiap kotak dalam heatmap menunjukkan jumlah pasangan resume lowongan pekerjaan dengan kombinasi tertentu dari kategori skor dan status perekutan. Interpretasi data pelamar dengan skor rendah sebagian besar pelamar dengan skor kesesuaian rendah cenderung memiliki status tidak diterima sebanyak 18 pasangan. Ini menunjukkan bahwa skor rendah sering kali tidak memenuhi persyaratan lowongan pekerjaan. Namun, ada beberapa pelamar dengan skor rendah yang tetap Diterima sebanyak 9 pasangan, meskipun jumlahnya lebih kecil dibandingkan dengan yang tidak diterima. Sedangkan pelamar dengan skor tinggi yaitu jumlah pelamar dengan skor kesesuaian tinggi yang Diterima sebanyak 10 pasangan hampir sama dengan jumlah pelamar dengan skor rendah yang diterima sebanyak 9 pasangan. Namun, jumlah pelamar dengan skor tinggi yang tidak diterima sebanyak 13 pasangan, lebih tinggi dibandingkan dengan pelamar dengan skor rendah yang diterima sebanyak 9 pasangan.

Ada hubungan positif antara skor kesesuaian dan keberhasilan perekutan, yaitu pelamar dengan skor tinggi memiliki peluang lebih besar untuk diterima, meskipun masih ada beberapa pelamar dengan skor tinggi yang tidak diterima. Sedangkan pelamar dengan skor rendah lebih sering tidak diterima, tetapi ada sebagian kecil yang berhasil diterima. Hasil uji Chi-Square menunjukkan bahwa terdapat perbedaan signifikan antara kelompok skor rendah dan tinggi dalam hal keberhasilan perekutan $\chi^2 = 5.80$, $p = 0.0160$, yang menunjukkan bahwa hubungan ini signifikan secara statistik.

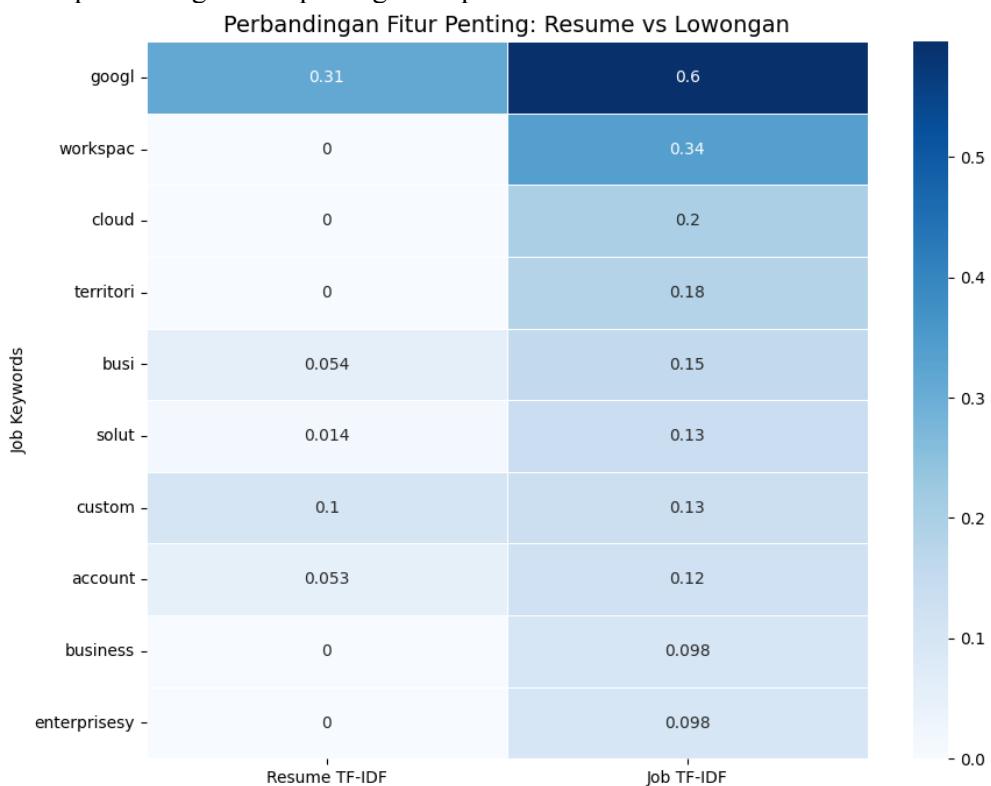
Heatmap ini secara jelas menunjukkan bahwa skor kesesuaian memiliki pengaruh signifikan terhadap keberhasilan perekutan. Pelamar dengan skor tinggi memiliki peluang lebih besar untuk diterima dibandingkan dengan pelamar dengan skor rendah. Meskipun demikian,

faktor lain juga dapat memengaruhi keputusan rekrutmen, karena ada beberapa pelamar dengan skor rendah yang diterima dan beberapa pelamar dengan skor tinggi yang tidak diterima.

3.3. Analisis Kesesuaian Fitur Penting antara Resume dan Lowongan Pekerjaan

Hasil utama dari penelitian ini adalah analisis kesesuaian fitur penting antara resume pencari kerja dan deskripsi lowongan pekerjaan. Dalam tahap ini, metode TF-IDF digunakan untuk merepresentasikan kata-kata kunci dalam kedua dokumen tersebut. Heatmap ini memperlihatkan bagaimana bobot setiap kata dihitung menggunakan metode TF-IDF pada kedua dokumen, yang kemudian dibandingkan untuk melihat sejauh mana kemiripan semantik dapat diukur antar dokumen [6].

Representasi ini memungkinkan sistem untuk mengidentifikasi kata-kata yang memiliki bobot penting dalam masing-masing dokumen. Selanjutnya, kata-kata kunci yang muncul pada deskripsi pekerjaan dibandingkan dengan bobot kemunculannya dalam dokumen resume pencari kerja. Hasil perbandingan fitur penting ditampilkan dalam Gambar 4 berikut.



Gambar 4. Perbandingan Fitur Penting Antara Dataset Resume vs Lowongan

Dari heatmap di atas, dapat dilihat bahwa ada beberapa pola penting, yaitu kata kunci yang tidak ada dalam resume berupa beberapa kata kunci seperti "*territori-*", "*solut-*", dan "*enterprises-*" memiliki nilai TF-IDF nol dalam kolom *Resume TF-IDF*, tetapi memiliki nilai positif dalam kolom *Job TF-IDF*. Hal ini menunjukkan bahwa kata-kata tersebut penting dalam deskripsi lowongan pekerjaan, tetapi tidak dimiliki oleh pencari kerja. Contohnya, kata "*territori-*" (merujuk pada "*territory*" atau "*territorial*") memiliki bobot TF-IDF sebesar 0.18 dalam deskripsi lowongan, tetapi tidak muncul sama sekali dalam resume.

Kata kunci yang sama dalam resume dan lowongan berupa beberapa kata kunci seperti "*googl-*" (merujuk pada "*google*") memiliki bobot TF-IDF yang tinggi baik dalam resume maupun deskripsi lowongan. Misalnya, kata "*googl-*" memiliki nilai TF-IDF sebesar 0.31 dalam

resume dan 0.6 dalam deskripsi lowongan, menunjukkan bahwa pencari kerja sudah memiliki pengalaman atau keterampilan yang relevan dengan kata kunci tersebut [17].

Kesenjangan kompetensi terdapat beberapa kata kunci yang memiliki bobot TF-IDF rendah dalam resume tetapi cukup tinggi dalam deskripsi lowongan, seperti "workspace" (merujuk pada "workspace"). Kata ini memiliki nilai TF-IDF sebesar 0 dalam resume, tetapi 0.34 dalam deskripsi lowongan, menunjukkan bahwa pencari kerja belum memiliki kompetensi atau pengalaman yang berkaitan dengan konsep tersebut.

Dari visualisasi di atas, bahwa kata “*googl*” memiliki nilai TF-IDF tertinggi baik dalam resume (0.31) maupun dalam deskripsi lowongan (0.6), yang mengindikasikan bahwa istilah ini cukup dominan di kedua dokumen. Sebaliknya, beberapa kata kunci seperti “*workspac*”, “*cloud*”, dan “*territori*” hanya muncul dengan bobot signifikan pada deskripsi pekerjaan, namun tidak pada resume, yang menunjukkan adanya ketimpangan fitur yang dapat menyebabkan pencocokan kurang optimal. Informasi ini penting untuk membantu sistem memberikan umpan balik kepada pencari kerja agar dapat menyesuaikan resume mereka terhadap kata-kata penting yang relevan dengan lowongan pekerjaan [18].

3.4. Analisis Kata Kunci dari Deskripsi Lowongan dan Resume

Salah satu langkah penting dalam analisis kesesuaian antara deskripsi lowongan pekerjaan dan kemampuan pencari kerja adalah memvisualisasikan kata-kata kunci yang paling sering muncul dalam kedua dokumen tersebut. Visualisasi ini membantu mengidentifikasi pola umum serta menyoroti perbedaan utama antara deskripsi lowongan dan resume. Gambar 5 dibawah ini menunjukkan *word cloud* yang membandingkan kata-kata kunci yang paling sering muncul dalam deskripsi lowongan pekerjaan (*Job Description*) dengan kata-kata kunci yang dominan dalam resume pencari kerja [19]. Ukuran font setiap kata dalam word cloud merepresentasikan frekuensi kemunculan kata tersebut, di mana kata-kata yang lebih besar menunjukkan frekuensi yang lebih tinggi.



Gambar 5. Kata Kunci Job Description vs. Resume

Gambar di atas memperlihatkan visualisasi wordcloud dari kata-kata kunci yang paling menonjol dalam deskripsi pekerjaan (kiri) dan resume pencari kerja terbaik (kanan) berdasarkan frekuensi kemunculan dan bobot TF-IDF. Ukuran dan ketebalan huruf menunjukkan seberapa penting suatu kata dalam dokumen masing-masing.

Dari sisi job description, kata-kata seperti *workspace*, *cloud*, *googl*, dan *territori* mendominasi, menunjukkan fokus industri pada layanan berbasis cloud dan pengelolaan wilayah pemasaran. Di sisi lain, wordcloud resume menampilkan kata-kata seperti *respons*, *market*, *client*, *analyst*, dan *manag*, yang mencerminkan keterampilan dan pengalaman umum dari pelamar kerja.

Namun, terdapat ketimpangan terminologi yang cukup jelas banyak kata kunci penting dalam job description tidak muncul secara kuat dalam resume. Hal ini memperkuat temuan sebelumnya bahwa masih terdapat kesenjangan semantik antara profil pelamar dan kebutuhan perusahaan. Visualisasi ini tidak hanya membantu dalam memahami distribusi fitur penting,

tetapi juga berfungsi sebagai dasar untuk memberikan rekomendasi perbaikan isi resume agar lebih selaras dengan tuntutan pasar kerja yang dinamis.

Dalam resume, kata-kata seperti "*internship*", "*experience*", dan "*management*" tampak dominan. Hal ini menunjukkan bahwa banyak pencari kerja menekankan pengalaman magang, pengalaman kerja sebelumnya, dan keterampilan manajerial dalam profil mereka. Kata-kata seperti "*custom*", "*solution*", dan "*enterprise*" juga muncul, tetapi dengan ukuran yang relatif lebih kecil dibandingkan dengan kata-kata dalam deskripsi lowongan. Ini menunjukkan bahwa beberapa keterampilan atau konsep yang dimiliki oleh pencari kerja belum sepenuhnya sesuai dengan permintaan industri.

Dari word cloud dapat dilihat beberapa pola penting dari perbandingan antar job description dan resume, yaitu Ada beberapa kata yang muncul baik dalam deskripsi lowongan maupun resume, seperti "*google*", "*workspace*", dan "*cloud*". Hal ini menunjukkan bahwa ada sejumlah pencari kerja yang sudah memiliki keterampilan atau pengalaman yang relevan dengan persyaratan pekerjaan. Beberapa kata kunci yang sering muncul dalam deskripsi lowongan, seperti "*territory*", "*account*", dan "*digital*", tidak begitu dominan dalam resume. Ini menunjukkan bahwa pencari kerja masih kurang memiliki keterampilan atau pengalaman yang spesifik dalam area-area tersebut.

Hasil word cloud memberikan wawasan berharga tentang kesenjangan kompetensi antara pencari kerja dan persyaratan pekerjaan dengan kata-kata seperti "*google*", "*cloud*", dan "*digital*" menunjukkan bahwa industri saat ini sangat membutuhkan keterampilan teknologi modern. Namun, banyak pencari kerja belum sepenuhnya fokus pada keterampilan ini dalam resume mereka. Ada beberapa kata yang hanya muncul dalam salah satu dokumen saja, seperti "*territory*" dalam deskripsi lowongan dan "*internship*" dalam resume. Hal ini menunjukkan bahwa pencari kerja perlu lebih fokus pada keterampilan yang spesifik untuk posisi tertentu, sementara industri perlu menyediakan informasi yang lebih jelas tentang keterampilan yang dibutuhkan.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis kesesuaian antara deskripsi lowongan pekerjaan dan kemampuan pencari kerja menggunakan metode *Cosine Similarity*. Berdasarkan eksperimen yang dilakukan dengan dataset publik "*Resume and Job Description*" dari platform Kaggle, sistem berhasil mengukur tingkat kesesuaian berdasarkan representasi teks menggunakan TF-IDF dan perhitungan *cosine similarity*.

Hasil menunjukkan bahwa sebagian besar resume memiliki skor kesesuaian rendah terhadap deskripsi lowongan pekerjaan, dengan rata-rata skor berada di kisaran 0 hingga 0.05. Hal ini menunjukkan bahwa banyak pencari kerja belum sepenuhnya memenuhi kriteria atau kompetensi yang diminta dalam lowongan pekerjaan. Meskipun demikian, sistem mampu mengidentifikasi beberapa pasangan resume dan job description yang memiliki kesesuaian sedang hingga tinggi, dengan skor tertinggi mencapai 0.21.

Dengan pendekatan yang digunakan, yaitu kombinasi praproses teks, vektorisasi TF-IDF, dan perhitungan *cosine similarity*, sistem memberikan hasil yang objektif dan dapat diinterpretasikan secara intuitif. Selain itu, sistem juga mampu memberikan umpan balik berupa daftar keterampilan tambahan yang diperlukan oleh pencari kerja untuk meningkatkan relevansi mereka terhadap posisi tertentu. Pendekatan ini membantu pencari kerja dalam merancang strategi pengembangan karier yang lebih tepat sasaran serta membantu pemberi kerja dalam menyaring kandidat secara lebih efisien dan objektif.

Meskipun sistem yang dikembangkan dalam penelitian ini sudah memberikan hasil yang cukup baik, masih ada ruang untuk pengembangan lebih lanjut agar sistem menjadi lebih responsif dan akurat. Beberapa saran pengembangan seperti integrasi NLP dan embedding model, penyempurnaan visualisasi dan umpan balik, implementasi sistem rekomendasi dinamis. Dengan

potensi implementasi yang luas dan manfaat yang nyata, penelitian ini layak dikembangkan lebih lanjut untuk menjadi bagian dari solusi cerdas dalam dunia rekrutmen dan pengembangan karier.

DAFTAR PUSTAKA

- [1] A. D. Wibawa, A. M. Amri, A. Mas, and S. Iman, “Text Mining for Employee Candidates Automatic Profiling Based on Application Documents,” *Emit. Int. J. Eng. Technol.*, pp. 47–62, Apr. 2022, doi: 10.24003/emitter.v10i1.679.
- [2] H. Artajaya, Julieta, J. Giancarlos, J. V. Moniaga, and A. Chowanda, “Job Recommendation System based on Resume using Natural Language Processing and Distance-based Algorithm,” in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Bandung, Indonesia: IEEE, Feb. 2024, pp. 1–6. doi: 10.1109/AIMS61812.2024.10512474.
- [3] S. Kumari, “Job Recommendation System Using NLP,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 5, pp. 2721–2728, May 2023, doi: 10.22214/ijraset.2023.52183.
- [4] Sandeep Kulkarni, Prathmesh Rahul Kurumkar, Vansh Sanjeev Kadam, and Vinut Prabhu Maradur, “AI-powered resume screening system,” *World J. Adv. Eng. Technol. Sci.*, vol. 15, no. 1, pp. 2263–2277, Apr. 2025, doi: 10.30574/wjaets.2025.15.1.0413.
- [5] J. Rosenberger, L. Wolfrum, S. Weinzierl, M. Kraus, and P. Zschech, “CareerBERT: Matching resumes to ESCO jobs in a shared embedding space for generic job recommendations,” *Expert Syst. Appl.*, vol. 275, p. 127043, May 2025, doi: 10.1016/j.eswa.2025.127043.
- [6] M. J. Lavin, “Analyse de documents avec TF-IDF,” *Program. Hist. En Fr.*, no. 4, Jun. 2022, doi: 10.46430/phfr0022.
- [7] Y. Chen, X. Zhan, W. Yang, X. Yan, Y. Du, and T. Zhao, “A NLP analysis of digital demand for healthcare jobs in China,” *Sci. Rep.*, vol. 15, no. 1, p. 14518, Apr. 2025, doi: 10.1038/s41598-025-98552-5.
- [8] R. Zbib *et al.*, “Learning Job Titles Similarity from Noisy Skill Labels,” 2022, *arXiv*. doi: 10.48550/ARXIV.2207.00494.
- [9] D. O. Sihombing, “Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis,” *J. Sist. Komput. Dan Inform. JSON*, vol. 4, no. 2, p. 396, Dec. 2022, doi: 10.30865/json.v4i2.5374.
- [10] R. Moura, M. Richetto, D. Luche, L. Tozi, and M. Silva, “New Professional Competencies and Skills Leaning towards Industry 4.0,” in *Proceedings of the 14th International Conference on Computer Supported Education*, Online Streaming, --- Select a Country ---: SCITEPRESS - Science and Technology Publications, 2022, pp. 622–630. doi: 10.5220/0011047300003182.
- [11] W. Bittremieux, R. Schmid, F. Huber, J. J. Van Der Hooft, M. Wang, and P. C. Dorrestein, “Comparison of Cosine, Modified Cosine, and Neutral Loss Based Spectrum Alignment For Discovery of Structurally Related Molecules,” Jun. 02, 2022. doi: 10.1101/2022.06.01.494370.
- [12] Dr. Irfan Landge, Shaikh Mohammad Mehtab, Shaikh Farhan, and Shaikh Mohd Zahed, “Smart Resume Analyzer,” *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 116–119, May 2023, doi: 10.48175/ijarsct-9724.
- [13] M. W. Putri, A. Muchayan, and M. Kamisutara, “Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF,” *JOINTECS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 3, p. 229, Sep. 2020, doi: 10.31328/jointecs.v5i3.1563.
- [14] Ylber Januzaj and Artan Luma, “Cosine Similarity – A Computing Approach to Match Similarity Between Higher Education Programs and Job Market Demands Based on Maximum Number of Common Words,” *Int. J. Emerg. Technol. Learn. IJET*, vol. 17, no. 12, pp. 258–268, Jun. 2022, doi: 10.3991/ijet.v17i12.30375.

- [15] S. G. M. and S. K. Suganthi, "AI based suitability measurement and prediction between job description and job seeker profiles," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100109, Nov. 2022, doi: 10.1016/j.jjimei.2022.100109.
- [16] T. K. U. V. S. M. Kadiwal, and S. Revanna, "Design and development of machine learning based resume ranking system," *Glob. Transit. Proc.*, vol. 3, no. 2, pp. 371–375, Nov. 2022, doi: 10.1016/j.gltp.2021.10.002.
- [17] A. Espinal, Y. Haralambous, D. Bedart, and J. Puentes, "Uncertainty-Oriented Textual Marker Selection for Extracting Relevant Terms from Job Offers," in *Artificial Intelligence and Fuzzy Logic System*, Academy and Industry Research Collaboration Center (AIRCC), Sep. 2022, pp. 1–16. doi: 10.5121/csit.2022.121601.
- [18] S. Rojas-Galeano, J. Posada, and E. Ordoñez, "A Bibliometric Perspective on AI Research for Job-Résumé Matching," *Sci. World J.*, vol. 2022, pp. 1–15, Oct. 2022, doi: 10.1155/2022/8002363.
- [19] P. Rasal, "Resume Parser Analysis Using Machine Learning and Natural Language Processing," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 5, pp. 2840–2844, May 2023, doi: 10.22214/ijraset.2023.52202.