PENGEMBANGAN MODEL REKOMENDASI LOWONGAN KERJA BERDASARKAN KESESUAIAN KOMPETENSI PENCARI KERJA MENGGUNAKAN CONTENT-BASED FILTERING PADA PLATFORM PENCARIAN KERJA

Erlina Ayunda Cahyanti[1], Muhammad Galang Satrio Wicaksono[2]
[1,2] Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur
E-mail: erlinaayundac@gmail.com, muhammad_galang.sada@upnjatim.ac.id

ABSTRACT

Matching job seekers' competencies with job openings is one of the key aspects of the job search process. This study aims to apply and evaluate the Content-Based Filtering method to analyze the compatibility of candidate competency profiles with job openings on the Tloker platform. The contribution of this study lies in the application of competency term normalization to standardize variations in skill terminology, as well as the utilization of the Tloker dataset, which consists of 7,125 candidate profiles and 10,000 job opening records. Candidate profiles were constructed from educational and skill information, while job posting data included job titles, requirements, and job descriptions. The matching process utilized Term Frequency-Inverse Document Frequency (TF-IDF) weighting and Cosine Similarity to generate the top five recommendations. An evaluation was conducted on 30 candidates through manual relevance assessment using the Precision@5 and Mean Average Precision@5 (MAP@5) metrics. The test results showed an average Precision@5 of 66.67% and MAP@5 of 83.30% based on the implementation formula. The results show that the model is capable of generating fairly relevant recommendations, but text similarity does not yet fully reflect the alignment of competencies with job requirements. Thus, term normalization and matching based on TF-IDF and cosine similarity can aid the job vacancy recommendation process, although relevance evaluation remains necessary to assess the suitability of the recommended results.

Keywords: *Content-Based Filtering, Cosine Similarity, Competency Profile, Job Recommendations, TF-IDF*

ABSTRAK

Pencocokan antara kompetensi pencari kerja dan kebutuhan lowongan pekerjaan merupakan salah satu aspek penting dalam proses pencarian kerja. Penelitian ini bertujuan menerapkan dan mengevaluasi metode Content-Based Filtering untuk menganalisis kesesuaian profil kompetensi kandidat dengan lowongan pekerjaan pada platform Tloker. Kontribusi penelitian ini terletak pada penerapan normalisasi istilah kompetensi untuk menyeragamkan variasi penyebutan keterampilan serta pemanfaatan dataset Tloker yang terdiri atas 7.125 profil kandidat dan 10.000 data

Article History
Received: November 2026
Reviewed: November 2026
Published: November 2026

Plagiarism Checker No 237.u873247.fh8374

Copyright : Author
Publish by : Kohesi

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

lowongan pekerjaan. Profil kandidat dibentuk dari informasi pendidikan dan keterampilan, sedangkan data lowongan mencakup judul, persyaratan, dan deskripsi pekerjaan. Proses pencocokan dilakukan menggunakan pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) dan pengukuran kemiripan Cosine Similarity untuk menghasilkan lima rekomendasi teratas. Evaluasi dilakukan terhadap 30 kandidat melalui penilaian relevansi secara manual menggunakan metrik Precision@5 dan Mean Average Precision@5 (MAP@5). Hasil pengujian menunjukkan rata-rata Precision@5 sebesar 66,67% dan MAP@5 sebesar 83,30% berdasarkan rumus implementasi. Hasil tersebut menunjukkan bahwa model mampu menghasilkan rekomendasi yang cukup relevan, tetapi kemiripan teks belum sepenuhnya mencerminkan kesesuaian kompetensi dengan kebutuhan pekerjaan. Dengan demikian, normalisasi istilah dan pencocokan berbasis TF-IDF serta Cosine Similarity dapat membantu proses rekomendasi lowongan pekerjaan, meskipun evaluasi relevansi tetap diperlukan untuk menilai kesesuaian hasil rekomendasi.
Kata Kunci: *Content-Based Filtering, Cosine Similarity,* Profil Kompetensi, Rekomendasi Lowongan Pekerjaan, *TF-IDF*

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong perubahan dalam proses pencarian dan perekrutan tenaga kerja. Transformasi teknologi informasi telah mengubah lingkup pencarian pekerjaan. Banyaknya individu yang mencari pekerjaan melalui *platform online* (Dillahunt et al., 2021). Kemudahan akses terhadap informasi lowongan pekerjaan membuat jumlah lowongan yang tersedia semakin banyak dan beragam sehingga memberikan lebih banyak pilihan bagi pencari kerja (Ramadhani et al., 2026). Untuk menjembatani antara perusahaan dan calon pekerja, berbagai *platform* pencarian kerja digital hadir dan berkembang pesat menjadi pilihan utama masyarakat, salah satunya adalah Tloker. Tloker merupakan platform pencarian kerja berbasis digital yang mempertemukan pencari kerja dengan berbagai informasi lowongan pekerjaan dari perusahaan di Indonesia. Platform ini menyediakan informasi lowongan berdasarkan posisi, lokasi, bidang pekerjaan, serta kualifikasi yang dibutuhkan. Selain itu, pencari kerja dapat melengkapi profil karier yang mencakup informasi pribadi, pendidikan, pengalaman, keterampilan, dan preferensi pekerjaan. Sementara itu, informasi lowongan pekerjaan memuat sejumlah atribut, seperti judul pekerjaan, nama perusahaan, lokasi, jenis pekerjaan, persyaratan atau kualifikasi kandidat, serta deskripsi pekerjaan (*Tloker - Cari Lowongan Pekerjaan Terbaru*, n.d.).

Banyaknya tawaran pekerjaan yang ditawarkan oleh platform, menyebabkan banyaknya informasi, terutama bagi lulusan baru yang belum memiliki banyak pengalaman (Sarsenbay et al., 2025). Sebagian besar platform hanya menampilkan informasi deskriptif dan daftar lowongan, tanpa memberikan konteks yang jelas tentang seberapa dekat profil kompetensi seseorang dengan persyaratan pekerjaan (Tang et al., 2025). Pengguna diharuskan melakukan pencocokan secara manual terhadap dekripsi pekerjaan, kualifikasi, pengalaman, serta

persyaratan lainnya. Proses tersebut memerlukan waktu yang cukup lama dan berpotensi menghasilkan pemilihan pekerjaan yang kurang sesuai dengan komptensi pengguna (Ramadhani et al., 2026).

Oleh karena itu, diperlukan sebuah solusi teknologi cerdas yang mampu mengotomatisasi penyaringan informasi agar proses perncarian kerja menjadi lebih efektif. Salah satu pendekatan yang dapat digunakan untuk menganalisis kesesuaian rekomendasi tersebut adalah *Content-Based Filtering*. Pendekatan ini terbukti efektif dalam merekomendasikan pekerjaan dengan mencocokkan keterampilan pengguna terhadap deskripsi pekerjaan (Koloman et al., 2023). Secara umum, penerapan *Content-Based Filtering* dapat dilakukan melalui berbagai pendekatan, antara lain pembobotan berbasis teks menggunakan TF-IDF (Hersianty et al., 2025). Masing-masing pendekatan tersebut telah menunjukkan hasil yyang positif dalam memberikan rekomendasi pekerjaan yang relevan.

Penelitian yang dilakukan oleh Fitria et al. Mengembangkan sistem rekomendasi menggunakan ontent-Based Filtering dengan TF-IDF dan Cosine Similarity, tetapi masih menemukan permasalahan dalam pencocokan pekerjaan ketika kesamaan keterampilan menyebabkan rekomendasi yang berbeda dari posisi pekerjaan yang diuji (Fitria et al., 2024). Selanjutnya, Hersianty et al. menerapkan TF-IDF dan *Cosine Similarity* untuk mencocokan keahlian pencari kerja dengan pengujian yang melibatkan lima profil pencari kerja dan lima belas data lowongan, serta menyarankan penggunaan data yang lebih bervariasi dan pengembangan evaluasi lebih lanjut (Hersianty et al., 2025). Sementara itu, Rahmawati et al. mengembangkan pendekatan yang lebih kompleks dengan mengintegrasikan *Content-Based Filtering, Sentence-BERT*, analisis kesenjangan keterampi*lan (skill gap analysis),* dan *Retrieval-Augmented Generation* (RAG), tetapi masih mengidentifikasi perlunya perluasan sampel pengujian, penambahan jumlah penilai dari praktisi *Human Resource* (HR), serta pengujian pada data pekerjaan dengan konteks industri yang lebih beragam (Rahmawati et al., 2026).

Berdasarkan penelitian-penelitian tersebut, masih terdapat ruang untuk mengkaji penerapan dan evaluasi pencocokan profil kompetensi pencari kerja dengan informasi karakteristik yang berbeda khususnya pada platform Tloker. Oleh karena itu, penelitian ini berfokus pada ppenerapan TF-IDF dan *Cosine Similarity* dengan penambahan normalisasi istilah kompetensi serta evaluasi relevansi pemeringkatan rekomendasi menggunakan *Precision@5* dan *Mean Average Precision@5* berdasarkan penilaian manual.

2. METODE PENELITIAN

2.1 Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komputasional yang berfokus pada penerapan dan evaluasi model sistem rekomendasi lowongan pekerjaan. Pendekatan ini dipilih karena penelitian melibatkan pengolahan data teks, perhitungan tingkat kemiripan antar profil komptensi pencari kerja dan informasi lowongan pekerjaan, serta pengukuran relevansi hasil rekomendasi menggunakan metrik evaluasi. Rancangan penelitian meliputi eksplorasi data, preprocessing, normalisasi istilah, pembentukan representasi teks, pemodelan rekomendasi, dan evaluasi hasil rekomendasi.

2.2 Objek dan Sumber Data Penelitian

Objek penelitian ini adalah data profil pencari kerja dan data lowongan pekerjaan yang tersedia pada platform Tloker. Data penelitian diperoleh melalui proses penarikan data secara langsung dari platform Tloker (*data extraction*. Data yang dikumpulkan berupa data sekunder dalam format terstruktur yang kemudian digunakan untuk membangun model rekomendasi lowongan pekerjaan berdasarkan kesesuaian kompetensi kandidat.

Data yang digunakan dalam penelitian terdiri atas dua kelompok utama, yaitu data profil kandidat dan data lowongan pekerjaan. Data profil kandidat diperoleh dari tabel profil, pendidikan, dan keterampilan yang tersedia dalam dataset Tloker. Informasi pendidikan digunakan untuk menggambarkan latar belakang akademik kandidat, sedangkan informasi keterampilan digunakan untuk merepresentasikan kompetensi yang dimiliki. Dalam penelitian ini, representasi profil kandidat dibentuk dengan menggabungkan informasi pendidikan dan keterampilan. Adapun data lowongan pekerjaan diperoleh dari tabel lowongan yang memuat informasi judul pekerjaan (*title*), persyaratan (*requirement*), dan deskripsi pekerjaan (*description*). Ketiga atribut tersebut digabungkan untuk membentuk representasi teks setiap lowongan pekerjaan. Dataset yang digunakan dalam proses pemodelan terdiri atas 7.125 profil kandidat dan 10.000 data lowongan pekerjaan. Karakteristik data yang digunakan dalam penelitian disajikan pada Tabel berikut.

Tabel 2.1 Variabel dan Sumber Data

Variabel	Jenis variabel	Sumber data	Keterangan
Pendidikan kandidat	Masukan	Data pendidikan kandidat	Informasi latar belakang pendidikan kandidat
Keterampilan kandidat	Masukan	Data keterampilan kandidat	Informasi keterampilan atau kompetensi yang dimiliki kandidat
Judul pekerjaan	Masukan	Data lowongan pekerjaan	Nama atau posisi pekerjaan yang ditawarkan
Persyaratan pekerjaan	Masukan	Data lowongan pekerjaan	Kualifikasi dan persyaratan yang dibutuhkan perusahaan
Deskripsi pekerjaan	Masukan	Data lowongan pekerjaan	Informasi mengenai tugas dan tanggung jawab pekerjaan

2.3 Langkah-Langkah Penelitian

Penelitian dilakukan melalui beberapa tahapan. Alur penelitian secara garis besar disajikan pada Gambar 2.1 berikut.



Gambar 2.1 Diagram Alur Penelitian

Tahapan penelitian dimulai dengan pengumpulan dataset yang terdiri atas data profil pencari kerja dan data lowongan pekerjaan, dilanjutkan dengan eksplorasi data untuk memeriksa struktur, kelengkapan, serta hubungan antar dataset. Selanjutnya, dilakukan preprocessing yang meliputi pembersihan data dan normalisasi teks. Data teks profil kandidat

dan informasi lowongan kemudian ditransformasikan menjadi representasi numerik menggunakan metode TF-IDF melalui proses vektorisasi. Selanjutnya, dilakukan perhitungan kemiripan antara vektor profil kandidat dan vektor lowongan menggunakan *Cosine Similarity*. Berdasarkan skor tersebut, lowongan diurutkan dari tingkat kemiripan tertinggi untuk menghasilkan lima rekomendasi teratas. Tahap terakhir adalah evaluasi model untuk menganalisis proporsi serta kualitas urutan rekomendasi yang relevan.

2.4 Preprocessing Data

Sebelum proses pemodelan dilakukan, data terlebih dahulu melalui tahap pengolahan untuk membentuk representasi teks dari masing-masing kelompok data. Data pendidikan dan keterampilan kandidat yang berasal dari tabel berbeda digabungkan berdasarkan identitas kandidat (*user_id*) sehingga menghasilkan satu representasi teks untuk setiap kandidat. Sementara itu, data lowongan pekerjaan dibentuk dengan menggabungkan atribut judul, persyaratan, dan deskripsi pekerjaan. Data yang tidak memiliki informasi teks yang dapat digunakan dalam proses pencocokan tidak disertakan dalam pemodelan. Adapun data yang memiliki sebagian atribut kosong tetap dapat digunakan selama atribut lainnya masih menyediakan informasi yang dapat direpresentasikan sebagai teks.

Penanganan data duplikat dilakukan untuk menghindari adanya rekomendasi dengan informasi lowongan yang sama. Proses penghapusan duplikasi diterapkan pada tahap pemeringkatan dengan mengidentifikasi kesamaan gabungan teks antara lowongan pekerjaan. Selain itu, hanya lowongan dengan skor *Cosine Similarity* lebih dari nol yang dipertimbangkan sebagai kandidat rekomendasi. Sistem kemudian mengurutkan lowongan berdasarkan skor kemiripan tertinggi dan mengambil maksimal lima lowongan sebagai hasil rekomendasi. Apabila jumlah lowongan unik yang memenuhi kriteria kurang dari lima, jumlah rekomendasi yang dihasilkan disesuaikan dengan data yang tersedia.

2.5 Pemodelan *Content-Based Filtering*

Pemodelan dilakukan menggunakan pendekatan *Content-Based Filtering,* yaitu metode rekomendasi yang memberikan saran item berdasarkan seberapa mirip karakteristik antara item yang direkomendasikan dan item pengguna (Firdausa, 2021). Dalam penelitian ini, profil kompetensii kandidat dibandingkan denga teks informasi lowongan pekerjaan melalui beberapa tahapan.

2.6 Pembobotan TF-IDF

Algoritma *Term Frequency Inverse Document Frequency* (TF-IDF) merupakan bidang information retrieval yang digunakan untuk menghitung seberapa relevan sebuah kata dalam rangkaian korpus dengan sebuah teks. Algoritma ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan dokumen yang mengandung kata. Algoritma ini akan menghitung bobot setiap *term* di dokumen dengan persamaan 1 sebagai berikut (Sintia et al., 2021):

$$tf_{t,d} = 1 + log_{tf} \quad (2.1)$$

Keterangan:
tf = *term* frekuensi atau banyaknya kata pada dokumen
$tf_{(t,d)}$ = *term* frekuensi atau banyaknya kata t pada dokumen d atau pembobotan lokal

Lalu persamaan 2 untuk mencari nilai *inverse document frequency:*

$$idf_{(t)} = \log\left(\frac{N}{df_{(t)}}\right) \quad (2.2)$$

Keterangan:
$idf_{(t)}$: *inverse document frequency* dari term t.
N: jumlah keseluruhan dokumen.
$df_{(t)}$: jumlah dokumen yang memuat term t.

Berdasarkan persamaan 1 dan 2 diatas, dapat ditentukan nilai bobotnya dengan mengalikan kedua persamaan sehingga:

$$W_{(t,d)} = tf_{(t,d)} * idf_t \quad (2.3)$$

Keterangan:
$W_{(t,d)}$: bobot term t pada dokumen d.
$tf_{(t,d)}$: *term frequency*.
idf_t: *inverse document frequency* dari term t.
$df_{(t)}$: jumlah dokumen yang memuat term t.

2.7 Perhitungan *Cosine Similarity*

Cosine Similarity adalah salah satu ukuran kemiripan paling popular yang diterapkan di dokumen teks. Algoritma ini digunakan untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek. Secara umum perhitungan algoritma ini didasarkan pada *vector space similarity measure*. Algoritma *Cosine Similarity* ini menghitung kemiripan antar dua buah objek yang dinyatakan dalam dua buah vektor menggunakan kata kunci dari sebuah dokumen sebagai ukuran. Rumus perhitungan *Cosine Similarity:*

$$\text{Cos } a = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i \times B}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (2.4)$$

Keterangan:
A: Vektor *A*, yang akan dibandingkan kemiripannya.
B: Vektor *B*, yang akan dibandingkan kemiripannya.
$A \cdot B$: *dot product* antara vektor *A* dan *B*.
$|A|$: Panjang (norma) vektor *A*.
$|B|$: Panjang (norma) vektor *B*.
$|A||B|$: Perkalian antara panjang vektor $|A|$ dan $|B|$.
n: Jumlah dimensi/fitur dari vektor.
A_i: Bobot/nilai komponen ke-i pada vektor *A*.
B_i: Bobot/nilai komponen ke-i pada vektor *B*.

2.8 Pembentukan Rekomendasi

Hasil perhitungan *Cosine Similarity* diurutkan berdasarkan skor kemiripan dari nilai tertinggi hingga terendah. Lowongan dengan skor positif dipertimbangkan sebagai kandidat rekomendasi. Selanjutnya, lima lowongan dengan skor kemiripan tertinggi dipilih sebagai rekomendasi utama. Untuk menghindari pengulangan informasi, dilakukan penyaringan terhadap rekomendasi yang memiliki teks lowongan identik pada masing-masiing kandidat. Apabila jumlah lowongan unik yang memenuhi kriteria kurang dari lima, sistem menampilkan jumlah rekomendasi yang tersedia.

2.9 Evaluasi Model

Evaluasi model dilakukan untuk mengukur kesesuaian hasil rekomendasi terhadap kompetensi kandidat. *Precision@5* digunakan untuk mengukur proporsi lowongan relevan dari lima rekomendasi teratas yang dihasilkan untuk setiap kandidat. Perhitungan dilakukan dengan membagi jumlah rekomendasi relevan dalam lima posisi teratas dengan nilai lima. Apabila kandidat memperoleh kurang dari lima rekomendasi, jumlah rekomendasi yang tidak tersedia tetap diperhitungkan sebagai posisi yang tidak relevan.

Average Precision@5 (AP@5) digunakan untuk mengukur kualitas urutan rekomendasi dengan mempertimbangkan posisi kemunculan lowongan yang relevan. Perhitungan dilakukan dengan menjumlahkan nilai presisi kumulatif pada setiap posisi rekomendasi yang relevan, kemudian membaginya dengan jumlah rekomendasi relevan yang ditemukan dalam lima posisi teratas. Apabila tidak ditemukan rekomendasi relevan, nilai AP@5 ditetapkan sebesar nol (Yulita et al., 2023).

$$\text{Precision@}K = \frac{\text{Jumlah rekomendasi relevan pada top-}K}{K} \tag{2.5}$$

Keterangan:
Precision@K: nilai *precision* pada K rekomendasi paling atas.
K: banyaknya rekomendasi yang ditampilkan (contohnya 5).

$$\text{AP} = \frac{1}{R}\sum_{k=1}^{n} \text{Precision@}\, k \times \text{rel}(k) \tag{2.6}$$

Keterangan:
AP: nilai *Average Precision* untuk satu *query*.
R: jumlah dokumen relevan pada hasil.
n: jumlah total dokumen yang ditampilkan.
Precision@K: nilai *precision* pada posisi ke-k.
rel(k): bernilai 1 apabila dokumen pada posisi ke-k dinyatakan relevan, dan 0 apabila tidak relevan.

Mean Average Precision@5 (MAP@5) merupakan rata-rata nilai AP@5 dari seluruh kandidat yang dievaluasi. Metrik ini digunakan untuk memperoleh gambaran kualitas pengurutan rekomendasi pada keseluruhan sampel kandidat. Perhitungan dilakukan dengan menjumlahkan nilai AP@5 setiap kandidat dan membaginya dengan jumlah kandidat yang dievaluasi (Sibuea & Widodo, 2024).

$$\text{MAP} = \frac{1}{Q}\sum_{q=1}^{Q} \text{AP}_q \tag{2.7}$$

Keterangan:
MAP: nilai *Mean Average Precision*.
Q: keseluruhan jumlah *query* yang diujikan.
AP$_q$: nilai *AP* pada *query* ke-q.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Persiapan Data Pemodelan

Berdasarkan hasil preprocessing, data yang digunakan dalam pemodelan terdiri atas 7.125 profil kandidat dan 10.000 data lowongan pekerjaan. Profil kandiidat dibentuk dari

penggabungan informasi pendidikan dan keterampilan, sedangkan informasi lowongan dibentuk dari judul, persyaratan, dan deskripsi pekerjaan. Pemeriksaan menunjukkan bahwa seluruh profil kandidat dan teks lowongan telah memiliki teks yang dapat digunakan dalam proses pemodelan.

Tabel 3.1 Ringkasan Data Pemodelan

Data	Jumlah
Profil kandidat	7.125
Lowongan pekerjaan	10.000
Kolom kandidat model	5
Kolom lowongan model	9
Profil kandidat kosong	0
Teks lowongan kosong	0

3.2 Hasil Normalisasi Istilah

Normalisasi istilah kompetensi dilakukan untuk menyamakan variasi penyebutan keterampilan yang memiliki makna serupa ke dalam satu istilah baku. Proses normalisasi menggunakan kamus yang disusun secara manual berdasarkan variasi istilah yang ditemukan pada data profil kandidat dan lowongan pekerjaan. Penyusunan kamus mempertimbangkan kesamaan makna kompetensi, penggunaan sinonim, serta perbedaan istilah dalam bahasa Indonesia dan bahasa Inggris. Berdasarkan implementasi yang digunakan, kamus normalisasi terdiri atas 16 variasi istilah yang dikelompokkan ke dalam lima istilah kompetensi baku, yaitu MENGEMUDI, MENJAHIT, MENGELAS, MEMASAK, dan AKUNTANSI. Rincian pemetaan istilah disajikan pada Tabel 3.2.

Tabel 3.2 Contoh Normalisasi Istilah

Istilah Sebelum	Istilah Setelah
DRIVER, SUPIR, SOPIR	MENGEMUDI
MENYETIR	MENGEMUDI
MENGENDARAI KENDARAAN	MENGEMUDI
PENJAHIT, SEWING	MENJAHIT
WELDING, WELDER	MENGELAS
JURU LAS	MENGELAS
COOK, CHEF	MEMASAK
JURU MASAK	MEMASAK
ACCOUNTING	AKUNTANSI

Proses penggantian dilakukan menggunakan pencocokan pola kata utuh (*whole-word pattern matching*) untuk memastikan penggantian hanya dilakukan pada istilah yang sesuai, sedangkan istilah di luar kamus tetap dipertahankan. Proses ini bertujuan meningkatkan konsistensi representasi teks sebelum pembobotan TF-IDF dan perhitungan *Cosine Similarity*. Namun, normalisasi masih terbatas pada istilah yang tercantum dalam kamus sehingga variasi lain yang belum terdaftar belum dapat diseragamkan.

3.3 Hasil Pembobotan TF-IDF

Teks profil kandidat dan teks lowongan ditransformasikan menjadi representasi numerik menggunakan metoode TF-IDF. Proses vektorisasi dilakukan menggunakan pustaka *Scikit-learn*

dengan parameter *TfidfVectorizer*. Proses pembobotan dilakukan terhadap gabungan teks kandidat dan lowongan untuk membangun kosakata yang sama. Matriks hasil transformasi kemudian dipisahkan menjadi matriks kandidat dan matriks lowongan.

Tabel 3.3 Ukuran Matriks TF-IDF

Matriks	Jumlah Baris	Jumlah Fitur
TF-IDF kandidat	7.125	20.000
TF-IDF lowongan	10.000	20.000

Hasil tersebut menunjukkan bahwa seluruh profil kandidat dan lowongan direpresentasikan dalam ruang fitur yang sama. Dengan demikian, setiap vektor profil kandidat dapat dibandingkan dengan seluruh vektor lowongan menggunakan perhitungan kemiripan.

3.4 Hasil Perhitungan *Cosine Similarity* dan Rekomendasi Lowongan

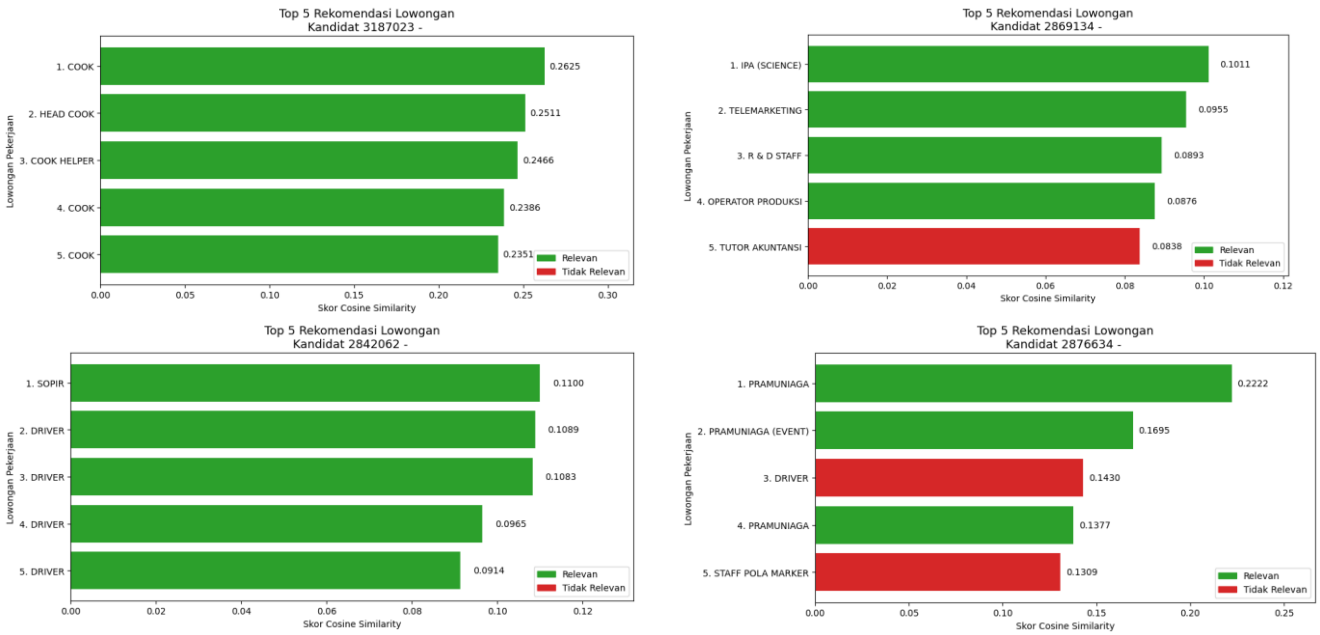
Perhitungan *Cosine Similarity* dilakukan dengan membandingkan vektor TF-IDF setiap kandidat terhadap seluruh vektor lowongan. Hasil perhitungan berupa skor kemirippan yang diguunakan untuk mengurutkan lowongan dari nilai tertinggi hingga terendah.

Pengujian awal dilakukan terhadap enam kandidat dengan karakteristik kompetensi berbeda, yaitu mengemudi dan menjahit, komputer, mengelas, memasak, marketing, serta komputer dan akuntansi.

Tabel 3.4 Contoh Hasil Rekomendasi Lowongan

Profil Kandidat	Rekomendasi Peringkat Pertama	Skor Kemiripan
MENGEMUDI / MENJAHIT	PENJAHIT	0,347931
KOMPUTER	GURU KOMPUTER	0,167259
MENGELAS	ENGINEERING SUPERVISOR	0,309160
MEMASAK	COOK	0,292829
MARKETING	MARKETING COMMUNICATION OFFICER	0,174041
KOMPUTER AKUNTANSI	STOCK CONTROLLER SPV	0,339017

Berdasarkan Tabel 3.4, model menghasilkan rekomendasi dengan skor kemiripan berbeda sesuai karakteristik teks profil kandidat. Sebagai contoh, kandidat dengan profil MENGEMUDI / MENJAHIT memperoleh posisi PENJAHIT sebagai rekomendasi pertama dengan skor 0,347931. Sementara itu, kandidat dengan profil MEMASAK memperoleh posisi COOK dengan skor 0,292829. Meskipun demikian, skor kemiripan yang tinggi belum selalu menunjukkan kesesuaian pekerjaan secara langsung. Kandidat dengan kompetensi MENGELAS memperoleh *ENGINEERING SUPERVISOR* sebagai rekomendasi pertama (0,309160), sedangkan *WELDING* berada pada peringkat keempat (0,153682). Hal ini terjadi karena model menghitung kemiripan berdasarkan keseluruhan teks profil kandidat dan informasi lowongan, bukan hanya istilah kompetensi pada judul pekerjaan. TF-IDF dan *Cosine Similarity* mengukur kesamaan kata tanpa memahami konteks serta tanggung jawab pekerjaan. Akibatnya, skor kemiripan tertinggi belum tentu menunjukkan lowongan yang paling sesuai dengan kompetensi kandidat.



Gambar 3.1 Contoh Hasil Rekomendasi Lowongan

Pada kandidat ID 3187023, lima lowongan yang direkomendasikan, yaitu COOK, HEAD COOK, COOK HELPER, dan COOK lainnya, seluruhnya dinilai relevan dengan profil kandidat yang memiliki keterampilan memasak. Kandidat ID 2869134 memperoleh empat rekomendasi relevan, yaitu IPA (SCIENCE), TELEMARKETING, R&D STAFF, dan OPERATOR PRODUKSI, sedangkan TUTOR AKUNTANSI dinilai tidak relevan. Selanjutnya, kandidat ID 2842062 memperoleh lima rekomendasi yang seluruhnya relevan, dengan posisi SOPIR dan DRIVER mendominasi daftar rekomendasi sesuai kompetensi mengemudi yang dimiliki. Sementara itu, kandidat ID 2876634 memperoleh tiga rekomendasi relevan, yaitu PRAMUNIAGA dan PRAMUNIAGA (EVENT), sedangkan DRIVER dan STAFF POLA MARKER dinilai tidak relevan.

Perbedaan hasil tersebut menunjukkan bahwa metode TF-IDF dan Cosine Similarity mampu menghasilkan rekomendasi yang sesuai dengan kompetensi tertentu, tetapi belum sepenuhnya menjamin seluruh lowongan yang direkomendasikan relevan. Dengan demikian, skor kemiripan digunakan untuk menentukan urutan rekomendasi, sedangkan penilaian manual tetap diperlukan untuk memastikan kesesuaian antara kompetensi kandidat dan kebutuhan pekerjaan.

3.5 Hasil Evaluasi Model

Evaluasi model dilakukan melaui pengujian terhadao 30 kandidat yang dipilih secara acak. Setiap kandidat memperoleh maksimal lima rekomendasi pberdasarkan skor kemiripan. Dari 30 kandidat yang diuji, dihasilkan 149 rekomendasi. Sebanyak 29 kandidat memperoleh lima rekomendasi, sedangkan satu kandidat hanya memperoleh empat rekomendasi karena jumlah lowongan unik dengan skor kemiripan positif yang tersedia setelah penyaringan duplikasi tidak mencapai lima. Berdasarkan label relevansi yang tercatat dalam notebook, terdapat 100 rekomendasi yang dinilai relevan dan 49 rekomendasi yang dinilai tidak relevan.

Tabel 3.5 Distribusi Label Relevansi

Label	Keterangan	Jumlah
1	Relevan	100

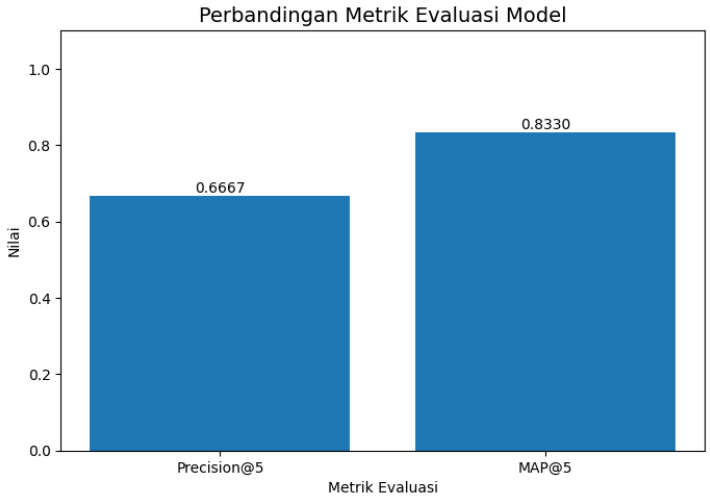
0	Tidak relevan	49
Total		**149**

Hasil evaluasi kemudian dihitung menggunakan Precision@5 dan Average Precision@5 (AP@5), yang selanjutnya dirata-ratakan untuk memperoleh MAP@5.

Tabel 3.6 Ringkasan Evaluasi Model

Metrik	Nilai
Jumlah kandidat	30
Jumlah rekomendasi	149
Rata-rata Precision@5	0,6667
Rata-rata Precision@5 (%)	66,67%
MAP@5 berdasarkan rumus implementasi	0,8330
MAP@5 (%)	83,30%

Nilai Precision@5 sebesar 0,6667 diperoleh dari rata-rata proporsi rekomendasi relevan dalam lima posisi teratas pada setiap kandidat. Sementara itu, nilai MAP@5 sebesar 0,8330 merupakan rata-rata AP@5 dari 30 kandidat.



Gambar 3.2 Perbandingan Metrik Evaluasi Model

Hasil tersebut menunjukkan bahwa model dapat menghasilkan rekomendasi yang dinilai relevan berdasarkan sebagian besar penilaian manual pada sampel pengujian. Namun, nilai evaluasi tersebut perlu diinterpretasikan sesuai dengan karakteristik sampel dan prosedur penilaian relevansi yang digunakan.

KESIMPULAN

Penelitian ini berhasil menerapkan metode *Content-Based Filtering* menggunakan TF-IDF dan *Cosine Similarity* untuk menganalisis kesesuaian profil kompetensi pencari kerja dengan lowongan pekerjaan pada platform Tloker. Model menggunakan 7.125 profil kandidat berdasarkan informasi pendidikan dan keterampilan serta 10.000 data lowongan pekerjaan yang mencakup judul, persyaratan, dan deskripsi pekerjaan. Normalisasi istilah kompetensi diterapkan untuk menyamakan variasi penyebutan keterampilan sehingga proses pencocokan teks dapat dilakukan secara lebih konsisten.

Hasil pengujian terhadap 30 kandidat memperoleh rata-rata *Precision@5* sebesar 66,67% dan *MAP@5* berdasarkan rumus implementasi sebesar 83,30%. Hasil tersebut menunjukkan bahwa model mampu menghasilkan rekomendasi yang cukup relevan, meskipun tingkat

kesesuaiannya berbeda pada setiap kandidat. Skor *Cosine Similarity* dapat digunakan untuk mengurutkan lowongan berdasarkan kemiripan teks, tetapi belum sepenuhnya menjamin kesesuaian kompetensi kandidat dengan kebutuhan pekerjaan. Penelitian ini masih memiliki keterbatasan, yaitu belum menggunakan informasi pengalaman kerja, normalisasi istilah masih bergantung pada kamus manual, dan evaluasi relevansi hanya dilakukan terhadap 30 kandidat.

Untuk pengembangan selanjutnya, disarankan agar informasi pengalaman kerja, sertifikasi, dan kompetensi lainnya ditambahkan ke dalam profil kandidat agar pencocokan menjadi lebih menyeluruh. Normalisasi istilah juga dapat dikembangkan menggunakan pendekatan pemrosesan bahasa alami, sedangkan evaluasi perlu melibatkan lebih banyak kandidat dan penilai dari praktisi Human Resource agar hasilnya lebih komprehensif. Selain itu, metode TF-IDF dan *Cosine Similarity* dapat dibandingkan dengan pendekatan berbasis representasi semantik untuk meningkatkan kesesuaian rekomendasi lowongan pekerjaan.

DAFTAR PUSTAKA

Dillahunt, T. R., Israni, A., Jiahong Lu, A., Cai, M., & Chiao-Yin Hsiao, J. (2021). *Examining the Use of Online Platforms for Employ-ment: A Survey of U.S. Job Seekers*. https://doi.org/10.1145/3411764.3445350

Firdausa, M. B. (2021). *Pengembangan Sistem Informasi Manajemen Keuangan Lesehan & Kolam Pancing Kresna Berbasis Website Menggunakan Model Evolutionary Prototype*.

Fitria, A., #2, Z., Ainul, M., & #3, Y. (2024). Sistem Rekomendasi Lowongan Pekerjaan Menggunakan Content-Based Filtering. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, *10*(3), 421–427. https://doi.org/10.26418/JP.V10I3.83801

Hersianty, V. M., Amalia, E. L., Puspitasari, D., & Wibowo, D. W. (2025). PENERAPAN ALGORITMA TF-IDF DAN COSINE SIMILARITY DALAM SISTEM REKOMENDASI LOWONGAN PEKERJAAN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(1), 1619–1625. https://doi.org/10.36040/JATI.V9I1.12406

Koloman, C., Maulana, R., Dwi, R., Putri, Z., & Harahap, W. A. (2023). Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering. *Journal of Creative Student Research*, *1*(6), 78–88. https://doi.org/10.55606/JCSRPOLITAMA.V1I6.2992

Rahmawati, I. Y., Mufarihati, F., Sri, C., & Aditya, K. (2026). Job Recommendation for Fresh Graduates to Reduce Competency Gaps Using Content-Based Filtering and Retrieval-Augmented Generation. *Jurnal Teknik Informatika (Jutif)*, *7*(2), 1886–1904. https://doi.org/10.52436/1.JUTIF.2026.7.2.5723

Ramadhani, S., Fiska, R. R., Informatika, T., Bengkalis, N., Alam, J. B., Alam, D. S., Bengkalis, K., Bengkalis, K., & Riau, P. (2026). SISTEM REKOMENDASI PEMILIHAN PEKERJAAN YANG RELEVAN BERBASIS CONTENT-BASED FILTERING DAN SEMANTIC MATCHING. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *10*(5), 8311–8315. https://doi.org/10.36040/JATI.V10I5.19200

Sarsenbay, S., Kabdiyev, A., Varlamis, I., Sardianos, C., Turan, C., Razhametov, B., & Kazym, Y. (2025). Generating Job Recommendations Based on User Personality and Gallup Tests. *Algorithms 2025, Vol. 18, Page 275*, *18*(5), 275. https://doi.org/10.3390/A18050275

Sibuea, S., & Widodo, Y. B. (2024). Pengembangan Model Machine Learning untuk Rekomendasi Produk Berdasarkan Analisis Pola Pembelian. *Jurnal Teknologi Informatika Dan Komputer*, *10*(2), 567–583. https://doi.org/10.37012/JTIK.V10I2.2354

Sintia, Defit, S., & Nurcahyo, G. W. (2021). Product Codefication Accuracy With Cosine Similarity And Weighted Term Frequency And Inverse Document Frequency (TF-IDF). *Journal of Applied Engineering and Technological Science (JAETS)*, *2*(2), 62–69. https://doi.org/10.37385/JAETS.V2I2.210

Tang, F., Zhu, R., Yao, F., Wang, J., Luo, L., & Li, B. (2025). Explainable person–job recommendations: challenges, approaches, and comparative analysis. *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/FRAI.2025.1660548/

Tloker - Cari Lowongan Pekerjaan Terbaru. (n.d.). Retrieved November 4, 2026, from https://Tloker.com/

Yulita, W., Cahyo Untoro, M., Praseptiawan, M., Ashari, I. F., Afriansyah, A., Bin, A. N., & Pee, C. (2023). Automatic Scoring Using Term Frequency Inverse Document Frequency Document Frequency and Cosine Similarity. *Scientific Journal of Informatics*, *10*(2), 93–104. https://doi.org/10.15294/SJI.V10I2.42209