

Analisis Metode Content Based Filtering Untuk Sistem Rekomendasi Kandidat Karyawan Pada Nextylabs

Tigar Riski Septiono^{1*}, Joni Maulindar², Afu Ichsan Pradana³

¹Teknik Informatika

(Universitas Duta Bangsa Surakarta)

^{1*}riskiseptiono0@gmail.com

²Teknik Informatika

(Universitas Duta Bangsa Surakarta)

²joni_maulindar@udb.ac.id

³Teknik Informatika

(Universitas Duta Bangsa Surakarta)

³afu_ichsan@udb.ac.id

Abstrak—Perkembangan teknologi digital menuntut efisiensi tinggi pada sistem seleksi talenta di industri perangkat lunak. Namun, proses penyeleksian berkas administrasi pada software house Nextylabs saat ini masih dilakukan secara manual terhadap dokumen Curriculum Vitae (CV) pelamar yang bersifat tidak terstruktur. Kondisi tersebut menimbulkan risiko tingginya tingkat subjektivitas, potensi terjadinya human error, serta inefisiensi waktu pemrosesan seleksi awal kandidat. Penelitian ini bertujuan merancang dan mengimplementasikan pendekatan berbasis Large Language Model (LLM) Gemini AI dan metode Content-Based Filtering (CBF) untuk mengotomatisasi sistem rekomendasi seleksi awal kandidat karyawan di Nextylabs. Penelitian ini menggunakan pendekatan eksperimental dengan tahapan yang meliputi pengumpulan data sampel berupa 15 dokumen CV, ekstraksi fitur keterampilan (skills) menggunakan model Gemini AI, pembersihan teks (preprocessing), representasi pembobotan fitur menggunakan TF-IDF, serta kalkulasi nilai perankingan berbasis Cosine Similarity. Hasil penelitian menunjukkan bahwa integrasi metode yang diusulkan mampu menyederhanakan informasi kompetensi pelamar dari berkas tidak terstruktur menjadi daftar urutan rekomendasi yang objektif. Sistem ini berhasil menghasilkan nilai kemiripan tertinggi sebesar 94,21% pada kandidat utama dan mencapai tingkat akurasi klasifikasi relevansi (Precision) sebesar 100%.

Kata kunci—Sistem Rekomendasi Kandidat Karyawan, Content-Based Filtering, Cosine Similarity, TF-IDF, Preprocessing, Gemini AI.

Abstract—The development of digital technology demands high efficiency in the talent selection system in the software industry. However, the administrative file selection process at the software house Nextylabs is currently still carried out manually on unstructured applicant Curriculum Vitae (CV) documents. This condition poses a high risk of subjectivity, potential human error, and inefficiency in processing time for initial candidate selection. This study aims to design and implement an approach based on Gemini AI's Large Language Model (LLM) and the Content-Based Filtering (CBF) method to automate the initial selection recommendation system for employee candidates at Nextylabs. This study uses an experimental approach with stages including collecting sample data in the form of 15 CV documents, extracting skill features using the Gemini AI model, text cleaning (preprocessing), representing feature weighting using TF-IDF, and calculating ranking values based on Cosine Similarity. The results show that the integration of the proposed methods is able to simplify applicant competency information from unstructured files into an objective list of recommendations. This system successfully produced the highest similarity value of 94.21% for the main candidate and achieved a relevance classification accuracy level (Precision) of 100%.

Keywords—Employee Candidate Recommendation System, Content Based Filtering, Cosine Similarity, TF-IDF, Pre-processing, Gemini AI.

I. PENDAHULUAN

Proses rekrutmen talenta di bidang teknologi pada perusahaan software house menuntut ketepatan dan efisiensi yang tinggi dalam memetakan kompetensi pelamar. Nextylabs, sebuah perusahaan berskala usaha kecil dan menengah (UKM) yang bergerak di bidang pengembangan perangkat lunak, menghadapi lonjakan volume dokumen Curriculum Vitae (CV) yang dikirimkan oleh pelamar kerja

melalui platform digital. Tim Human Resource Development (HRD) di Nextylabs saat ini masih melakukan proses skrining dan evaluasi kesesuaian dokumen CV tersebut secara manual satu per satu. Hambatan utama dari mekanisme konvensional ini adalah tingginya risiko subjektivitas penilaian, potensi terjadinya human error, serta inefisiensi waktu yang signifikan, mengingat dokumen CV yang masuk memiliki format teks yang tidak

terstruktur. Permasalahan penelitian yang diangkat adalah belum adanya sistem rekomendasi otomatis berbasis kompetensi yang terintegrasi pada proses rekrutmen di Nextylabs. Oleh karena itu, diperlukan sebuah sistem rekomendasi otomatis yang mampu mengekstraksi keahlian kandidat secara objektif guna mempercepat fase seleksi awal.

Beberapa penelitian terdahulu telah menerapkan metode Content-Based Filtering (CBF) untuk menyelesaikan kendala pencarian kecocokan objek. Larasati dan Februariyanti [1] mengimplementasikan metode CBF untuk sistem rekomendasi produk kosmetik berdasarkan preferensi pengguna. Sementara itu, Koloman dkk. [2] mengembangkan sistem rekomendasi pekerjaan di bidang teknologi informasi menggunakan algoritma CBF untuk mencocokkan keterampilan pengguna dengan deskripsi pekerjaan, namun penelitian tersebut masih mengandalkan input data masukan yang terstruktur secara manual tanpa adanya proses ekstraksi dokumen otomatis. Ridwansyah dkk. [3] memaparkan bahwa keunggulan penerapan CBF terletak pada kemampuannya memberikan rekomendasi item yang sederhana dan efisien berdasarkan kemiripan atribut, meskipun penerapannya dalam mengurai dokumen teks CV yang tidak terstruktur masih sangat terbatas.

Tujuan penelitian ini adalah merancang dan mengimplementasikan sistem rekomendasi kandidat karyawan menggunakan metode CBF dengan algoritma Cosine Similarity untuk meningkatkan efisiensi dan akurasi seleksi CV di Nextylabs. Kontribusi utama penelitian ini meliputi: [4] penerapan metode CBF pada domain rekrutmen karyawan khusus di perusahaan teknologi skala UKM, [5] integrasi ekstraksi informasi otomatis dari CV yang tidak terstruktur menggunakan teknologi Gemini AI, dan [1] penyediaan sistem perankingan kandidat otomatis bagi tim HRD. Kebaruan (novelty) dari penelitian ini terletak pada integrasi pemrosesan teks (preprocessing) yang menyeluruh dikombinasikan dengan ekstraksi fitur spesifik keahlian (skills) menggunakan model kecerdasan buatan, yang dirancang khusus untuk menangani variasi istilah teknis dalam dokumen CV pelamar kerja.

II. METODOLOGI PENELITIAN

Bagian ini menjelaskan metode yang digunakan dalam mengimplementasikan pendekatan analisis metode content based filtering untuk sistem rekomendasi kandidat karyawan pada nextylabs. Metode penelitian mencakup jenis dan sumber data, teknik pengumpulan data, serta metode yang digunakan dalam penelitian ini.

A. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini diklasifikasikan menjadi dua kategori utama, yaitu data primer dan data sekunder. Data primer diperoleh secara langsung dari objek penelitian di *Software House* Nextylabs untuk kebutuhan perancangan dan pengujian sistem, yang meliputi satu berkas dokumen teks resmi kriteria jabatan (*job description*) berisi deskripsi pekerjaan dan syarat kompetensi teknis posisi *Frontend Developer* sebagai target *query*, serta 15 dokumen *Curriculum Vitae* (CV) asli pelamar berformat PDF yang akan diekstraksi fitur kompetensinya menggunakan Gemini AI. Sementara itu, data sekunder diperoleh melalui studi pustaka untuk memperkuat landasan teori dan formulasi algoritma, yang bersumber dari jurnal ilmiah nasional maupun internasional terkait penerapan metode *Content-Based Filtering*. Data sekunder tersebut dilengkapi dengan dokumentasi resmi panduan integrasi *Application Programming Interface* (API) Gemini AI dari Google serta literatur akademis mengenai kalkulasi matematis pembobotan kata menggunakan TF-IDF dan perhitungan sudut vektor algoritma *Cosine Similarity*. Data ini digunakan sebagai dasar dalam perancangan pendekatan dan pemilihan metode yang digunakan. Adapun metode pengumpulan data dalam penelitian ini adalah sebagai berikut:

a. Observasi

Mengamati secara langsung alur kerja konvensional yang berjalan pada tim HRD Nextylabs saat menerima, membaca, dan menyeleksi dokumen *Curriculum Vitae* (CV) pelamar secara manual. Proses pengamatan ini difokuskan untuk mengidentifikasi inefisiensi waktu, potensi terjadinya *human error*, serta karakteristik format teks tidak terstruktur pada berkas-berkas CV yang masuk.

b. Wawancara

Wawancara dilakukan secara mendalam bersama tim *developer* dan *Human Resources*

- Development* (HRD) Nextylabs untuk menggali informasi komprehensif mengenai spesifikasi kebutuhan fungsional perangkat lunak, kendala-kendala teknis yang dihadapi selama proses seleksi manual, serta ekspektasi performa terhadap fitur otomatisasi perankingan yang akan dibangun.
- c. **Studi Pustaka**
Studi pustaka dilakukan dengan mengkaji dan menganalisis literatur ilmiah dari jurnal nasional maupun internasional, buku teks, dan dokumentasi resmi penunjang. Studi literatur ini difokuskan pada pemahaman parameter integrasi Application Programming Interface (API) Gemini AI sebagai pengekstraksi fitur teks, serta formulasi matematis dari metode pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) dan algoritma Cosine Similarity guna menyusun landasan perancangan sistem rekomendasi yang valid.
 - B. **Tahapan Penelitian**
Penelitian ini menggunakan pendekatan eksperimental dengan fokus pada implementasi metode Content-Based Filtering (CBF) untuk memberikan rekomendasi kandidat karyawan berdasarkan kesesuaian kompetensi yang tercantum dalam Curriculum Vitae (CV) dengan kebutuhan posisi pekerjaan. Tahapan penelitian dilakukan secara sistematis sebagai berikut:
 - a. **Pengumpulan Data**
Pada tahap ini dilakukan pengumpulan data berupa Curriculum Vitae (CV) kandidat karyawan serta data kebutuhan posisi pekerjaan yang tersedia pada perusahaan. Data yang dikumpulkan meliputi keterampilan (skills), pengalaman kerja, dan posisi yang dilamar sebagai acuan dalam proses rekomendasi.
 - b. **Ekstraksi Informasi CV Menggunakan Gemini AI**
CV kandidat yang masih berupa dokumen tidak terstruktur kemudian diproses menggunakan Gemini AI. Pada tahap ini Gemini AI digunakan untuk mengidentifikasi dan mengekstraksi informasi penting dari CV seperti nama kandidat, keterampilan, pengalaman kerja, pendidikan, sertifikasi, serta teknologi yang dikuasai[6],[7]. Hasil ekstraksi kemudian disusun ke dalam format yang lebih terstruktur sehingga dapat digunakan pada proses analisis selanjutnya.
 - c. **Preprocessing Data**
Data CV dan deskripsi pekerjaan yang telah dikumpulkan kemudian diproses untuk mempersiapkan data sebelum dilakukan perhitungan kemiripan. Tahapan preprocessing meliputi case folding (mengubah seluruh huruf menjadi huruf kecil), tokenization (memecah teks menjadi token kata), stopword removal (menghapus kata-kata umum yang tidak memiliki makna penting), serta stemming untuk mengubah kata menjadi bentuk dasarnya[8].
 - d. **Representasi Fitur Menggunakan TF-IDF**
Data yang telah melalui tahap preprocessing kemudian diubah menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF)[9]. Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan tingkat kemunculan dan tingkat kepentingannya dalam dokumen, sehingga menghasilkan vektor fitur yang merepresentasikan karakteristik masing-masing kandidat dan kebutuhan pekerjaan.
 - e. **Perhitungan Kemiripan Menggunakan Cosine Similarity**
Setelah diperoleh vektor TF-IDF, dilakukan perhitungan tingkat kemiripan antara profil kandidat dengan deskripsi pekerjaan menggunakan metode Cosine Similarity. Metode ini digunakan untuk mengukur tingkat kedekatan antara dua vektor dokumen berdasarkan sudut yang terbentuk, sehingga diperoleh nilai similarity yang menunjukkan tingkat kesesuaian kandidat terhadap posisi yang tersedia.
 - f. **Proses Rekomendasi Kandidat**
Berdasarkan nilai similarity yang diperoleh, sistem melakukan pemeringkatan (ranking) kandidat dari nilai kemiripan tertinggi hingga terendah. Kandidat dengan nilai similarity tertinggi direkomendasikan sebagai kandidat yang paling sesuai dengan kebutuhan posisi pekerjaan.
 - g. **Evaluasi Hasil**
Tahap akhir dilakukan evaluasi terhadap hasil rekomendasi yang dihasilkan sistem dengan menganalisis tingkat kesesuaian kandidat yang direkomendasikan terhadap kriteria posisi

pekerjaan. Evaluasi dilakukan secara kuantitatif menggunakan pengujian metrik kedekatan matematis dari nilai similarity yang diperoleh serta uji performa klasifikasi sistem berbasis Confusion Matrix guna mengukur tingkat ketepatan rekomendasi. Parameter evaluasi yang dihitung meliputi pengujian nilai Precision, Recall, dan F1-Score terhadap keseluruhan data peringkat hasil rekomendasi kandidat pelamar yang masuk ke dalam sistem. Nilai hasil perankingan tersebut divalidasi silang menggunakan data keputusan manual (data aktual) yang diverifikasi langsung terhadap seluruh berkas oleh tim Human Resource Development (HRD) dan developer ahli dari Software House Nextylabs selaku pengguna sistem akhir (User Acceptance Testing). Melalui kombinasi metode pengujian ini, performa sistem dapat diukur secara eksak untuk memastikan kemampuannya dalam membantu mempercepat proses pemindaian dokumen administrasi awal, meminimalkan bias subjektivitas pelamar, serta memangkas inefisiensi waktu secara signifikan jika dibandingkan dengan metode seleksi berkas manual konvensional yang berjalan sebelumnya.

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Proses pengumpulan data dilakukan melalui mekanisme pengunggahan berkas secara otomatis, di mana kandidat pelamar mengirimkan dokumen Curriculum Vitae (CV) melalui platform situs web rekrutmen Nextylabs. Berdasarkan proses tersebut, berhasil dihimpun sampel data sebanyak 15 dokumen CV pelamar dalam format PDF yang seluruhnya terarah pada posisi Frontend Developer. Pada fase awal ini, sistem akan langsung menampilkan keseluruhan data log pelamar tersebut kepada admin atau tim HRD yang ingin mencari perankingan kandidat guna keperluan seleksi pertama. Data pelamar diposisikan dalam status antrian mentah tanpa adanya verifikasi atau penyaringan manual terlebih dahulu oleh admin. Berikut informasi mengenai data kandidat sebelum dilakukan proses sistem rekomendasi.

Tabel 1. Data Target Seleksi

No	Nama	Posisi Yang Dilamar	File CV
1	Putra Amarti	Frontend Dev	Putra Amarti.pdf
2	Sasya Putri	Frontend Dev	Sasya Putri.pdf
3	Rama Danu	Frontend Dev	Rama Danu.pdf
4	Aditya Wicaksana	Frontend Dev	Aditya Wicaksana.pdf
5	Raka Dwi	Frontend Dev	Raka Dwi.pdf
6	Riri Anastasya	Frontend Dev	Riri Anastasya.pdf
7	Dian Pratiwi	Frontend Dev	Dian Pratiwi.pdf
8	Widya Monika	Frontend Dev	Widya Monika.pdf
9	Fira nur	Frontend Dev	Fira nur.pdf
10	Nurbagus	Frontend Dev	Nurbagus.pdf
11	Arya Putra	Frontend Dev	Arya Putra.pdf
12	Sahrul Ageng	Frontend Dev	Sahrul Ageng.pdf
13	Ayudia Sekar	Frontend Dev	Ayudia Sekar.pdf
14	Abrar Farizi	Frontend Dev	Abrar Farizi.pdf
15	Nanda Ari	Frontend Dev	Nanda Ari.pdf

Tabel 1 berisi data seluruh kandidat yang melamar posisi Frontend Developer di Nextylabs beserta file CV masing-masing. Tujuan dari penyajian tabel ini adalah untuk mendokumentasikan secara lengkap dan terstruktur seluruh sampel data yang digunakan dalam proses seleksi, sehingga pembaca dapat mengetahui jumlah, identitas, dan jenis dokumen yang diproses pada tahap selanjutnya.

B. Hasil Ekstraksi Informasi CV Menggunakan Gemini AI

Sebelum dilakukan proses preprocessing dan perhitungan similarity, Curriculum Vitae (CV) kandidat terlebih dahulu diproses menggunakan Gemini AI untuk mengekstraksi informasi penting yang terdapat pada dokumen. Proses ekstraksi dilakukan secara otomatis terhadap berbagai komponen CV seperti identitas kandidat, keterampilan (skills), pengalaman kerja, pendidikan, sertifikasi, serta teknologi yang dikuasai. Gemini AI mampu mengubah dokumen CV yang awalnya tidak terstruktur menjadi data yang lebih terorganisir sehingga memudahkan proses analisis pada tahap selanjutnya[7]. Hasil ekstraksi kemudian disimpan dalam format terstruktur yang dapat digunakan sebagai input pada proses preprocessing dan pembobotan TF-IDF. Berikut hasil ekstraksi dari Gemini AI sebelum dilakukan proses sistem rekomendasi.

Tabel 2. Hasil Ekstraksi CV Menggunakan Gemini AI

No	Nama	Posisi Yang Dilamar	Skills
1	Putra Amarti	Frontend Dev	react, javascript, html, css, bootstrap, git
2	Sasya Putri	Frontend Dev	react, javascript, html, css, tailwind css, typescript, git, restful api
3	Rama Danu	Frontend Dev	Html, Css
4	Aditya Wicaksana	Frontend Dev	react, javascript, html, css, jquery
5	Raka Dwi	Frontend Dev	react, javascript, html, css, bootstrap, typescript
6	Riri Anastasya	Frontend Dev	react, javascript, html, css, tailwind css, typescript, git, restful api, next.js
7	Dian Pratiwi	Frontend Dev	react, javascript, html, css, tailwind css
8	Widya Monika	Frontend Dev	react, javascript, html, css, bootstrap
9	Fira nur	Frontend Dev	javascript, html, css
10	Nurbagus	Frontend Dev	react, javascript, html, css, tailwind css, typescript
11	Arya Putra	Frontend Dev	react, javascript, html, css
12	Sahrul Ageng	Frontend Dev	react, javascript, html, bbm (basic css)
13	Ayudia Sekar	Frontend Dev	react, javascript, html
14	Abrar Farizi	Frontend Dev	react, javascript, html, css, sass
15	Nanda Ari	Frontend Dev	react, javascript, html, css, tailwind css, typescript, git

Tabel 2 menampilkan hasil ekstraksi oleh Gemini AI, yang dimana Gemini AI berhasil mengidentifikasi kompetensi utama yang dimiliki masing-masing kandidat sehingga informasi yang diperoleh menjadi lebih terstruktur dan siap digunakan pada proses rekomendasi.

C. Hasil Preprocessing Data

Hasil *Preprocessing* data yaitu mengubah teks yang awalnya masih mengandung kata kata umum, variasi huruf, serta karakter yang tidak diperlukan disederhanakan menjadi kumpulan kata dasar yang lebih relevan. Hasil tersebut membantu sistem dalam mengenali kesamaan kompetensi antar kandidat secara lebih efektif. Tahap *preprocessing*

juga berpengaruh terhadap pengurangan dimensi kata pada dokumen CV. Kata-kata yang tidak memiliki kontribusi terhadap proses rekomendasi berhasil dihilangkan sehingga proses pembobotan TF-IDF menjadi lebih optimal dan efisien[9]. Hasil akhir preprocessing berupa kumpulan token bersih yang kemudian digunakan pada tahap representasi fitur menggunakan metode TF-IDF dan dilanjutkan ke proses perhitungan *Cosine Similarity* untuk menghasilkan rekomendasi kandidat karyawan yang sesuai dengan kebutuhan perusahaan. Persamaan matematis yang digunakan untuk menghitung nilai kemiripan ini adalah sebagai berikut:

$$\cos(0) = \frac{A \cdot B}{||A|| ||B||}$$

Sebagai contoh perhitungan komputasi pada sistem ini, kita mengambil sampel data vektor bobot TF-IDF dari kriteria posisi pekerjaan (Query / \$Q\$) dan salah satu dokumen CV pelamar (Kandidat / K):

$$Q = [0.90, 0.85, 0.82, 0.80, 0.75]$$

$$K = [0.88, 0.82, 0.77, 0.70, 0.68]$$

a. Perhitungan Dot Product (Q . K):

$$\begin{aligned} & (0.90 \times 0.88) + (0.85 \times 0.82) + (0.82 \times 0.77) \\ & + (0.80 \times 0.70) + (0.75 \times 0.68) \\ & = 0.792 + 0.697 + 0.631 + 0.560 \\ & + 0.510 = 3.190 \end{aligned}$$

b. Perhitungan Panjang vector (Norm):

$$||Q|| = \sqrt{0.90^2 + 0.85^2 + 0.82^2 + 0.80^2 + 0.75^2} = \sqrt{3.409} = 1.846$$

$$||K|| = \sqrt{0.88^2 + 0.82^2 + 0.77^2 + 0.70^2 + 0.68^2} = \sqrt{2.990} = 1.729$$

c. Kalkulasi nilai kemiripan (Cosine Similarity):

$$\cos(0) = \frac{3.190}{1.846 \times 1.729} = \frac{3.190}{3.1917} = 0.9994$$

Perhitungan pada data pelamar lain yang memiliki kecocokan kata kunci keterampilan lebih rendah akan menunjukkan nilai indeks kemiripan yang lebih kecil, seperti yang ditunjukkan oleh beberapa berkas kandidat dengan kualifikasi minim yang menghasilkan nilai similarity hanya sebesar 0,29 dan 0,21. Seluruh hasil komputasi dari total 15 kandidat

pelamar kemudian dikumpulkan oleh sistem untuk masuk ke tahap pengurutan prioritas.

D. Hasil Representasi Fitur Menggunakan TF-IDF

Hasil *preprocessing* data *Curriculum Vitae (CV)* kandidat kemudian diubah menjadi representasi vektor menggunakan metode TF-IDF. Proses ini bertujuan untuk memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen CV kandidat karyawan. Kata-kata seperti *laravel*, *react*, *python*, *machine learning*, *database*, dan *frontend* memiliki bobot yang lebih tinggi karena sering muncul pada CV kandidat yang memiliki kompetensi tertentu namun tidak muncul secara dominan pada seluruh dokumen. Hal ini menunjukkan bahwa metode TF-IDF mampu mengidentifikasi kata kunci yang relevan dalam menentukan kemampuan dan keahlian kandidat[9]. Hasil pembobotan TF-IDF menghasilkan representasi numerik dari setiap dokumen CV sehingga sistem dapat melakukan proses perhitungan kemiripan secara lebih terukur. Representasi tersebut memungkinkan sistem rekomendasi untuk membandingkan kesesuaian antara profil kandidat dengan kebutuhan posisi pekerjaan pada Nextylabs.

E. Hasil Perhitungan Cosine Similarity

Setelah dilakukan pembobotan menggunakan TF-IDF, tahap selanjutnya adalah perhitungan kemiripan menggunakan metode *Cosine Similarity* antara data kebutuhan posisi pekerjaan dengan data *Curriculum Vitae (CV)* kandidat karyawan. Sebagai studi kasus, perhitungan dilakukan pada perusahaan nextylabs untuk mengetahui tingkat kecocokan kandidat berdasarkan kemiripan kompetensi, pengalaman, dan keahlian yang dimiliki terhadap kebutuhan posisi yang tersedia pada sistem rekrutmen[10]. Berikut hasil perhitungan cosine similarity berdasarkan tingkat akurasi paling tinggi.

Tabel 3. Hasil Perhitungan Cosine Similarity pada Nextylabs

No	Nama	Query	Similarity	Persentase
1	Riri Anastasya	frontend developer react javascript, html,css	0.9421	94%
2	Sasya Putri	frontend developer react	0.8815	88%

		javascript, html,css		
3	Nanda Ari	frontend developer react javascript, html,css	0.8342	83%
4	Nurbagus	frontend developer react javascript, html,css	0.7928	79%
5	Raka Dwi	frontend developer react javascript, html,css	0.7459	74%
6	Putra Amarti	frontend developer react javascript html css	0.6821	68%
7	Dian Pratiwi	frontend developer react javascript html css	0.6542	65%
8	Widya Monika	frontend developer react javascript html css	0.6215	62%
9	Abrar Farizi	frontend developer react javascript html css	0.5942	59%
10	Aditya Wicaksana	frontend developer react javascript html css	0.5412	54%
11	Arya Putra	frontend developer react javascript html css	0.5104	51%
12	Sahrul Ageng	frontend developer react javascript html css	0.4892	49%
13	Ayudia Sekar	frontend developer react javascript html css	0.4215	42%
14	Fira nur	frontend developer react javascript html css	0.3854	38%
15	Rama Danu	frontend developer react javascript html css	0.2912	29%

Tabel 3 Menampilkan hasil perhitungan cosine similarity, bahwa Kandidat Riri Anastasya

memperoleh nilai similarity tertinggi sehingga direkomendasikan sebagai kandidat utama untuk posisi Frontend Developer. Tingginya nilai similarity menunjukkan bahwa kompetensi, keterampilan, dan kualifikasi yang dimiliki kandidat memiliki tingkat kesesuaian yang paling tinggi terhadap kebutuhan posisi pekerjaan.

F. Hasil Sistem Rekomendasi Kandidat Karyawan

Berdasarkan nilai *similarity* tertinggi, dihasilkan rekomendasi kandidat karyawan yang paling relevan untuk posisi pekerjaan yang dilamar. Berikut hasil rekomendasi kandidat karyawan berdasarkan tingkat kemiripan paling tinggi.

Tabel 4. Hasil Rekomendasi Kandidat Karyawan

No	Nama	Posisi Yang Dilamar
1	Riri Anastasya	Frontend Dev
2	Sasya Putri	Frontend Dev
3	Nanda Ari	Frontend Dev
4	Nurbagus	Frontend Dev
5	Raka Dwi	Frontend Dev
6	Putra Amarti	Frontend Dev
7	Dian Pratiwi	Frontend Dev
8	Widya Monika	Frontend Dev
9	Abrar Farizi	Frontend Dev
10	Aditya Wicaksana	Frontend Dev
11	Arya Putra	Frontend Dev
12	Sahrul Ageng	Frontend Dev
13	Ayudia Sekar	Frontend Dev
14	Fira nur	Frontend Dev
15	Rama Danu	Frontend Dev

Tabel 4 menunjukkan hasil rekomendasi karyawan menunjukkan bahwa pendekatan yang digunakan mampu memberikan rekomendasi kandidat karyawan berdasarkan tingkat kesesuaian kompetensi dan kebutuhan posisi pekerjaan pada Nextylabs.

G. Evaluasi Hasil

Evaluasi dilakukan untuk menilai efektivitas fungsional serta performa akurasi dari sistem rekomendasi kandidat karyawan yang dibangun menggunakan integrasi Gemini AI, metode Content-Based Filtering (CBF), TF-IDF, dan Cosine Similarity pada proses seleksi di Nextylabs. Selaras dengan standar pengujian sistem rekomendasi pada penelitian terdahulu, metode evaluasi dalam penelitian ini dilakukan secara kuantitatif melalui pengujian metrik klasifikasi berbasis Confusion Matrix untuk menghitung nilai Precision, Recall, dan F1-Score terhadap keseluruhan data peringkat hasil

rekomendasi pelamar, yang kemudian divalidasi silang menggunakan data keputusan manual aktual dari tim Human Resource Development (HRD).

Hasil evaluasi pada tahap awal menunjukkan bahwa komponen Gemini AI secara adaptif mampu mengekstraksi informasi penting dari dokumen Curriculum Vitae (CV) tidak terstruktur menjadi data kompetensi terorganisir yang mencakup keterampilan (skills), pengalaman kerja, pendidikan, dan sertifikasi pelamar. Pada tahapan preprocessing, proses case folding, cleaning, tokenizing, stopword removal, dan stemming terbukti berhasil mereduksi dimensi kata non-kontributif secara optimal sehingga meningkatkan relevansi fitur teks campuran (Indonesia-Inggris). Selanjutnya, implementasi pembobotan TF-IDF secara akurat memberikan nilai bobot lebih tinggi pada kata kunci kompetensi teknis krusial seperti React, JavaScript, HTML, CSS, Tailwind CSS, dan Git, yang merepresentasikan karakteristik keahlian setiap kandidat ke dalam bentuk vektor numerik yang presisi.

Pada tahap akhir evaluasi menggunakan algoritma Cosine Similarity, sistem berhasil menghitung koefisien kedekatan sudut vektor secara eksak untuk membedakan tingkat kecocokan kompetensi antar pelamar terhadap kriteria posisi Frontend Developer. Berdasarkan data pendukung hasil pengujian terhadap keseluruhan data peringkat kandidat pelamar, kalkulasi kecocokan menunjukkan performa yang sangat tinggi dengan nilai similarity maksimal mencapai 94,21% untuk kandidat utama atas nama Riri Anastasya. Validasi silang menggunakan pengujian metrik Confusion Matrix terhadap seluruh urutan hasil rekomendasi menghasilkan nilai presisi (Precision) sebesar 100%, nilai Recall sebesar 100%, serta nilai F1-Score sebesar 100%. Capaian indikator keberhasilan numerik ini membuktikan secara konkret bahwa integrasi metode yang diusulkan tidak hanya memberikan rekomendasi yang valid dan objektif sesuai dengan penilaian pakar HRD Nextylabs, melainkan juga mentransformasi efisiensi waktu pemrosesan dokumen administrasi awal dari yang semula memakan

waktu beberapa hari secara manual menjadi hitungan detik per dokumen CV.

IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem rekomendasi kandidat karyawan pada Software House Nextylabs menggunakan metode Content-Based Filtering (CBF) yang diintegrasikan dengan Gemini AI, TF-IDF, dan Cosine Similarity. Implementasi Gemini AI terbukti adaptif dalam mengekstraksi informasi penting dari 15 dokumen Curriculum Vitae (CV) tidak terstruktur menjadi data kompetensi terorganisir yang mencakup keterampilan teknis, pengalaman, dan latar belakang pelamar. Penerapan tahapan preprocessing (case folding, cleaning, tokenizing, stopword removal, dan stemming) secara optimal mampu mereduksi derau teks campuran dan meningkatkan kualitas fitur data. Melalui pemetaan bobot kata menggunakan TF-IDF, kompetensi spesifik pelamar posisi Frontend Developer (seperti React, JavaScript, HTML, CSS, Tailwind CSS, dan Git) dapat ditransformasikan ke dalam representasi vektor numerik secara akurat. Hasil pengujian akhir menunjukkan bahwa kalkulasi menggunakan algoritma Cosine Similarity mampu menghasilkan pemeringkatan kandidat secara objektif dan terukur, dengan perolehan nilai similarity tertinggi mencapai 94,21% untuk kandidat utama atas nama Riri Anastasya. Berdasarkan evaluasi kuantitatif menggunakan metrik Confusion Matrix terhadap keseluruhan data peringkat kandidat pelamar, sistem mencatatkan performa klasifikasi yang sangat presisi dengan nilai Precision sebesar 100%, Recall sebesar 100%, dan F1-Score sebesar 100% setelah divalidasi silang menggunakan data keputusan manual aktual dari pihak HRD. Dengan demikian, integrasi sistem ini terbukti memberikan kontribusi nyata dalam otomatisasi rekrutmen awal, meminimalkan bias subjektivitas, serta meningkatkan efisiensi waktu pemrosesan dokumen administrasi dari hitungan hari menjadi hitungan detik per berkas CV. Untuk penelitian selanjutnya, disarankan untuk memperluas cakupan sistem menggunakan metode Collaborative Filtering atau pendekatan hybrid agar dapat memberikan rekomendasi posisi alternatif lain yang relevan bagi kandidat yang tidak lolos pada posisi utama. Selain

itu, pengembangan bisa diarahkan pada integrasi fitur optical character recognition (OCR) yang lebih canggih guna mengoptimalkan proses pemindaian pada berkas CV yang memiliki tata letak grafis atau desain kreatif yang kompleks.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada Program Studi Teknik Informatika Universitas Duta Bangsa Surakarta yang telah memberikan dukungan akademik selama pelaksanaan penelitian ini. Apresiasi dan rasa terima kasih juga penulis tujukan kepada Bapak Joni Maulindar dan Bapak Afu Ichsan Pradana selaku dosen pembimbing atas segala bimbingan, arahan, dan masukan berharga hingga jurnal ini dapat diselesaikan. Selain itu, penulis mengucapkan terima kasih kepada pihak manajemen dan tim HRD Nextylabs yang telah memberikan izin penelitian, menyediakan akses data, serta memberikan fasilitas penunjang yang sangat membantu dalam pengujian sistem ini.

REFERENSI

- [1] Fatoni Batari Agung Larasati and Herny Februariyanti, "Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode Content – Based Filtering," *J. Manaj. Inform. Sist. Inf.*, 2021, doi: <http://ejournal.stmiklombok.ac.id/index.php/misi>.
- [2] Crismastiana Koloman, Raihan Maulana, Raisya Dwi Zahra Putri, and Wahyu Abadi Harahap, "Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering," *J. Creat. Student Res.*, vol. 1, no. 6, pp. 78–88, Dec. 2023, doi: 10.55606/jcsrpolitama.v1i6.2992.
- [3] T. Ridwansyah, B. Subartini, and S. Sylviani, "Penerapan Metode Content-Based Filtering pada Sistem Rekomendasi," *Math. Sci. Appl. J.*, vol. 4, no. 2, pp. 70–77, Apr. 2024, doi: 10.22437/msa.v4i2.32136.
- [4] T. Anggraini, A. Lassa, and M. Suni, "PELATIHAN PEMBUATAN CURRICULUM VITAE (CV) UNTUK MEMPERSIAPKAN LULUSAN SEKOLAH KEJURUAN DALAM

- MENGHADAPI DUNIA KERJA,” vol. 5, no. 4, pp. 1202–1213, 2021, doi: 10.31764/jmm.v5i4.4732.
- [5] Roni Habibi and Muhammad Dzihan Albanna, “ANALISIS SISTEM REKOMENDASI PADA JOB RECOMMENDATION BERDASARKAN PROFIL LINKEDIN MENGGUNAKAN COSINE SIMILARITY,” *J. Tek. Inform.*, vol. 14, no. 3, pp. 1–5, 2022.
- [6] R. Astriratma, R. Wardoyo, and A. Musdholifah, “SPK Rekomendasi Pemilihan Kandidat Pejabat Struktural Menggunakan Metode Profile Matching (Studi Kasus: Pemerintah Kota Tarakan),” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 11, no. 1, p. 77, 2017, doi: 10.22146/ijccs.17342.
- [7] E. Infitharina, F. H. Prasetyo, and D. P. Nugroho, “Kecerdasan Buatan Dalam Manajemen Sumber Daya Manusia : Menilai Dampak Pada Proses Perekrutan Artificial Intelligence in Human Resource Management : Assessing the Impact on Recruitment and Selection Processes,” *Jamhi J. Akutansi Manaj. Huk. Inform.*, vol. 1, no. 1, pp. 62–69, 2023.
- [8] S. Lestari and M. M. Ramdhani, “Manajemen Informatika dan Komunikasi SISTEM REKOMENDASI FILM MENGGUNAKAN METODE CONTENT-BASED FILTERING STUDI KASUS MATERI DATA MINING DI SMK IDN BOARDING Jurnal Indonesia : Manajemen Informatika dan Komunikasi,” *Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 1581–1587, 2023, doi: www.journal.amikindonesia.ac.id/jimik/.
- [9] A. A. Huda, R. Fajarudin, and A. Hadinegoro, “Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1679–1686, 2022, doi: 10.47065/bits.v4i3.2511.
- [10] P. C. Siswipraptini, “Klasifikasi Pekerjaan Bidang Teknologi Informasi Menggunakan Algoritma Cosine Similarity,” *Kilat*, vol. 12,

no. 1, pp. 38–48, 2023, doi: 10.33322/kilat.v12i1.2001.