

PENERAPAN CONTENT-BASED FILTERING UNTUK REKOMENDASI PENEMPATAN DIVISI MAGANG BERDASARKAN PROGRAM STUDI, MINAT, DAN KEAHLIAN

Yosua Aremathea Novantoro, Moh. Muhtarom, Joni Maulindar
Program Studi Teknik Informatika, Universitas Duta Bangsa Surakarta
Jl. Bhayangkara No.55 Kota Surakarta, Jawa Tengah, Indonesia
yosuanovantoro@gmail.com

ABSTRAK

Program magang merupakan kegiatan pembelajaran yang memberikan kesempatan kepada mahasiswa untuk memperoleh pengalaman kerja secara langsung di dunia industri. Namun, proses penempatan peserta magang masih sering dilakukan secara manual dengan informasi yang terbatas, sehingga berpotensi menimbulkan ketidaksesuaian antara program studi, minat, dan keahlian peserta dengan kebutuhan divisi. Penelitian ini bertujuan untuk menerapkan metode *Content-Based Filtering* dalam memberikan rekomendasi penempatan divisi magang berdasarkan profil peserta. Data yang digunakan meliputi program studi, minat, keahlian peserta, dan kebutuhan divisi. Tahapan penelitian terdiri dari pengumpulan data, *preprocessing*, ekstraksi fitur menggunakan TF-IDF, serta perhitungan tingkat kemiripan menggunakan *cosine similarity*. Hasil pengujian terhadap 12 data peserta menunjukkan bahwa sistem mampu menghasilkan rekomendasi berdasarkan nilai kemiripan tertinggi. Peserta dengan kemampuan pengembangan aplikasi cenderung direkomendasikan ke divisi *IT Support*, peserta dengan kemampuan jaringan ke divisi Teknisi, dan peserta dengan kemampuan administrasi serta pengolahan data ke divisi *Shared Service*. Nilai kemiripan tertinggi diperoleh oleh Taufik Hidayat sebesar 0,5399 pada divisi Teknisi. Dengan demikian, metode *Content-Based Filtering* dapat mendukung proses penempatan divisi magang secara lebih terarah dan objektif.

Kata kunci : *Content-Based Filtering, sistem rekomendasi, penempatan divisi magang, TF-IDF, cosine similarity*

1. PENDAHULUAN

Program magang merupakan salah satu bentuk pembelajaran yang memberikan kesempatan kepada mahasiswa untuk mengenal dunia kerja secara langsung. Melalui kegiatan magang, mahasiswa dapat menerapkan ilmu yang diperoleh selama perkuliahan sekaligus mengembangkan kemampuan sesuai dengan kebutuhan dunia industri. Dalam pelaksanaannya, penempatan mahasiswa pada divisi yang sesuai menjadi hal penting karena dapat memengaruhi efektivitas kegiatan magang, baik bagi mahasiswa maupun perusahaan. Sistem rekomendasi secara umum dapat digunakan untuk membantu pengguna memperoleh pilihan yang paling relevan berdasarkan karakteristik atau preferensi tertentu [1].

Selain itu, kualitas data juga menjadi faktor penting dalam sistem rekomendasi karena ketepatan hasil rekomendasi sangat dipengaruhi oleh data yang digunakan [11].

Permasalahan yang sering terjadi dalam proses penempatan peserta magang adalah penentuan divisi yang masih dilakukan secara manual dengan mempertimbangkan informasi yang terbatas. Proses tersebut dapat menyebabkan ketidaksesuaian antara program studi, minat, dan keahlian mahasiswa dengan kebutuhan divisi yang tersedia. Akibatnya, peserta magang berpotensi ditempatkan pada divisi yang kurang sesuai dengan kemampuan yang dimiliki, sehingga proses pembelajaran dan kontribusi selama magang menjadi kurang optimal. Sistem rekomendasi telah banyak diterapkan untuk membantu proses pencocokan profil pengguna dengan kebutuhan tertentu. Penerapan sistem rekomendasi dapat

digunakan pada rekomendasi pekerjaan di bidang teknologi informasi [2].

Sistem rekomendasi juga dapat dimanfaatkan dalam pemilihan program bagi mahasiswa [3].

Selain itu, metode rekomendasi berbasis konten dapat digunakan untuk membantu pencocokan profil pencari kerja dengan lowongan pekerjaan [4].

Pencocokan antara keterampilan pengguna dan kebutuhan pekerjaan juga dapat membantu menghasilkan rekomendasi yang lebih sesuai [9].

Pada bidang pendidikan vokasi, model rekomendasi dapat digunakan untuk mencocokkan kompetensi pengguna dengan kebutuhan kerja [10].

Salah satu metode yang dapat digunakan untuk mengatasi permasalahan tersebut adalah *Content-Based Filtering*. Metode ini bekerja dengan membandingkan atribut pengguna dengan atribut *item* sehingga rekomendasi diberikan berdasarkan tingkat kemiripan konten [8].

Dalam penelitian ini, atribut pengguna berupa program studi, minat, dan keahlian mahasiswa, sedangkan atribut *item* berupa data divisi yang mencakup deskripsi dan kriteria kebutuhan divisi. *Content-Based Filtering* juga telah diterapkan pada sistem rekomendasi artikel jurnal berbasis *machine learning* [5].

Metode yang sama dapat digunakan untuk membantu pencocokan topik mahasiswa dengan bidang dosen pembimbing secara lebih tepat [6]. Untuk menghasilkan rekomendasi berbasis kemiripan teks, penelitian ini menggunakan TF-IDF dan *cosine similarity*. TF-IDF digunakan untuk memberikan bobot pada kata berdasarkan tingkat kepentingannya,

sedangkan *cosine similarity* digunakan untuk menghitung tingkat kemiripan antara profil peserta dan data divisi. TF-IDF dan *Content-Based Filtering* dapat digunakan pada sistem rekomendasi pekerjaan berbasis kesamaan konten [7].

Penerapan *Content-Based Filtering* juga dapat digunakan dalam sistem rekomendasi objek wisata dengan mempertimbangkan kesamaan karakteristik data [12].

Dengan pendekatan tersebut, sistem dapat menentukan divisi yang paling sesuai berdasarkan nilai kemiripan tertinggi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menerapkan metode *Content-Based Filtering* pada sistem rekomendasi penempatan divisi magang berdasarkan program studi, minat, dan keahlian mahasiswa. Rencana penyelesaian masalah dilakukan melalui beberapa tahapan, yaitu mengumpulkan data peserta dan data divisi, melakukan *preprocessing* data, mengubah data teks menjadi bentuk numerik menggunakan TF-IDF, menghitung tingkat kemiripan menggunakan *cosine similarity*, serta menentukan divisi dengan nilai kemiripan tertinggi sebagai hasil rekomendasi. Dengan adanya sistem ini, proses penempatan divisi magang diharapkan dapat dilakukan secara lebih sesuai, terarah, dan objektif.

2. TINJAUAN PUSTAKA

Tinjauan pustaka memuat konsep-konsep yang mendukung penelitian, meliputi sistem rekomendasi, metode *Content-Based Filtering*, TF-IDF, *cosine similarity*, dan penelitian terdahulu yang relevan. Pembahasan pada bagian ini digunakan sebagai landasan teoritis dalam pengembangan sistem rekomendasi penempatan divisi magang.

2.1. Sistem Rekomendasi

Sistem rekomendasi merupakan sistem yang dirancang untuk membantu pengguna memperoleh pilihan yang paling sesuai berdasarkan kebutuhan, preferensi, atau karakteristik tertentu. Dalam berbagai bidang, sistem rekomendasi digunakan untuk mengurangi banyaknya informasi yang harus dipilih secara manual sehingga proses pengambilan keputusan dapat dilakukan dengan lebih cepat dan terarah [1].

Perkembangan sistem rekomendasi saat ini juga menunjukkan bahwa data menjadi komponen penting karena kualitas data sangat berpengaruh terhadap ketepatan hasil rekomendasi yang dihasilkan [11].

Dalam penelitian ini, sistem rekomendasi digunakan untuk membantu proses penempatan divisi magang agar lebih sesuai dengan profil mahasiswa.

2.2. Content-Based Filtering

Content-Based Filtering merupakan salah satu pendekatan dalam sistem rekomendasi yang bekerja dengan mencocokkan atribut pengguna dengan atribut item yang tersedia. Rekomendasi diberikan

berdasarkan tingkat kemiripan antara karakteristik pengguna dan karakteristik item [8].

Pada penelitian sebelumnya, metode CBF telah digunakan pada rekomendasi pekerjaan di bidang teknologi informasi dan menghasilkan rekomendasi yang relevan berdasarkan kecocokan kemampuan pengguna dengan deskripsi pekerjaan [2].

Selain itu, metode yang sama juga diterapkan pada rekomendasi lowongan pekerjaan dan dinilai efektif dalam memberikan hasil yang sesuai dengan profil pencari kerja [4].

Penelitian lain menunjukkan bahwa pencocokan antara keterampilan pengguna dan deskripsi pekerjaan dapat meningkatkan relevansi hasil rekomendasi yang diberikan [9].

Berdasarkan hal tersebut, CBF dinilai relevan untuk diterapkan pada rekomendasi penempatan divisi magang.

2.3. TF-IDF dan Cosine Similarity

Dalam sistem rekomendasi berbasis konten, data teks perlu direpresentasikan dalam bentuk numerik agar dapat dihitung tingkat kemiripannya. Salah satu metode yang umum digunakan adalah TF-IDF, yaitu teknik pembobotan kata berdasarkan frekuensi kemunculannya dalam dokumen dan tingkat kepentingannya terhadap keseluruhan dokumen [7].

Setelah data direpresentasikan menggunakan TF-IDF, tingkat kemiripan antara profil mahasiswa dan kebutuhan divisi dapat dihitung menggunakan *cosine similarity*. Metode ini digunakan untuk mengukur kedekatan antara dua vektor, di mana nilai yang semakin tinggi menunjukkan bahwa dua data yang dibandingkan memiliki tingkat kemiripan yang semakin besar [5].

Kombinasi TF-IDF dan *cosine similarity* telah digunakan dalam beberapa penelitian dan dinilai cukup efektif dalam menghasilkan rekomendasi berbasis kesamaan konten [6], [7], [12].

2.4. Penelitian Terdahulu

Beberapa penelitian terdahulu menunjukkan bahwa *Content-Based Filtering* telah diterapkan pada berbagai kasus rekomendasi. Penelitian Winardi dkk. menggunakan TF-IDF dan *cosine similarity* pada sistem rekomendasi pekerjaan dan memperoleh hasil yang cukup baik [7].

Penelitian Elnursa dkk. menerapkan sistem rekomendasi pada pemilihan program MSIB bagi mahasiswa pendidikan informatika dan menunjukkan bahwa pendekatan rekomendasi dapat membantu proses seleksi program yang lebih sesuai [3].

Sementara itu, penelitian Fitria dkk. menunjukkan bahwa *Content-Based Filtering* dapat digunakan untuk mencocokkan profil pengguna dengan lowongan pekerjaan [4].

Penelitian Astuti dan Andriyani juga menunjukkan bahwa pendekatan CBF dapat membantu proses pencocokan topik mahasiswa

dengan bidang dosen pembimbing secara lebih tepat [6].

Selain itu, penelitian Farell dkk. memperlihatkan bahwa model rekomendasi juga relevan digunakan untuk membantu pencocokan kompetensi pengguna dengan kebutuhan kerja pada bidang pendidikan vokasi [10].

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa metode CBF relevan untuk diterapkan pada rekomendasi penempatan divisi magang berbasis atribut program studi, minat, dan keahlian.

3. METODE PENELITIAN

Pada bab ini dijelaskan tahapan-tahapan yang dilakukan dalam penelitian, mulai dari proses pengumpulan data, metode pengembangan sistem, hingga penerapan metode *Content-Based Filtering* dalam menghasilkan rekomendasi penempatan divisi magang. Penjelasan pada bagian ini bertujuan untuk memberikan gambaran mengenai alur penelitian yang digunakan agar proses penelitian dapat dilakukan secara terarah dan sistematis.

3.1. Jenis dan Sumber Data

Pada penelitian ini, data yang digunakan terdiri dari data primer dan data sekunder. Data primer merupakan data yang diperoleh secara langsung dari sumber utama yang berkaitan dengan objek penelitian, sedangkan data sekunder merupakan data pendukung yang diperoleh dari dokumen, jurnal, artikel ilmiah, dan referensi lain yang relevan dengan penelitian.

a. Data Primer

Data primer dalam penelitian ini diperoleh secara langsung dari lingkungan penelitian, yaitu data yang berkaitan dengan profil mahasiswa magang dan kebutuhan divisi. Data ini diperoleh melalui pengamatan terhadap proses penempatan magang serta informasi dari pihak yang terlibat dalam kegiatan magang.

b. Data Sekunder

Data sekunder merupakan data pendukung yang diperoleh dari berbagai sumber tertulis, seperti jurnal, artikel ilmiah, buku, dokumen, dan referensi lain yang berkaitan dengan sistem rekomendasi, *Content-Based Filtering*, TF-IDF, dan *cosine similarity*. Data sekunder ini digunakan untuk memperkuat landasan teori dan mendukung proses analisis dalam penelitian.

3.2. Metode Pengumpulan Data

Untuk memperoleh data yang dibutuhkan, penelitian ini menggunakan beberapa teknik pengumpulan data, yaitu observasi, wawancara, dan studi pustaka.

a. Observasi

Observasi dilakukan untuk memahami kondisi secara langsung terkait proses penempatan mahasiswa magang dan kebutuhan divisi yang ada. Melalui observasi, peneliti memperoleh gambaran

mengenai alur penempatan magang yang berlangsung di lapangan.

b. Wawancara

Wawancara dilakukan dengan pihak yang berkaitan dengan proses penempatan magang, seperti pembimbing atau pihak yang memahami kebutuhan divisi. Wawancara ini bertujuan untuk memperoleh informasi yang lebih rinci mengenai kriteria penempatan mahasiswa magang.

c. Studi Pustaka

Studi pustaka dilakukan dengan mengumpulkan dan mempelajari berbagai referensi yang relevan, seperti jurnal, buku, dan artikel ilmiah. Teknik ini digunakan untuk memperkuat dasar teori serta mendukung pemilihan metode yang digunakan dalam penelitian.

3.3. Metode Pengembangan Sistem

Tahapan penelitian ini disusun sebagai langkah yang sistematis untuk menghasilkan sistem rekomendasi penempatan divisi magang yang sesuai dengan profil mahasiswa. Proses penelitian dimulai dari pengumpulan data, kemudian dilanjutkan dengan pengolahan data, penerapan metode, hingga evaluasi hasil rekomendasi yang dihasilkan sistem. Dengan tahapan yang terstruktur, penelitian dapat dilakukan secara lebih terarah dan memudahkan proses analisis pada setiap tahap.

a. Pengumpulan Data

Tahap awal penelitian dilakukan dengan mengumpulkan data yang dibutuhkan, yaitu data mahasiswa dan data divisi. Data mahasiswa meliputi program studi, minat, dan keahlian, sedangkan data divisi meliputi nama divisi, deskripsi tugas, dan kompetensi yang dibutuhkan. Data ini menjadi dasar utama dalam proses pencocokan antara karakteristik mahasiswa dengan kebutuhan setiap divisi.

b. Preprocessing Data

Setelah data terkumpul, langkah berikutnya adalah melakukan *preprocessing* data. Tahap ini bertujuan untuk menyiapkan data agar lebih terstruktur dan siap digunakan dalam proses perhitungan. Pada tahap ini dilakukan pembersihan data, penyusunan format, serta penyesuaian atribut agar data mahasiswa maupun data divisi dapat diproses dengan lebih baik oleh sistem.

c. Ekstraksi Fitur Menggunakan TF-IDF

Data teks yang telah diproses kemudian diubah ke dalam bentuk numerik menggunakan metode TF-IDF. Proses ini dilakukan untuk memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya di dalam dokumen. Hasil dari ekstraksi fitur ini menjadi representasi data yang akan digunakan dalam proses penghitungan kemiripan antara profil mahasiswa dan kebutuhan divisi.

d. Menghitung Kemiripan Menggunakan *Cosine Similarity*

Setelah data direpresentasikan dalam bentuk numerik, tahap berikutnya adalah menghitung tingkat kemiripan antara profil mahasiswa dan kebutuhan divisi menggunakan *cosine similarity*. Perhitungan ini bertujuan untuk mengetahui tingkat kesesuaian antara dua data yang dibandingkan. Nilai kemiripan yang lebih tinggi menunjukkan bahwa divisi tersebut lebih sesuai untuk direkomendasikan kepada mahasiswa.

e. Evaluasi Hasil

Tahap terakhir adalah melakukan evaluasi terhadap hasil rekomendasi yang dihasilkan sistem. Evaluasi dilakukan untuk melihat apakah rekomendasi yang diberikan sudah sesuai dengan karakteristik mahasiswa dan kebutuhan divisi. Dari tahap ini dapat diketahui sejauh mana metode yang digunakan mampu mendukung proses penempatan divisi magang secara lebih relevan, efisien, dan terarah.

3.4. Tahapan Penerapan Content-Based Filtering

Tahapan penerapan *Content-Based Filtering* pada penelitian ini dilakukan untuk menghasilkan rekomendasi penempatan divisi magang berdasarkan kesesuaian antara profil peserta dan data divisi. Profil peserta dibentuk dari atribut program studi, minat, dan keahlian, sedangkan profil divisi dibentuk dari deskripsi divisi, kriteria program studi, kriteria minat, dan kriteria keahlian. Seluruh data tersebut diproses melalui beberapa tahapan, yaitu pembentukan profil teks, *preprocessing*, pembobotan menggunakan TF-IDF, serta perhitungan tingkat kemiripan menggunakan *cosine similarity*. Divisi dengan nilai kemiripan tertinggi akan dipilih sebagai rekomendasi utama bagi peserta magang.

a. Pembentukan Profil Peserta dan Profil Divisi

Pada tahap awal, sistem membentuk profil peserta dari gabungan atribut program studi, minat, dan keahlian yang dimiliki masing-masing peserta. Selanjutnya, sistem juga membentuk profil divisi berdasarkan nama divisi, deskripsi, serta kriteria yang dibutuhkan. Penggabungan atribut tersebut bertujuan agar setiap peserta dan divisi memiliki representasi data teks yang dapat dibandingkan pada proses berikutnya.

b. *Preprocessing* Data Teks

Data teks yang telah dibentuk kemudian melalui tahap *preprocessing* untuk menyederhanakan isi teks agar lebih mudah diproses oleh sistem. Tahap ini meliputi perubahan seluruh huruf menjadi huruf kecil, penghapusan tanda baca, serta penyusunan kata-kata penting dari profil peserta maupun profil divisi. Hasil dari tahap ini adalah data teks yang lebih bersih dan terstruktur sehingga dapat digunakan dalam proses pembobotan.

c. Pembobotan Kata Menggunakan TF-IDF

Setelah melalui tahap *preprocessing*, data teks diubah menjadi bentuk numerik menggunakan metode TF-IDF. Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan

frekuensi kemunculannya dalam dokumen dan tingkat kepentingannya terhadap seluruh dokumen. Rumus TF-IDF yang digunakan adalah sebagai berikut:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \\ = \left(\frac{f_{t,d}}{\sum f_d} \right) \times \log \left(\frac{N}{df(t)} \right)$$

Keterangan:

$f_{t,d}$ adalah jumlah kemunculan kata t pada dokumen d , $\sum f_d$ adalah jumlah seluruh kata pada dokumen d , N adalah jumlah seluruh dokumen, dan $df(t)$ adalah jumlah dokumen yang mengandung kata t . Hasil pembobotan ini digunakan untuk merepresentasikan profil peserta dan profil divisi dalam bentuk vektor numerik.

d. Perhitungan Kemiripan Menggunakan Cosine Similarity

Setelah diperoleh bobot TF-IDF, tahap berikutnya adalah menghitung tingkat kemiripan antara profil peserta dan profil divisi menggunakan metode *cosine similarity*. Metode ini digunakan untuk mengetahui tingkat kesesuaian antara dua vektor berdasarkan kedekatan nilainya. Rumus *cosine similarity* yang digunakan adalah sebagai berikut:

$$\text{Cosine Similarity}(A, B) = \frac{\sum (A_i \times B_i)}{(\sqrt{\sum A_i^2} \times \sqrt{\sum B_i^2})}$$

Keterangan:

A merupakan vektor profil peserta, B merupakan vektor profil divisi, A_i adalah bobot kata ke- i pada vektor peserta, dan B_i adalah bobot kata ke- i pada vektor divisi. Nilai *cosine similarity* berada pada rentang 0 sampai 1. Semakin tinggi nilai yang diperoleh, maka semakin besar tingkat kemiripan antara profil peserta dengan divisi yang dibandingkan.

e. Penentuan Rekomendasi Divisi

Pada tahap akhir, sistem membandingkan nilai *cosine similarity* antara satu peserta dengan seluruh data divisi. Divisi yang memiliki nilai kemiripan paling tinggi akan dipilih sebagai rekomendasi utama. Dengan demikian, hasil rekomendasi yang diberikan sistem didasarkan pada tingkat kesamaan konten antara karakteristik peserta dan kebutuhan masing-masing divisi.

4. HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan hasil dari penerapan metode *Content-Based Filtering* pada sistem rekomendasi penempatan divisi magang. Pembahasan difokuskan pada implementasi sistem, proses perhitungan rekomendasi, hasil yang diperoleh, serta analisis terhadap keluaran sistem dalam menentukan divisi yang sesuai dengan profil mahasiswa.

4.1. Hasil Pengumpulan Data

Pengumpulan data dilakukan dengan menyiapkan data peserta magang dan data divisi yang digunakan dalam proses rekomendasi. Data peserta terdiri dari nama, program studi, minat, dan keahlian. Data tersebut digunakan sebagai profil peserta yang

akan dibandingkan dengan kebutuhan masing-masing divisi.

Pada penelitian ini digunakan 12 data peserta magang sebagai data pengujian. Data tersebut

memiliki variasi program studi, minat, dan keahlian agar sistem dapat diuji pada beberapa karakteristik peserta yang berbeda, seperti bidang web, UI/UX, jaringan, administrasi, konstruksi, dan data analyst.

Tabel 1. Data Mahasiswa

No	Nama Peserta	Program Studi	Minat	Keahlian
1	Vina Oktaviani	Akuntansi	audit	excel
2	Taufik Hidayat	Teknik Elektro	jaringan	troubleshooting
3	Putri Ayu	Manajemen	HR	administrasi
4	Luthfi Ramadhan	Teknik Sipil	konstruksi	AutoCAD
5	Kiki Amelia	Sistem Informasi	UI/UX	Figma, prototyping
6	Joko Susilo	Teknik Informatika	Backend, API	Laravel, PHP
7	Indah Permata	Sistem Informasi	Data analyst	Python, Excel
8	Hendra Wijaya	Teknik Telekomunikasi	jaringan, konfigurasi	Mikrotik, routing
9	Gina Maharani	Teknik Informatika	frontend development	HTML, CSS, React
10	Fajar Nugroho	Sistem Informasi	UI/UX design	Figma, Adobe XD
11	Eko Prasetyo	Teknik Komputer	hardware, jaringan	Cisco, LAN, troubleshooting
12	Dewi Anggraini	Teknik Informatika	mobile development	Flutter, Dart

Berdasarkan Tabel 1, data peserta memiliki karakteristik yang cukup beragam. Peserta dengan latar belakang Teknik Informatika dan Sistem Informasi lebih banyak memiliki minat pada bidang pengembangan aplikasi, UI/UX, *backend*, *frontend*, dan *data analyst*. Sementara itu, peserta dari Teknik Elektro, Teknik Telekomunikasi, dan Teknik Komputer cenderung memiliki minat pada bidang jaringan dan troubleshooting.

4.2. Hasil Preprocessing Data

Tahap *preprocessing* dilakukan untuk menyiapkan data peserta dan data divisi agar dapat

diproses oleh sistem rekomendasi. Data yang digunakan masih berbentuk teks, sehingga perlu disederhanakan terlebih dahulu agar proses perhitungan dapat berjalan lebih baik.

Pada tahap ini, sistem mengubah seluruh teks menjadi huruf kecil, menghapus tanda baca, memisahkan kalimat menjadi kumpulan kata, serta menghilangkan kata yang tidak diperlukan. Data peserta yang terdiri dari program studi, minat, dan keahlian digabungkan menjadi satu profil teks. Data divisi juga digabungkan dari deskripsi, kriteria program studi, kriteria minat, dan kriteria keahlian.

Tabel 2. Hasil *Preprocessing* Data

No	Nama Peserta	Data Awal	Hasil Preprocessing
1	Vina Oktaviani	Akuntansi, audit, Excel	akuntansi audit excel
2	Taufik Hidayat	Teknik Elektro, jaringan, troubleshooting	teknik elektro jaringan troubleshooting
3	Putri Ayu	Manajemen, HR, administrasi	manajemen hr administrasi
4	Luthfi Ramadhan	Teknik Sipil, konstruksi, AutoCAD	teknik sipil konstruksi autocad
5	Kiki Amelia	Sistem Informasi, UI/UX, Figma, prototyping	sistem informasi ui ux figma prototyping
6	Joko Susilo	Teknik Informatika, backend, API, Laravel, PHP	teknik informatika backend api laravel php
7	Indah Permata	Sistem Informasi, data analyst, Python, Excel	sistem informasi data analyst python excel
8	Hendra Wijaya	Teknik Telekomunikasi, jaringan, konfigurasi, Mikrotik, routing	teknik telekomunikasi jaringan konfigurasi mikrotik routing
9	Gina Maharani	Teknik Informatika, frontend development, HTML, CSS, React	teknik informatika frontend development html css react
10	Fajar Nugroho	Sistem Informasi, UI/UX design, Figma, Adobe XD	sistem informasi ui ux design figma adobe xd
11	Eko Prasetyo	Teknik Komputer, hardware, jaringan, Cisco, LAN, troubleshooting	teknik komputer hardware jaringan cisco lan troubleshooting
12	Dewi Anggraini	Teknik Informatika, mobile development, Flutter, Dart	teknik informatika mobile development flutter dart

4.3. Hasil Implementasi Sistem

Sistem rekomendasi penempatan divisi magang dikembangkan berbasis web menggunakan Next.js dan *Firebase Firestore*. Sistem ini digunakan untuk mengolah data peserta berupa program studi, minat, dan keahlian, kemudian menghasilkan rekomendasi

divisi berdasarkan tingkat kemiripan dengan data divisi.

Sistem terdiri dari halaman pengguna dan halaman admin. Halaman pengguna digunakan untuk mengisi data peserta dan melihat hasil rekomendasi. Sementara itu, halaman admin digunakan untuk

mengelola data divisi dan melihat riwayat hasil rekomendasi yang tersimpan pada *Firebase Firestore*.

Tabel 3. Hasil Implementasi Sistem

No	Fitur	Keterangan
1	Form Rekomendasi	Digunakan untuk mengisi data peserta berupa nama, program studi, minat, dan keahlian.
2	Proses Rekomendasi	Sistem menghitung kecocokan peserta dengan divisi menggunakan TF-IDF dan Cosine Similarity.
3	Hasil Rekomendasi	Sistem menampilkan divisi terbaik berdasarkan skor similarity tertinggi.
4	Login Admin	Admin masuk ke sistem menggunakan Firebase Authentication.
5	Kelola Divisi	Admin dapat menambah, mengedit, dan menghapus data divisi.
6	Riwayat Rekomendasi	Sistem menyimpan dan menampilkan kembali hasil rekomendasi peserta.

Berdasarkan Tabel 3, sistem telah memiliki fitur utama yang dibutuhkan dalam proses rekomendasi penempatan divisi magang. Data divisi digunakan sebagai pembanding terhadap profil peserta, sedangkan hasil rekomendasi disimpan agar dapat dilihat kembali oleh admin.

4.4. Hasil Ekstraksi Fitur Menggunakan TF-IDF

Setelah data peserta dan data divisi melalui tahap *preprocessing*, sistem melakukan pembobotan kata menggunakan metode TF-IDF sebagaimana telah dijelaskan pada Bab 3. Tahap ini bertujuan untuk merepresentasikan data teks ke dalam bentuk fitur yang dapat digunakan dalam proses perhitungan kemiripan. Pembobotan dilakukan terhadap kata-kata

yang terdapat pada profil peserta dan profil divisi, sehingga setiap data memiliki representasi yang dapat dibandingkan secara sistematis.

Pada sistem ini, profil peserta dibentuk dari gabungan atribut program studi, minat, dan keahlian. Sementara itu, profil divisi dibentuk dari deskripsi divisi, kriteria program studi, kriteria minat, dan kriteria keahlian. Hasil pembobotan TF-IDF menghasilkan kumpulan fitur yang merepresentasikan karakteristik utama dari masing-masing peserta. Karena sistem tidak menampilkan nilai bobot TF-IDF secara langsung pada antarmuka, hasil pada tahap ini ditunjukkan dalam bentuk representasi fitur utama yang digunakan dalam proses perhitungan *similarity*.

Tabel 4. Hasil Ekstraksi Fitur

No	Nama Peserta	Representasi Fitur
1	Vina Oktaviani	akuntansi, audit, excel
2	Taufik Hidayat	teknik elektro, jaringan, troubleshooting
3	Putri Ayu	manajemen, hr, administrasi
4	Luthfi Ramadhan	teknik sipil, konstruksi, autocad
5	Kiki Amelia	sistem informasi, ui ux, figma, prototyping
6	Joko Susilo	teknik informatika, backend, api, laravel, php
7	Indah Permata	sistem informasi, data analyst, python, excel
8	Hendra Wijaya	teknik telekomunikasi, jaringan, konfigurasi, mikrotik, routing
9	Gina Maharani	teknik informatika, frontend development, html, css, react
10	Fajar Nugroho	sistem informasi, ui ux design, figma, adobe xd
11	Eko Prasetyo	teknik komputer, hardware, jaringan, cisco, lan, troubleshooting
12	Dewi Anggraini	teknik informatika, mobile development, flutter, dart

Berdasarkan Tabel 4, setiap peserta memiliki representasi fitur yang berbeda sesuai dengan program studi, minat, dan keahlian masing-masing. Fitur-fitur tersebut menjadi dasar dalam proses pembobotan dan selanjutnya digunakan untuk menghitung tingkat kemiripan antara profil peserta dan profil divisi. Semakin relevan kata kunci yang dimiliki peserta terhadap kebutuhan divisi, maka semakin besar kemungkinan sistem menghasilkan rekomendasi yang sesuai.

4.5. Hasil Perhitungan Cosine Similarity

Setelah diperoleh representasi fitur dari hasil pembobotan TF-IDF, tahap berikutnya adalah menghitung tingkat kemiripan antara profil peserta dan

profil divisi menggunakan metode *cosine similarity*. Proses ini dilakukan untuk mengetahui tingkat kesesuaian antara karakteristik peserta dengan kebutuhan masing-masing divisi. Nilai *similarity* yang dihasilkan berada pada rentang 0 sampai 1, di mana nilai yang lebih tinggi menunjukkan tingkat kemiripan yang lebih besar.

Pada penelitian ini, sistem membandingkan setiap profil peserta dengan seluruh data divisi yang tersedia. Selanjutnya, divisi dengan nilai *similarity* tertinggi dipilih sebagai rekomendasi utama. Hasil perhitungan menunjukkan adanya perbedaan tingkat kecocokan antar peserta berdasarkan kata kunci yang dimiliki pada masing-masing profil.

Tabel 5. Hasil Perhitungan *Cosine Similarity*

No	Nama Peserta	Profil Utama Peserta	Divisi dengan Similarity Tertinggi	Skor Similarity
1	Vina Oktaviani	audit, excel	Shared Service	0.0517
2	Taufik Hidayat	jaringan, troubleshooting	Teknisi	0.5399
3	Putri Ayu	hr, administrasi	Shared Service	0.2474
4	Luthfi Ramadhan	konstruksi, autocad	Konstruksi	0.1331
5	Kiki Amelia	ui ux, figma, prototyping	IT Support	0.2303
6	Joko Susilo	backend, api, laravel, php	IT Support	0.1009
7	Indah Permata	data analyst, python, excel	Shared Service	0.2981
8	Hendra Wijaya	jaringan, konfigurasi, mikrotik, routing	Teknisi	0.1817
9	Gina Maharani	frontend, html, css, react	IT Support	0.3157
10	Fajar Nugroho	ui ux design, figma, adobe xd	IT Support	0.1872
11	Eko Prasetyo	hardware, jaringan, troubleshooting	Teknisi	0.2749
12	Dewi Anggraini	mobile development, flutter, dart	IT Support	0.1009

Berdasarkan Tabel 5, hasil perhitungan menunjukkan bahwa peserta dengan minat dan keahlian yang berkaitan dengan jaringan dan *troubleshooting* cenderung memperoleh nilai *similarity* tertinggi pada divisi Teknisi. Peserta yang memiliki kemampuan pada bidang pengembangan aplikasi, seperti *frontend*, *backend*, *UI/UX*, dan *mobile development*, cenderung memperoleh rekomendasi pada divisi *IT Support*. Sementara itu, peserta dengan kemampuan administrasi, audit, dan pengolahan data lebih banyak direkomendasikan ke divisi *Shared Service*.

Nilai *similarity* tertinggi diperoleh oleh Taufik Hidayat sebesar 0.5399 pada divisi Teknisi. Hasil ini menunjukkan bahwa kata kunci pada profil peserta memiliki tingkat kesesuaian yang cukup tinggi dengan kebutuhan divisi tersebut. Sebaliknya, nilai *similarity* terendah diperoleh Vina Oktaviani sebesar 0.0517 pada divisi *Shared Service*. Meskipun nilainya relatif rendah, divisi tersebut tetap menjadi rekomendasi terbaik karena memiliki nilai kemiripan paling tinggi dibandingkan divisi lainnya. Hasil ini menunjukkan bahwa sistem mampu memberikan rekomendasi berdasarkan tingkat kesamaan konten antara profil peserta dan karakteristik divisi yang tersedia.

4.6. Alur Proses Rekomendasi

Alur proses rekomendasi pada sistem ini dimulai dari pengisian data peserta magang melalui *form* rekomendasi. Data yang dimasukkan meliputi nama, program studi, minat, dan keahlian. Selanjutnya, sistem menggabungkan atribut program studi, minat, dan keahlian menjadi satu profil teks peserta.

Setelah profil peserta terbentuk, sistem mengambil data divisi yang tersimpan pada *Firestore*. Data divisi tersebut terdiri dari nama divisi, deskripsi, kriteria program studi, kriteria minat, dan kriteria keahlian. Data peserta dan data divisi kemudian melalui tahap *preprocessing* agar format teks menjadi lebih sederhana dan siap diproses.

Tahap berikutnya adalah ekstraksi fitur menggunakan TF-IDF untuk memberikan bobot pada setiap kata. Setelah itu, sistem menghitung tingkat kemiripan antara profil peserta dan setiap divisi menggunakan *Cosine Similarity*. Divisi dengan nilai

similarity tertinggi dipilih sebagai rekomendasi utama dan ditampilkan kepada pengguna. Hasil rekomendasi tersebut juga disimpan ke *Firestore* sebagai riwayat rekomendasi.



Gambar 1. Alur Proses Rekomendasi

Alur tersebut menunjukkan bahwa sistem bekerja secara berurutan mulai dari *input* data sampai menghasilkan rekomendasi divisi. Dengan adanya proses penyimpanan riwayat, hasil rekomendasi yang sudah diperoleh juga dapat dilihat kembali oleh *admin* melalui halaman riwayat rekomendasi.

4.7. Evaluasi Hasil

Evaluasi hasil dilakukan untuk melihat kemampuan sistem dalam memberikan rekomendasi divisi magang yang sesuai dengan karakteristik peserta. Evaluasi ini didasarkan pada hasil pengolahan 12 data peserta menggunakan metode *Content-Based Filtering* melalui tahapan *preprocessing*, pembobotan TF-IDF, dan perhitungan *cosine similarity*. Hasil rekomendasi yang diperoleh kemudian dianalisis berdasarkan kesesuaian antara profil peserta dan kebutuhan divisi.

Berdasarkan hasil pengujian, sistem menunjukkan bahwa peserta dengan minat dan keahlian pada bidang pengembangan aplikasi, seperti *frontend*, *backend*, *UI/UX*, dan *mobile development*, cenderung memperoleh rekomendasi pada divisi *IT Support*. Peserta yang memiliki kemampuan pada bidang jaringan, konfigurasi perangkat, *routing*, *LAN*, dan *troubleshooting* cenderung direkomendasikan ke divisi Teknisi. Sementara itu, peserta dengan kemampuan administrasi, audit, dan pengolahan data lebih banyak memperoleh rekomendasi pada divisi *Shared Service*. Hasil tersebut menunjukkan bahwa sistem mampu menyesuaikan rekomendasi berdasarkan kesamaan konten antara profil peserta dan karakteristik masing-masing divisi.

Nilai *similarity* tertinggi diperoleh oleh Taufik Hidayat sebesar 0.5399 pada divisi Teknisi. Nilai tersebut menunjukkan bahwa profil peserta yang memuat kata kunci jaringan dan *troubleshooting* memiliki tingkat kesesuaian yang cukup tinggi dengan kebutuhan divisi Teknisi. Di sisi lain, nilai *similarity* terendah diperoleh oleh Vina Oktaviani sebesar 0.0517 pada divisi *Shared Service*. Meskipun nilainya relatif rendah, divisi tersebut tetap menjadi rekomendasi terbaik karena memiliki nilai kemiripan tertinggi dibandingkan divisi lainnya. Kondisi ini menunjukkan bahwa hasil rekomendasi tidak hanya dipengaruhi oleh banyaknya kata yang dimiliki peserta, tetapi juga oleh tingkat relevansi kata-kata tersebut terhadap profil divisi.

Secara umum, hasil evaluasi menunjukkan bahwa metode *Content-Based Filtering* dapat digunakan untuk mendukung proses penempatan divisi magang secara lebih terarah dan objektif. Sistem mampu memberikan rekomendasi berdasarkan kemiripan konten antara data peserta dan data divisi yang tersedia. Namun demikian, kualitas hasil rekomendasi tetap dipengaruhi oleh kelengkapan dan ketepatan kata kunci pada data peserta maupun data divisi. Semakin baik representasi data yang digunakan, maka semakin baik pula hasil rekomendasi yang dapat dihasilkan oleh sistem.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, sistem rekomendasi penempatan divisi magang berbasis *Content-Based Filtering* berhasil diterapkan untuk membantu menentukan divisi yang sesuai dengan profil peserta. Sistem memanfaatkan data program studi, minat, dan keahlian peserta, kemudian membandingkannya dengan data kriteria divisi menggunakan metode TF-IDF dan *Cosine Similarity*. Hasil pengujian terhadap 12 data peserta menunjukkan bahwa sistem mampu memberikan rekomendasi yang sesuai dengan karakteristik peserta, seperti *IT Support* untuk peserta dengan minat web dan pengembangan aplikasi, Teknisi untuk peserta dengan keahlian jaringan dan *troubleshooting*, serta *Shared Service* untuk peserta dengan kemampuan administrasi dan pengolahan data. Sistem juga telah terintegrasi dengan *Firebase Firestore* untuk menyimpan data divisi dan riwayat rekomendasi. Untuk pengembangan selanjutnya, sistem dapat ditingkatkan dengan menambahkan proses *preprocessing* yang lebih lengkap, seperti *stemming* dan *stopword removal*, serta memperbanyak data divisi dan variasi kata kunci agar hasil rekomendasi menjadi lebih akurat.

DAFTAR PUSTAKA

- [1] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.
- [2] C. Koloman, R. Maulana, R. D. Z. Putri, and W. A. Harahap, "Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering," *Journal of Creative Student Research*, vol. 1, no. 6, pp. 78–88, 2023, doi: 10.55606/jcsrpolitama.v1i6.2992.
- [3] D. B. Elnursa, V. Nofriana, A. Syamsuri, and L. Cahyani, "Sistem Rekomendasi Pemilihan Program MSIB Bagi Mahasiswa Pendidikan Informatika," *Journal Software, Hardware and Information Technology*, vol. 3, no. 2, pp. 34–45, 2023, doi: 10.24252/shift.v3i2.92.
- [4] A. Fitria, S. Zaman, and M. A. Yaqin, "Sistem Rekomendasi Lowongan Pekerjaan Menggunakan Content-based Filtering," vol. 10, no. 3, pp. 421–427, 2024.
- [5] A. Rianti, N. W. A. Majid, and A. Fauzi, "Machine Learning Journal Article Recommendation System Using Content Based Filtering," 2024.
- [6] F. Dwi and W. Andriyani, "Pengembangan Sistem Rekomendasi Pembimbing Tugas Akhir Menggunakan Teknik Content Based Filtering," vol. 9, no. 2, pp. 474–483, 2025, doi: 10.26798/jiko.v9i2.1599.Type equation here.
- [7] S. Winardi, S. Megawan, N. P. Wong, R. Kurniawan, and F. A. Putra, "Utilizing TF-IDF Content-based Filtering for Job

- Recommendation Systems,” vol. 8, no. 5, pp. 2838–2845, 2025.
- [8] S. Hashim and J. Waden, “Content-based filtering algorithm in social media,” *Wasit Journal of Computer and Mathematics Science*, vol. 2, no. 1, pp. 14–17, 2023, doi: 10.31185/wjcm.112.
- [9] H. Mgarbi, M. Y. Chkouri, and A. Tahiri, “Building a recommendation system based on the job offers extracted from the web and the skills of job seekers,” *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6964–6971, 2023, doi: 10.11591/ijece.v13i6.pp6964-6971.
- [10] G. Farell, C. N. Z. Latt, N. Jalinus, A. Yulastri, and R. Wahyudi, “Analysis of Job Recommendations in Vocational Education Using the Intelligent Job Matching Model,” *International Journal on Informatics Visualization*, vol. 8, no. 1, pp. 361–367, 2024, doi: 10.62527/joiv.8.1.2201.
- [11] R. Lai, R. Chen, and C. Zhang, “A Survey on Data-Centric Recommender Systems,” 2024. [Online]. Available: <http://arxiv.org/abs/2401.17878>
- [12] D. Pratiwi, A. Asrianda, and L. Rosnita, “Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang,” *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 2, pp. 85–96, 2025, doi: 10.54082/jiki.169.