JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 203-209
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Sistem Rekomendasi Pelatihan Digitalent Scholarship Menggunakan Content Based Filetring dan Cossine Similarity

Agustina Latifatul Mutmainnah[1], Ahmad Yusuf[2], Bayu Adhi Nugroho[3]

[1,2,3] Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negri Sunan Ampel, Surabaya
Email: [1]09050622079@student.uinsby.ac.id, [2]ahmadyusuf@uinsa.ac.id, [3]bayu.lecture@gmail.com
Email Penulis Korespondensi: [1]09050622079@student.uinsby.ac.id

Abstrak− Transformasi digital nasional menuntut tersedianya talenta digital yang adaptif, sebagaimana diamanatkan dalam Peraturan Presiden Nomor 82 Tahun 2022. Program Digital Talent Scholarship (DTS) hadir untuk memenuhi kebutuhan tersebut, namun platform pelatihannya belum dapat memberikan rekomendasi yang sesuai profil pengguna. Penelitian ini mengembangkan sistem rekomendasi pelatihan berbasis hybrid yang menggabungkan rule-based filtering dan content-based filtering menggunakan TF-IDF dan Cosine Similarity. Data pelatihan dari beberapa program DTS diolah melalui prapengolahan teks dan penyaringan berdasarkan usia, pendidikan, pekerjaan, serta minat, kemudian dihitung kesamaan semantik antara input pengguna dan deskripsi pelatihan. Evaluasi menggunakan histori peserta tahun 2023 dengan metrik Top-3 Accuracy menunjukkan performa sebesar 72,73%, menandakan bahwa sebagian besar pengguna menerima minimal satu rekomendasi yang benar dalam tiga peringkat teratas. Temuan ini menunjukkan bahwa pendekatan hybrid mampu menghasilkan rekomendasi yang relevan dan personal. Ke depan, sistem dapat dikembangkan dengan embedding berbasis transformer dan pemodelan preferensi pengguna yang lebih dinamis.

Kata Kunci: Sistem Rekomendasi, Pelatihan Digital, Content-Based Filtering, TF-IDF, Cossine Similarity

Abstract− National digital transformation demands the availability of adaptive digital talent, as mandated by Presidential Regulation Number 82 of 2022. The Digital Talent Scholarship (DTS) program exists to meet this need, but its training platform is unable to provide recommendations tailored to user profiles. This research develops a hybrid training recommendation system that combines rule-based and content-based filtering using TF-IDF and Cosine Similarity. Training data from several DTS programs is processed through text preprocessing and filtering based on age, education, occupation, and interests, then semantic similarity is calculated between user input and training descriptions. Evaluation using historical participant data from 2023 with the Top-3 Accuracy metric shows a performance of 72.73%, indicating that most users receive at least one correct recommendation in the top three rankings. These findings demonstrate that the hybrid approach is capable of generating relevant and personalized recommendations. In the future, the system can be developed with transformer-based embedding and more dynamic user preference modeling.

Keywords: Recomender System, Digital training, Content-Based Filtering, TF-IDF, Cossine Similarity

1. PENDAHULUAN

Di era transformasi digital, teknologi menjadi pendorong utama globalisasi dan mendukung visi Indonesia Digital 2045. Pemerintah merespons kebutuhan ini melalui Peraturan Presiden Nomor 82 Tahun 2022 tentang Percepatan Transformasi Digital. Sebagai bagian dari implementasi kebijakan tersebut, Kominfo melalui BPSDM menyelenggarakan program Digital Talent Scholarship (DTS) sejak 2018 untuk meningkatkan keterampilan masyarakat di bidang teknologi digital, yang mencakup delapan akademi seperti DEA, TA, dan GTA. Meski telah memiliki aplikasi informasi pelatihan, sistem yang ada belum adaptif karena tidak menyesuaikan rekomendasi dengan profil pengguna seperti usia, pekerjaan, dan minat.

Permasalahan ini muncul di tengah masih lebarnya kesenjangan talenta digital nasional. Kominfo memperkirakan kekurangan sekitar 3 juta talenta digital hingga 2030[1], sementara minat masyarakat terhadap pelatihan teknologi meningkat pesat, termasuk pertumbuhan 237% pada pelatihan Generative AI. Selain itu, laporan BPSDM Kominfo [2] menegaskan bahwa akses dan relevansi pelatihan masih menjadi tantangan.

Berbagai penelitian telah mengusulkan sistem rekomendasi pelatihan untuk meningkatkan relevansi konten. Penelitian yang dilakukan Ma dkk. [3] membangun sistem DORIS yang menggabungkan data kursus dan profil pengguna, sedangkan Ahmadian Yazdi dkk. [4] menekankan pentingnya minat yang dinamis. Pendekatan lain menautkan rekomendasi dengan pekerjaan dan kebutuhan pasar kerja [5] [6]] Untuk mengatasi masalah cold-start, penelitian berbasis meta-learning, seperti jejaring sosial [7] juga telah dikembangkan.

Namun, sebagian besar pendekatan tersebut masih belum sepenuhnya mempertimbangkan konteks pekerjaan atau latar belakang pengguna. Collaborative Filtering menghadapi masalah cold-start, sedangkan Rule-Based Filtering cenderung kaku. Oleh karena itu, penelitian ini mengusulkan sistem rekomendasi pelatihan Digitalent Scholarship berbasis Content-Based Filtering menggunakan TF-IDF dan Cosine Similarity. Pendekatan ini menghubungkan profil

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **06/12/2025**; Accepted: **15/01/2026**; Published: **02/04/2026**

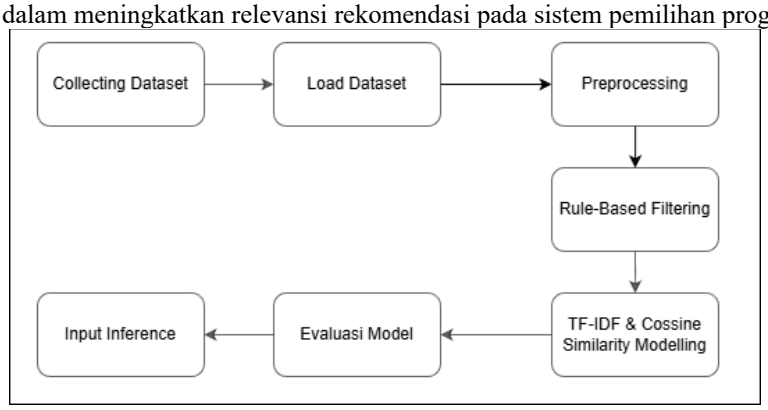
JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 203-209
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

pengguna (usia, pekerjaan, pendidikan, dan minat) dengan deskripsi pelatihan, sehingga menghasilkan rekomendasi yang lebih relevan, personal, dan adaptif untuk mendukung pengembangan talenta digital di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Metode penelitian ini dirancang untuk mengembangkan sistem rekomendasi pelatihan berbasis hybrid sederhana, yaitu menggabungkan pendekatan rule-based filtering dengan content-based filtering menggunakan Term Frequency–Inverse Document Frequency (TF-IDF) dan cosine similarity. Pendekatan ini dipilih karena mampu memfilter kandidat pelatihan secara ketat sesuai kriteria peserta sebelum menghitung kesesuaian semantik berdasarkan minat pengguna. Strategi ini sejalan dengan penelitian oleh Abdurrafi & Ningsih [8] yang menunjukkan bahwa kombinasi rule-based dan cosine similarity efektif dalam meningkatkan relevansi rekomendasi pada sistem pemilihan program pelatihan.



Gambar 1. Diagram Alir Penelitian

a) Collecting dataset
Collecting dataset adalah proses pengumpulan, pengukuran, dan analisis berbagai tipe informasi menggunakan teknik berstandar. Dalam penelitian ini, dataset diperoleh dari booklet resmi pelatihan akademi DEA, GTA, dan TA. Dataset disimpan dalam format .xlsx dengan delapan kolom utama: nama pelatihan, deskripsi, tema, usia minimal, usia maksimal, pekerjaan, pendidikan terakhir, jurusan, dan akademi penyelenggara. Total 47 data pelatihan dikumpulkan dan data histori peserta pelatihan tahun 2023 digunakan sebagai ground truth untuk mengukur keakuratan hasil rekomendasi.

b) Preprocessing Data
Setelah data dikumpulkan, proses selanjutnya yaitu preprocessing dataset. Pada tahap ini beberapa tahapan, yang meliputi normalisasi teks dengan mengubah semua huruf menjadi huruf kecil, penghapusan tanda baca, karakter non-alfabetik, spasi ganda, serta penanganan nilai kosong dengan mengisi string kosong. Kolom yang dibersihkan meliputi nama pelatihan, deskripsi, pekerjaan, dan pendidikan terakhir. Seluruh teks kemudian digabungkan menjadi satu dokumen untuk setiap baris data dan disiapkan menjadi representasi vector.

c) Rule-Based Filtering
Rule-Based System (RBS) merupakan jenis algoritma yang menggunakan teknik sederhana dengan aturan dasar yang memuat pengetahuan permasalahan. Menurut Jannachd kk. [9], pendekatan rule-based membantu mempersempit ruang pencarian rekomendasi dan meningkatkan relevansi awal sebelum dilakukan analisis semantic. Pengetahuan tersebut berupa kode dengan menjadi aturan IF-THEN dan disimpan dalam basis data. Sistem memeriksa aturan kondisi IF, menjalankan kondisi THEN jika syarat terpenuhi. Proses berlanjut hingga satu atau dua kondisi memenuhi dan jika tidak, sistem keluar dari perulangan untuk menyaring kandidat pelatihan yang sesuai dengan profil pengguna [10].

d) Penerapan algoritma TF-IDF & Cossine Similarity
Pada tahap ini algoritma TF-IDF dan Cossine Similarity digunakan untuk merepresentasikan dokumen. Setiap data pelatihan dibentuk menjadi satu dokumen yang merupakan gabungan dari kolom nama pelatihan, tema, dan deskripsi. Dokumen ini kemudian diubah menjadi vektor numerik menggunakan transformasi TF-IDF, sehingga setiap pelatihan memiliki representasi matematis berbasis bobot kata. Selanjutnya, teks minat pengguna juga ditransformasikan dengan model TF-IDF yang sama agar berada pada ruang vektor yang sebanding. Untuk menghitung tingkat kemiripan antara minat pengguna dan setiap pelatihan, digunakan cosine_similarity dari sklearn.metrics.pairwise, yang menghasilkan skor antara 0 hingga 1, di mana nilai yang lebih mendekati 1 menunjukkan tingkat kesamaan yang lebih tinggi. Terakhir, skor similarity ini dapat digabungkan dengan bobot rule-based seperti bobot berdasarkan pekerjaan atau pendidikan pengguna untuk menghasilkan skor total yang lebih relevan dengan profil pengguna secara keseluruhan.

e) Evaluasi model

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **06/12/2025**; Accepted: **15/01/2026**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 203-209
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Evaluasi sistem adalah tahap untuk menilai kinerja model rekomendasi. Data historis peserta pelatihan tahun 2023 digunakan sebagai ground truth untuk menghitung akurasi. Metode evaluasi yang digunakan adalah Top-3 Accuracy, yaitu persentase di mana salah satu dari tiga rekomendasi teratas sesuai dengan pelatihan aktual yang sudah diikuti peserta. Pendekatan top-K seperti ini dianggap relevan dalam evaluasi sistem Hybrid-Content based menggunakan hit rate@K dan precision@K untuk menguji apakah item relevan muncul di urutan atas rekomendasi. Evaluasi model ini cukup penting dilakukan untuk memvalidasi apakah model benar-benar berjalan pada data baru dengan baik.

f) Hasil dan interpretasi

Tahap terakhir adalah menyajikan model yang sudah dievaluasi. Sistem dijalankan dengan melakukan input interaktif profil pengguna yang memungkinkan sistem digunakan secara berulang dan personal. Pengguna diminta mengisi usia, pendidikan terakhir, pekerjaan, dan minat melalui antarmuka sederhana berbasis Python. Profil pengguna ini diproses secara otomatis melalui tahapan filtering dan perhitungan kesamaan semantik, kemudian menghasilkan daftar pelatihan yang relevan. Dengan pendekatan ini, sistem mampu memberikan rekomendasi yang lebih tepat sasaran berdasarkan kombinasi logika aturan dan analisis kesamaan konten.

2.2 Rule-Based Filtering

Penerapan algoritma ini menjadi konteks penyaringan awal system rekomendasi, metode ini digunakan untuk mempersempit kandidat pelatihan yang sesuai dengan profil pengguna. Aturan yang diterapkan dalam tahap ini meliputi penyaringan usia, kesesuaian dengan pendidikan terakhir, pencocokan jurusan dan kecocokan pekarjaan menggunakan cosine similarity. Sehingga rules IF-THEN yang diterapkan berupa:

$$R = \{p \in P | C_{usia}(p) \wedge C_{pendidikan}(p) \wedge C_{jurusan}(p) \wedge C_{pekerjaan}(p)\} \quad (1)$$

Keterangan:

P = seluruh data pelatihan

R = himpunan pelatihan yang lolos aturan

$C_{usia}(p)$ = *true* jika $usia_min \leq usia_user \leq usia_max$

$C_{pendidikan}(p)$ = *true* jika pendidikan pengguna memenuhi syarat

$C_{jurusan}(p)$ = *true* jika jurusan cocok (exact/fuzzy)

$C_{pekerjaan}(p)$ = *true* jika cosine similarity pekerjaan ≥ ambang yang ditentukan

2.3 Penerapan Algoritma TF-IDF dan Cossine Similarity

TF-IDF (Term Frequency–Inverse Document Frequency) adalah salah satu model statistik paling populer dan banyak digunakan dalam pemrosesan informasi teks dan penambangan data. Model ini berfungsi untuk mengevaluasi tingkat kepentingan atau signifikansi suatu kata terhadap suatu dokumen dalam kumpulan dokumen (corpus). Menurut Huda dkk. [11] Nilai signifikansi dari sebuah kata meningkat bersamaan dengan frekuensi kemunculan pada sebuah teks, namun berbanding terbalik dengan meningkatnya frekuensi kemunculan dalam kumpulan teks. TF-IDF adalah algoritma skema pemberatan yang paling populer.

Berikut adalah formula dari TF-IDF:

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

Dengan rumus TF:

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (3)$$

dan rumus IDF:

$$IDF(t) = \log\left(\frac{1+N}{1+N}\right) + 1 \quad (4)$$

Sedangkan Cossine similarity Metode cosine similarity merupakan metode perhitungan jarak antara A dan B yang menghasilkan sudut cosine x di antara kedua vektor. Nilai sudut ini menentukan kesamaan dua buah objek yang dibandingkan. Perhitungan dalam cossine sendiri mempunyai nilai terkecil 0 dan nilai terbesar 1. dalam penggunaan

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **06/12/2025**; Accepted: **15/01/2026**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 203-209
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

metode cosine similarity, 2 vektor yang dibandingkan disini adalah informasi profil pengguna dan deskripsi pelatihan. Dengan rumus perhitungan:

$$Cossine\ Similarity\ (A,B) = \frac{A.B}{\|A\|\|B\|} \quad (5)$$

3. HASIL DAN PEMBAHASAN

Implementasi algoritma Hybrid dengan Rule based filtering dan TF-IDF ini digunakan untuk membuat system rekomendasi berbasis teks berdasarkan profil pengguna calon peserta pelatihan. System ini bertujuan untuk memudahkan pengguna dalam memilih jenis tema pelatihan yang sesuai dengan profil mereka secara relevan dan mudah. Berikut beberapa langkah dan hasil dalam penerapan algoritma hybrid-content based.Pada bagian ini dijelaskan hasil-hasil penelitian dan sekaligus diberikan pembahasan yang komprehensif.

3.1 Load Dataset

Load dataset dengan menggunakan google drive dalam komputasi google colab. Penggunaan drive ini memudahkan dalam mengambil data dalam cloud. Dataset yang digunakan meliputi;

a) Dataset teks pelatihan

index	id_pelatihan	nama_pelatihan	deskripsi_pelatihan	tema	usia_min	usia_max	pendidikan terakhir	Jurusan	pekerjaan	Akademi
0	1	Aplikasi Perkantoran bagi Guru dan Tenaga Kependidikan	Meningkatkan keterampilan penggunaan aplikasi perkantoran untuk guru dan tenaga kependidikan.	Aplikasi Perkantoran	18	50	SMA/Sederajat	Semua Jurusan	Guru, Tenaga Kependidikan	Thematic Academy
1	2	Basic Cyber Security untuk Pelajar Sekolah Menengah	Pengenalan dasar keamanan siber untuk pelajar sekolah menengah.	Cyber Security	15	19	SMA/Sederajat	Semua Jurusan	Pelajar Sekolah Menengah	Thematic Academy
2	3	Data Analytics untuk Pelajar Sekolah Menengah	Pengenalan data analytics dan machine learning untuk pelajar sekolah menengah.	Data Analytics	15	19	SMA/Sederajat	Semua Jurusan	Pelajar Sekolah Menengah	Thematic Academy
3	4	Desain Grafis bagi Guru dan Tenaga Kependidikan	Pelatihan desain grafis untuk guru, tenaga kependidikan, dan pengelola instansi pendidikan.	Desain Grafis	18	50	SMA/Sederajat	Semua Jurusan	Guru, Tenaga Kependidikan, atau Pengelola Pendidikan	Thematic Academy
4	5	Enhancing Digital Talents: Mastering Generative AI	Mengembangkan talenta digital dalam bidang Generative AI.	Generative AI	17	49	SMA/Sederajat	Semua Jurusan	Terbuka untuk umum	Thematic Academy
5	6	Pemanfaatan AI untuk Pemasaran Digital di Sektor Pariwisata	Pemanfaatan AI untuk meningkatkan pemasaran digital di sektor pariwisata.	AI untuk Pemasaran Digital	18	50	SMA/Sederajat	Semua Jurusan	Pelaku usaha di sektor pariwisata	Thematic Academy

Gambar 2. Load Dataset Pelatihan

b) Dataset Histori Peserta Pelatihan 2023

	No.	Akademi	Minat	Pelatihan	usia	pekerjaan	pendidikan terakhir
0	6897	Government Transformation Academy	Digital	Digital Public Relations	29	Pegawai Negeri Sipil	S1
1	6897	Government Transformation Academy	Digital	Digital Public Relations	34	Pegawai Negeri Sipil	D3
2	6897	Government Transformation Academy	Digital	Digital Public Relations	31	Pegawai Negeri Sipil	D3
3	6897	Government Transformation Academy	Digital	Digital Public Relations	26	Pegawai Negeri Sipil	D3
4	6897	Government Transformation Academy	Digital	Digital Public Relations	41	Pegawai Negeri Sipil	D3
5	6897	Government Transformation Academy	Digital	Digital Public Relations	46	Pegawai Negeri Sipil	D3
6	6897	Government Transformation Academy	Digital	Digital Public Relations	44	Pegawai Negeri Sipil	D3
7	6897	Government Transformation Academy	Digital	Digital Public Relations	30	Pegawai Negeri Sipil	D3
8	6897	Government Transformation Academy	Digital	Digital Public Relations	46	Pegawai Negeri Sipil	D3
9	6897	Government Transformation Academy	Digital	Digital Public Relations	26	Pegawai Negeri Sipil	D3

Gambar 3. Load Dataset Histori Pelatihan

3.2 Preprocesing

Setelah load dataset, tahap selanjutnya yaitu preprocessing. Prepocessing yang dilakukan dalam sistem rekomendasi ini ada beberapa tahapan, yaitu:

1. Lowercasing (mengonversi data frame ke huruf kecil)
   text = text.lower()
2. Punctuation Removal (Menghapus Tanda baca)
   text = re.sub(rf"[{string.punctuation}]", "", text)
3. Whitespace Normalization (Menghapus Spasi Berlebih)
   text = re.sub(r"\s+", " ", text).strip()

Proses preprocessing ini digunakan untuk mempersiapkan data agar dapat diolah membersihkan dan menormalkan data teks sebelum digunakan untuk *modelling system* dan perhitungan *similarity*.

3.3 Rule-Based System

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **06/12/2025**; Accepted: **15/01/2026**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 203-209
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Dalam fungsi rule-based system, berfungsi untuk memeriksa apakah usia berada dalam rentang usia dataset pelatihan. Lalu data yang lolos dalam rentang usia akan difilter lebih lanjut untuk dicocokkan dengan *pendidikan* terakhir user. Pendidikan ini disyaratkan dengan syarat minimum dari dataset pelatihan. Lalu user yang *lolos* filter pendidikan akan diteruskan pada pekerjaan apakah termasuk kedalam *rule based* sebagai persyaratan pekerjaan atau tidak. Pendekatan rule-based ini berfungsi sebagai penyaring awal agar hanya pelatihan yang sesuai secara objektif (usia, pendidikan, pekerjaan) yang dilanjutkan ke tahap perhitungan kemiripan teks.

3.4 TF-IDF Vectorization

Dalam penerapan algoritma TF-IDF ini, system terlebih dahulu menggabungkan beberapa kolom teks (nama pelatihan, tema, dan deskripsi pelatihan) kedalam satu fitur baru "text_minat". Fitur ini kemudian diproses menggunakan *TfidfVectorizer* untuk menghasilkan representasi numerik berbasis bobot kata. Pada tahap ini, perintah dibawah akan membangun model TF-IDIF berdasarkan seluruh data pelatihan yang telah melewati proses penyaringan awal (berdasarkan usia, pendidikan, dan pekerjaan). Selanjutnya, minat pengguna yang diberikan dalam bentuk teks bebas (*user_minat*) dikonversi menjadi vektor dengan memanggil tfidf_minat = vectorizer.transform([user_minat]). Tahap *transform* ini memastikan bahwa masukan pengguna direpresentasikan dalam ruang vektor yang sama dengan dokumen pelatihan, sehingga memungkinkan perhitungan kemiripan antar-vektor pada tahap selanjutnya.

3.5 Cosine Similarity

Pada tahap perhitungan kemiripan, sistem menggunakan cosine similarity untuk membandingkan minat pengguna dengan seluruh pelatihan. Fungsi terlebih dahulu memeriksa apakah user_minat kosong; jika kosong, proses similarity dilewati dan seluruh pelatihan diberi skor 0. Jika tidak kosong, teks user_minat diubah menjadi vektor TF-IDF melalui `vectorizer.transform()`, sehingga berada dalam ruang vektor yang sama dengan dokumen pelatihan. Perhitungan `cosine_similarity (tfidf_minat, tfidf_matrix)[0]` menghasilkan array satu dimensi yang berisi skor kemiripan untuk setiap pelatihan; misalnya nilai [0. 87, 0.63, 0.12, 0.00] menunjukkan bahwa pelatihan pertama paling sesuai dengan minat pengguna. Indeks [0] digunakan karena keluaran asli fungsi cosine similarity berbentuk matriks dua dimensi berukuran (1 × jumlah_pelatihan).

3.6 Evaluasi Model

Pada tahap evaluasi akurasi, sistem terlebih dahulu mengambil kolom “Pelatihan” dari df_test sebagai data ground truth, kemudian melakukan reset index untuk menghasilkan kolom "id_user" yang akan digunakan sebagai penghubung dengan hasil rekomendasi. Selanjutnya, DataFrame hasil rekomendasi (df_rekomendasi) digabungkan dengan ground truth menggunakan fungsi merge berdasarkan "id_user" sehingga setiap rekomendasi dapat dipasangkan dengan jawaban pelatihan yang benar. Sistem kemudian menambahkan kolom baru bernama Benar, yang bernilai True apabila nama pelatihan dari ground truth ditemukan dalam nama pelatihan hasil rekomendasi; proses pencocokan dilakukan secara case-insensitive menggunakan metode `lower()`. Langkah berikutnya adalah menghitung akurasi Top-3 melalui `groupby("id_user")["Benar"].max().mean()`, di mana `max()` digunakan untuk memeriksa apakah minimal satu dari tiga rekomendasi teratas sesuai dengan ground truth untuk setiap pengguna, dan `.mean()` digunakan untuk memperoleh nilai rata-rata keseluruhan pengguna. Hasil akurasi kemudian ditampilkan dalam format persentase menggunakan `print(f"Akurasi Top-3: {akurasi_top3:.2%}")`. Nilai akurasi Top-3 ini menjadi indikator performa sistem rekomendasi dalam memberikan saran pelatihan yang relevan di posisi teratas.

3.7 Input Interaktif User

Contoh input interaktif *user profile*

```
# Contoh uji coba
hasil_dengan_skor = rekomendasi_rule_based_dengan_skor(
    user_usia=25,
    user_pendidikan="D3",
    user_pekerjaan="pns",
    user_minat="digital"
)

hasil_dengan_skor.head(5)
```

Gambar 4. Contoh Input Profile User

Dalam fungsi diatas, dapat diketahui merupakan fungsi untuk menginputkan profile user baru dari peserta pelatihan, sehingga hasil akan ditampilkan pelatihan pelatihan yang sesuai dengan sistem. Hasil akan menunjukkan naman pelatihan, deskripsi, dan score similairty dalam setiap pelatihan yang direkomendasikan

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **06/12/2025**; Accepted: **15/01/2026**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 203-209
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

	nama_pelatihan	tema	Akademi	pekerjaan	Skor_Minat	Skor_Pekerjaan	Skor_Total
0	Digital Public Relations	Relasi Publik	Government Transformation Academy	Aparatur Sipil Negara	0.405469	1.0	0.643281
1	Fasilitator Pembelajaran Digital	Fasilitator Pembelajaran	Government Transformation Academy	Aparatur Sipil Negara	0.344425	1.0	0.606655
2	Fasilitator Pembelajaran Digital Menengah	Fasilitator Pembelajaran	Government Transformation Academy	Aparatur Sipil Negara	0.195184	1.0	0.517110
3	Analisis Kota Cerdas - ISO 37122:2019	Analisis Kota Cerdas	Government Transformation Academy	Aparatur Sipil Negara	0.000000	1.0	0.400000

Gambar 5. Hasil Rekomendasi Profile User

Hasil diatas juga menunjukkan bahwa sistem dapat merekomendasikan pelatihan dengan hasil skor similarity yang urut dari skor similarity yang paling besar. Dalam gambar diatas juga dapat diketahui bahwa pekerjaan cocok dengan input user.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sebuah sistem rekomendasi pelatihan Digital Talent Scholarship (DTS) yang bersifat adaptif dan personal, sebagai solusi terhadap keterbatasan platform pelatihan yang sebelumnya belum mampu memberikan rekomendasi berdasarkan karakteristik dan kebutuhan peserta. Sistem rekomendasi yang dibangun mengintegrasikan pendekatan rule-based filtering dengan content-based filtering berbasis TF-IDF dan cosine similarity, sehingga mampu memadukan penyaringan awal yang berbasis aturan objektif dengan analisis semantik konten pelatihan.

Evaluasi kinerja dilakukan menggunakan histori peserta tahun 2023 sebagai ground truth dan mengacu pada metrik Top-3 Accuracy. Hasil evaluasi menunjukkan bahwa sistem mencapai akurasi Top-3 sebesar (masukkan nilai akurasi Anda, misalnya 78% atau 0.78), yang berarti sebagian besar pengguna menerima setidaknya satu rekomendasi yang benar dari tiga rekomendasi teratas. Nilai akurasi tersebut menegaskan bahwa mekanisme kombinasi rule-based dan similarity berbasis TF-IDF mampu memberikan rekomendasi yang efektif dan berkualitas.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan hybrid sederhana dapat menjadi solusi yang efisien untuk sistem rekomendasi pelatihan, terutama dalam upaya meningkatkan kesesuaian pelatihan dengan kebutuhan peserta. Sistem ini juga mendukung digitalisasi layanan serta peningkatan kualitas pelayanan pelatihan di bawah program DTS. Ke depan, hasil penelitian ini dapat dikembangkan lebih jauh melalui integrasi model semantik modern seperti transformer-based embeddings, pembobotan adaptif berdasarkan perilaku pengguna, maupun rekomendasi lintas-akademi secara real time

REFERENCES

[1] Intimedia, “Kominfo Releases the 2024 Indonesia Digital Society Index:” Diakses: 4 Desember 2025. [Daring]. Tersedia pada: https://intimedia.id/read/digital-talent-gap-in-indonesia-kominfo-releases-the-2024-indonesia-digital-society-index Intimedia

[2] BPSDM Kominfo, “Audiensi BPSDM Kominfo dengan Oracle terkait Pengembangan Talenta Digital.” Diakses: 4 Desember 2025. [Daring]. Tersedia pada: https://bpsdm.komdigi.go.id/berita-pengembangan-talenta-digital-dalam-menghadapi-tantangan-pada-era-perubahan-46-1926 BPSDMD Komunikasi & Digital

[3] Y. Ma, R. Ouyang, X. Long, Z. Gao, T. Lai, dan C. Fan, “DORIS: Personalized course recommendation system based on deep learning,” *PLoS One*, vol. 18, no. 6, hlm. e0284687-, Jun 2023, [Daring]. Tersedia pada: https://doi.org/10.1371/journal.pone.0284687

[4] H. Ahmadian Yazdi, S. J. Seyyed Mahdavi, dan H. Ahmadian Yazdi, “Dynamic educational recommender system based on Improved LSTM neural network,” *Sci Rep*, vol. 14, no. 1, Des 2024, doi: 10.1038/s41598-024-54729-y.

[5] M. Tavakoli, A. Faraji, J. Vrolijk, M. Molavi, S. T. Mol, dan G. Kismihók, “An AI-based open recommender system for personalized labor market driven education,” *Advanced Engineering Informatics*, vol. 52, Apr 2022, doi: 10.1016/j.aei.2021.101508.

[6] S. A. Alsaif, M. Sassi Hidri, H. A. Eleraky, I. Ferjani, dan R. Amami, “Learning-Based Matched Representation System for Job Recommendation,” *Computers*, vol. 11, no. 11, Nov 2022, doi: 10.3390/computers11110161.

[7] Journal of Big Data, “Accuracy improvements for cold-start recommendation problem using indirect relations in social networks.” Diakses: 4 Desember 2025. [Daring]. Tersedia pada: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00484-0 Intimedia

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **06/12/2025**; Accepted: **15/01/2026**; Published: **02/04/2026**

[8] M. Abdurrafi dan D. Untari Ningsih, "Content-based filtering using cosine similarity algorithm for alternative selection on training programs," *Journal of Soft Computing Exploration*, vol. 4, hlm. 204–212, Des 2023, doi: 10.52465/joscex.v4i4.232.

[9] D. Jannach, A. Manzoor, W. Cai, dan L. Chen, "A Survey on Conversational Recommender Systems," *ACM Comput. Surv.*, vol. 54, no. 5, Mei 2021, doi: 10.1145/3453154.

[10] R. R. Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, hlm. 302–311, Mei 2020, doi: 10.30812/matrik.v19i2.688.

[11] A. A. Huda, R. Fajarudin, dan A. Hadinegoro, "Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Des 2022, doi: 10.47065/bits.v4i3.2511.

This is an open access article under the CC–BY-SA license