

IMPLEMENTASI CONTENT-BASED FILTERING UNTUK REKOMENDASI PROGRAM STUDI BERDASARKAN MINAT DAN PRESTASI AKADEMIK SISWA

Rivaldi Faqih Rahmawan^{1*}, Aprilisa Arum Sari², Hanifah Permatasari³

¹Teknik Informatika

Universitas Duta Bangsa Surakarta

^{1*}220103072@mhs.udb.ac.id

²Teknik Informatika

Universitas Duta Bangsa Surakarta

²aprilisa_arumsari@udb.ac.id

³Sistem Informasi

Universitas Duta Bangsa Surakarta

³hanifah_permatasari@udb.ac.id

Abstrak— Pemilihan program studi kuliah merupakan salah satu keputusan penting bagi siswa SMA karena berpengaruh terhadap arah pendidikan dan karier di masa depan. Namun, siswa sering mengalami kesulitan dalam menentukan program studi yang sesuai dengan bidang minat dan kemampuan akademiknya. Penelitian ini bertujuan untuk menerapkan metode *Content-Based Filtering* pada sistem rekomendasi program studi berdasarkan data alumni SMA Negeri 1 Nguter tahun 2025. Data yang digunakan meliputi bidang minat, program studi alumni, mata pelajaran pendukung, nilai mata pelajaran pendukung, dan rata-rata rapor. Proses rekomendasi dilakukan dengan membentuk profil alumni dan profil siswa menjadi fitur teks, kemudian dilakukan pembobotan menggunakan *Term Frequency-Inverse Document Frequency* atau TF-IDF dan perhitungan kemiripan menggunakan *Cosine Similarity*. Sistem menghasilkan tiga rekomendasi program studi dengan nilai kemiripan tertinggi. Berdasarkan contoh pengujian pada bidang minat Teknologi, sistem merekomendasikan Sistem Informasi, Teknik Informatika, dan Ilmu Komputer dengan nilai *Precision@3* sebesar 100%. Hasil tersebut menunjukkan bahwa metode *Content-Based Filtering* dapat digunakan untuk memberikan rekomendasi program studi yang relevan berdasarkan bidang minat dan karakteristik akademik siswa.

Kata kunci— sistem rekomendasi, program studi, cosine similarity, content-based filtering, TF-IDF.

Abstract— Choosing a college major is one of the most critical decisions for high school students, as it significantly impacts their future education and career paths. However, students frequently face difficulties in determining a major that aligns with their field of interest and academic capabilities. This study aims to apply the Content-Based Filtering method to a college major recommendation system based on the data of SMA Negeri 1 Nguter alumni from 2025. The data used includes fields of interest, alumni majors, supporting subjects, supporting subject scores, and report card averages. The recommendation process is carried out by converting alumni and student profiles into text features, followed by weighting using Term Frequency-Inverse Document Frequency (TF-IDF) and similarity calculation using Cosine Similarity. The system generates the top three college major recommendations with the highest similarity scores. Based on a sample test within the Technology interest field, the system recommended Information Systems, Informatics Engineering, and Computer Science, achieving a *Precision@3* score of 100%. These results indicate that the Content-Based Filtering method can be effectively used to provide relevant college major recommendations based on students' fields of interest and academic characteristics.

Keywords— recommendation system, college major, cosine similarity, content-based filtering, TF-IDF.

I. PENDAHULUAN

Pemilihan jurusan kuliah merupakan salah satu keputusan penting bagi siswa Sekolah Menengah Atas karena berpengaruh terhadap arah Pendidikan lanjutan dan peluang karier di masa depan. Namun, dalam praktiknya masih banyak siswa yang mengalami kesulitan dalam menentukan jurusan kuliah yang sesuai dengan minat, kemampuan akademik, serta prospek karier yang diharapkan. Kesalahan dalam memilih jurusan dapat berdampak pada rendahnya motivasi belajar, ketidaksesuaian bidang studi dengan kemampuan siswa, bahkan

berpotensi menyebabkan mahasiswa merasa tidak cocok dengan jurusan yang telah dipilih. Kondisi tersebut menunjukkan bahwa siswa membutuhkan sistem yang dapat memberikan pertimbangan secara lebih objektif dalam proses pemilihan jurusan kuliah [1].

SMA Negeri 1 Nguter merupakan salah satu sekolah yang memiliki kebutuhan dalam membantu siswa menentukan pilihan jurusan kuliah yang sesuai dengan karakteristik masing-masing siswa. Selama ini, proses pemilihan jurusan umumnya masih dipengaruhi oleh saran orang tua, guru Bimbingan Konseling, teman

sebagai, maupun persepsi siswa terhadap jurusan tertentu. Dengan adanya sistem rekomendasi, siswa dapat memperoleh alternatif jurusan kuliah yang lebih sesuai berdasarkan data yang dimiliki [2].

Sistem rekomendasi merupakan sistem yang digunakan untuk memberikan saran atau pilihan kepada pengguna berdasarkan karakteristik tertentu. Salah satu metode yang dapat digunakan dalam sistem rekomendasi adalah *Content-Based Filtering*. Metode ini bekerja dengan membandingkan profil pengguna dengan karakteristik item yang direkomendasikan. Dalam konteks pemilihan jurusan kuliah, profil pengguna dapat berupa data minat dan nilai akademik siswa, sedangkan item yang direkomendasikan berupa data program studi atau jurusan kuliah. Pendekatan ini dapat digunakan untuk menghasilkan rekomendasi berdasarkan tingkat kesesuaian antara karakteristik siswa dan karakteristik jurusan [3].

Pada penelitian ini, metode *Content-Based Filtering* diterapkan dengan menggunakan pembobotan *Term Frequency-Inverse Document Frequency* atau TF-IDF dan perhitungan kemiripan menggunakan *Cosine Similarity*. TF-IDF digunakan untuk memberikan bobot terhadap fitur atau kata yang merepresentasikan karakteristik jurusan dan profil siswa, sedangkan *Cosine Similarity* digunakan untuk menghitung tingkat kemiripan antara profil siswa dengan data jurusan. Semakin tinggi nilai kemiripan yang diperoleh, maka semakin tinggi pula tingkat kesesuaian jurusan tersebut untuk direkomendasikan kepada siswa [4].

Beberapa penelitian sebelumnya telah menerapkan metode *Content-Based Filtering* dengan TF-IDF dan *Cosine Similarity* pada berbagai sistem rekomendasi, seperti rekomendasi artikel berita, objek wisata, dan komunitas akademik. Namun, penelitian-penelitian tersebut umumnya berfokus pada satu jenis data atau domain tertentu, sehingga belum secara khusus mengintegrasikan profil minat dan nilai akademik siswa dalam rekomendasi pemilihan jurusan kuliah. Oleh karena itu, penelitian ini berfokus pada pengembangan sistem rekomendasi pemilihan jurusan kuliah

dengan menggabungkan data minat dan nilai akademik siswa sebagai dasar dalam menentukan rekomendasi [5].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk merancang dan membangun sistem rekomendasi pemilihan jurusan kuliah berbasis *Content-Based Filtering* dengan integrasi profil minat dan nilai akademik siswa di SMA Negeri 1 Nguter. Sistem yang dikembangkan berbentuk *prototype* berbasis web yang dapat memberikan rekomendasi jurusan berdasarkan tingkat kesesuaian tertinggi. Selain itu, penelitian ini juga melakukan evaluasi terhadap hasil rekomendasi menggunakan metrik *precision* dan *recall* untuk mengetahui tingkat ketepatan sistem dalam memberikan rekomendasi jurusan kepada siswa [6].

II. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini mengikuti tahapan terstruktur mulai dari analisis data siswa dan jurusan, penerapan komputasi kemiripan fitur, hingga pengujian hasil rekomendasi yang dihasilkan.

A. Jenis dan Sumber Data

Penelitian ini menggunakan data primer dan data sekunder. Data primer diperoleh dari data siswa SMA Negeri 1 Nguter yang meliputi profil minat dan nilai akademik pada mata pelajaran tertentu yang relevan dengan pemilihan jurusan kuliah. Data tersebut digunakan sebagai dasar pembentukan profil siswa dalam sistem rekomendasi.

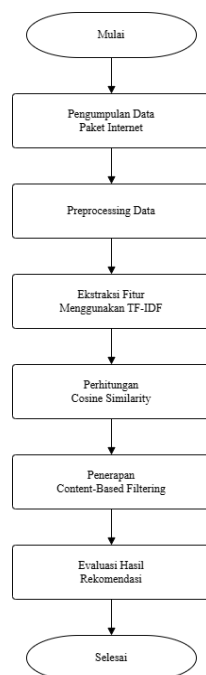
Data sekunder diperoleh dari informasi program studi atau jurusan kuliah yang bersumber dari situs resmi perguruan tinggi dan sumber informasi pendidikan yang relevan. Data jurusan meliputi nama jurusan, deskripsi jurusan, bidang keilmuan, serta kompetensi yang dibutuhkan. Data ini digunakan sebagai karakteristik item yang akan dibandingkan dengan profil siswa.

Metode yang digunakan dalam penelitian ini adalah *Content-Based Filtering*. Metode ini digunakan untuk mencocokkan profil minat dan nilai akademik siswa dengan karakteristik jurusan kuliah. Proses pembobotan fitur dilakukan menggunakan *Term Frequency-*

Inverse Document Frequency atau TF-IDF, sedangkan perhitungan tingkat kemiripan dilakukan menggunakan *Cosine Similarity*.

B. Tahapan Penelitian

Jumlah halaman paper yang diajukan antara 6 sampai dengan 12 halaman Tahapan penelitian dilakukan secara terstruktur mulai dari pengumpulan data hingga evaluasi hasil rekomendasi. Alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Flowchart Tahapan Penelitian

Pada gambar diatas menunjukkan bahwa tahapan penelitian dilakukan secara sistematis, seperti penjelasan berikut ini:

1. Pengumpulan Data

Tahap pertama adalah pengumpulan data yang dibutuhkan dalam sistem rekomendasi. Data yang dikumpulkan meliputi data profil minat siswa, nilai akademik siswa, serta data jurusan kuliah. Data profil minat dan nilai akademik digunakan sebagai profil pengguna, sedangkan data jurusan digunakan sebagai item yang akan direkomendasikan.

2. Preprocessing Data

Data yang telah dikumpulkan kemudian melalui tahap *preprocessing*. Pada tahap ini

dilakukan proses pengolahan teks yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Proses ini bertujuan untuk membersihkan data teks agar lebih terstruktur dan siap digunakan pada tahap pembobotan.

3. Ekstraksi Fitur Menggunakan TF-IDF

Setelah proses *preprocessing*, data teks diubah ke dalam bentuk vektor menggunakan metode TF-IDF. Metode ini digunakan untuk memberikan bobot pada setiap kata atau fitur berdasarkan tingkat kepentingannya dalam dokumen. Hasil dari proses ini berupa representasi vektor dari profil siswa dan karakteristik jurusan.

4. Perhitungan Kemiripan Menggunakan *Cosine Similarity*

Tahap selanjutnya adalah Menghitung tingkat kemiripan antara profil siswa dan karakteristik jurusan menggunakan *Cosine Similarity*. Nilai kemiripan digunakan untuk mengetahui seberapa sesuai suatu jurusan dengan profil siswa. Jurusan yang memiliki nilai kemiripan tertinggi akan menempati peringkat rekomendasi teratas.

5. Penerapan *Content-Based Filtering*

Metode *Content-Based Filtering* digunakan untuk menghasilkan daftar rekomendasi jurusan berdasarkan hasil perhitungan kemiripan. Sistem akan menampilkan beberapa jurusan dengan nilai kesesuaian tertinggi sebagai rekomendasi kepada siswa. Hasil rekomendasi ini digunakan sebagai alat bantu dalam pengambilan keputusan pemilihan jurusan kuliah.

6. Evaluasi Hasil

Tahap terakhir adalah evaluasi hasil rekomendasi. Evaluasi dilakukan untuk mengetahui tingkat ketepatan sistem dalam memberikan rekomendasi jurusan. Pengujian dilakukan menggunakan metrik *precision* dan *recall* dengan membandingkan hasil rekomendasi sistem terhadap data historis atau pilihan jurusan yang relevan.

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan menggunakan data alumni SMA Negeri 1 Nguter tahun 2025. Data yang digunakan berjumlah 35 data alumni yang telah

disusun ke dalam dataset. Dataset tersebut digunakan sebagai dasar dalam proses rekomendasi program studi menggunakan metode *Content-Based Filtering*.

Tabel 1. Contoh Struktur Dataset Alumni

id_alu mni	bidang_m inat	prodi_alu mni	mapel_ 1	nilai_ 1	mapel_ 2	nilai_ 2	rata_rata_r apor
A001	Kesehatan	Keperawat	Biologi	90	Kimia	89	86.50
A002	Teknologi	Sistem	Matemat	88	Informat	87	87.43
A003	Ekonomi	Akuntansi	Ekonom	87	Matemat	86	86.64
A004	Sosial dan	Psikologi	Sosiolog	90	Biologi	88	90.11
A005	Seni dan	Desain	Seni	86	Informat	85	88.81
	desain	Komunika	budaya		ika		
		si Visual					

Pada Tabel 1. Dapat diketahui bahwa data alumni yang dikumpulkan meliputi id alumni, bidang minat, program studi alumni, mata pelajaran pendukung pertama, nilai mata pelajaran pendukung pertama, mata pelajaran pendukung kedua, nilai mata pelajaran pendukung kedua, serta rata-rata nilai rapor.

B. Hasil Pengolahan Data

Tahap pengolahan data dilakukan dengan memeriksa dan menyeragamkan atribut pada dataset alumni, seperti bidang_minat, prodi_alumni, mapel_1, dan mapel_2. Data numerik berupa nilai_1, nilai_2, dan rata_rata_rapor disusun dalam bentuk angka agar dapat digunakan pada proses perhitungan. Nilai akademik kemudian dikategorikan ke dalam bentuk teks agar dapat diproses menggunakan metode TF-IDF. Karena data alumni yang digunakan memiliki nilai mata pelajaran pendukung minimal 85, maka kategori nilai dibuat lebih rapat seperti pada Tabel 2.

Tabel 2. Kategori Nilai Akademik

Rentang Nilai	Kategori
89-100	SangatTinggi
87-88	Tinggi
85-86	CukupTinggi

Pada Tabel 2, dapat diambil contoh seperti data alumni dengan bidang minat Teknologi, program studi Sistem Informasi, nilai Matematika 88, nilai Informatika 87, dan rata-rata rapor 86,75

dibentuk menjadi profil fitur: teknologi sistem informasi matematika tinggi informatika tinggi rata rapor cukup tinggi. Profil fitur tersebut digunakan sebagai dokumen dalam proses pembobotan TF-IDF dan perhitungan kemiripan menggunakan *Cosine Similarity*.

C. Hasil Pembobotan TF-IDF

Pada tahap ini dilakukan pembobotan setiap *term* menggunakan metode *Term Frequency-Inverse Document Frequency* terhadap profil alumni yang telah dibentuk pada tahap sebelumnya. Pembobotan TF-IDF bertujuan untuk mengetahui tingkat kepentingan setiap *term* dalam merepresentasikan karakteristik masing-masing profil alumni. Untuk mempermudah penyajian hasil, contoh perhitungan manual menggunakan satu data alumni, sedangkan proses perhitungan DF dan IDF tetap dilakukan terhadap seluruh 35 data alumni.

Tabel 3. Contoh Pembentukan Profil Alumni

ID Alumni	Profil
A001	Kesehatan Biologi 90 Kimia 85 CukupTinggi Keperawatan
A002	Teknologi Matematika 88 Informatika 87 Tinggi Sistem Informasi
A003	Teknologi Matematika 89 Fisika 86 Tinggi Teknik Elektro

Pada Tabel 3, dapat Dilihat bahwa setiap data alumni dibentuk menjadi sebuah profil teks dengan menggabungkan bidang minat, mata pelajaran pendukung, nilai mata pelajaran, kategori nilai akademik, dan program studi alumni.

Sebagai contoh digunakan profil alumni A002 yang terdiri atas tujuh *term*, yaitu Teknologi, Matematika, 88, Informatika, 87, CukupTinggi, dan SistemInformasi. Karena setiap *term* hanya muncul satu kali pada dokumen dengan jumlah keseluruhan tujuh *term*, maka nilai *Term Frequency* dihitung sebagai berikut.

$$TF(Teknologi) = \frac{1}{7} = 0,143$$

$$TF(Matematika) = \frac{1}{7} = 0,143$$

$$TF(88) = \frac{1}{7} = 0,143$$

$$TF(Informatika) = \frac{1}{7} = 0,143$$

$$TF(CukupTinggi) = \frac{1}{7} = 0,143$$

$$TF(SistemInformasi) = \frac{1}{7} = 0,143$$

Selanjutnya dilakukan perhitungan *Document Frequency* dan *Inverse Document Frequency*. Jumlah dokumen yang digunakan sebanyak 35 data alumni. Berdasarkan hasil perhitungan pada seluruh dokumen, *term* Teknologi muncul pada 10 dokumen, Matematika pada 17 dokumen, 88 pada 11 dokumen, Informatika pada 5 dokumen, 87 pada 11 dokumen, CukupTinggi pada 11 dokumen, dan SistemInformasi pada 1 dokumen. Nilai IDF dihitung menggunakan Persamaan sebagai berikut.

$$IDF(Teknologi) = \log\left(\frac{35}{10}\right) = 0,544$$

$$IDF(Matematika) = \log\left(\frac{35}{17}\right) = 0,314$$

$$IDF(88) = \log\left(\frac{35}{11}\right) = \log(2,5) = 0,503$$

$$IDF(Informatika) = \log\left(\frac{35}{5}\right) = 0,845$$

$$IDF(87) = \log\left(\frac{35}{11}\right) = \log(2,5) = 0,503$$

$$IDF(CukupTinggi) = \log\left(\frac{35}{11}\right) = 0,503$$

$$IDF(SistemInformasi) = \log\left(\frac{35}{1}\right) = 1,544$$

Berdasarkan hasil perhitungan TF dan IDF, diperoleh bobot TF-IDF setiap term sebagai berikut.

$$TF-IDF(Teknologi) = 0,143 \times 0,544 = 0,078$$

$$TF-IDF(Matematika) = 0,143 \times 0,314 = 0,045$$

$$TF-IDF(88) = 0,143 \times 0,314 = 0,072$$

$$TF-IDF(Informatika) = 0,143 \times 0,845 = 0,121$$

$$TF-IDF(87) = 0,143 \times 0,503 = 0,072$$

$$TF-IDF(CukupTinggi) = 0,143 \times 0,503 = 0,072$$

$$TF-IDF(SistemInformasi) = 0,143 \times 1,544 = 0,221$$

Tabel 4. Contoh Hasil Pembobotan TF-IDF

Term	TF	DF	IDF	TF-IDF
Teknologi	0,143	10	0,544	0,078
Matematika	0,143	17	0,314	0,045
88	0,143	11	0,503	0,072
Informatika	0,143	5	0,845	0,121
87	0,143	11	0,503	0,072
CukupTinggi	0,143	11	0,503	0,072
SistemInformasi	0,143	1	1,544	0,221

Berdasarkan Tabel 4, *term* SistemInformasi memiliki bobot TF-IDF tertinggi sebesar 0,221 karena hanya muncul pada satu dokumen sehingga menghasilkan nilai IDF yang paling besar. Sebaliknya, *term* Matematika memiliki bobot TF-IDF paling rendah sebesar 0,045 karena muncul pada lebih banyak dokumen. Hasil pembobotan ini menunjukkan bahwa metode TF-IDF mampu memberikan bobot yang lebih tinggi pada *term* yang lebih spesifik sehingga dapat merepresentasikan karakteristik setiap profil alumni dengan lebih baik. Bobot TF-IDF tersebut selanjutnya digunakan sebagai dasar dalam proses pengukuran tingkat kemiripan menggunakan metode *Cosine Similarity*.

D. Perhitungan Cosine Similarity

Setelah diperoleh bobot TF-IDF dari setiap profil alumni, tahap selanjutnya adalah menghitung tingkat kemiripan antara profil siswa dengan seluruh profil alumni menggunakan metode *Cosine Similarity*. Perhitungan ini bertujuan untuk mengetahui alumni yang memiliki karakteristik paling mirip dengan profil siswa sehingga dapat digunakan sebagai dasar dalam memberikan rekomendasi program studi.

Tabel 5. Data Input Siswa

Atribut	Nilai
Bidang Minat	Teknologi
Mata Pelajaran 1	Matematika
Nilai Mapel 1	88
Mata Pelajaran 2	Informatika
Nilai Mapel 2	87
Rata-rata Rapor	86,75

Berdasarkan Tabel 5, sistem membentuk profil siswa sebagai “Teknologi Matematika 88 Informatika 87 CukupTinggi”. Setelah itu

dilakukan perhitungan *Cosine Similarity* dengan rumus sebagai berikut:

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{|A||B|}$$

Keterangan:

A merupakan representasi vektor dari masukan pengguna. B merupakan representasi vektor dari data kursus. $|A|$ dan $|B|$ masing-masing menyatakan panjang vektor A dan B.

Berdasarkan hasil perhitungan menggunakan metode *Cosine Similarity*, diperoleh nilai kemiripan antara profil siswa dan seluruh profil alumni. Hasil perhitungan menunjukkan bahwa profil alumni A002 memperoleh nilai *similarity* sebesar 1,000, yang menunjukkan bahwa seluruh atribut profil siswa identik dengan profil alumni tersebut. Sementara itu, alumni A014 memperoleh nilai *similarity* sebesar 0,820, dan alumni A026 memperoleh nilai *similarity* sebesar 0,708. Nilai *similarity* yang semakin mendekati 1 menunjukkan tingkat kemiripan yang semakin tinggi antara profil siswa dan profil alumni.

Tabel 6. Tabel Hasil Similarity

ID Alumni	Profil	Similarity
A002	Sistem Informasi	1,000
A012	Teknik Informatika	0,820
A026	Ilmu Komputer	0,708
A005	Teknik Industri	0,601
A006	Teknik Sipil	0,581

Pada Tabel 6 dapat dilihat bahwa melalui implementasi metode *Cosine Similarity*, diperoleh matriks kedekatan antara profil siswa terhadap seluruh profil alumni yang tersedia. Hasil pengujian pada Tabel 6 memperlihatkan bahwa profil alumni A002 merepresentasikan kemiripan tertinggi dengan nilai *similarity* 1,000 yang berarti seluruh atributnya selaras dengan profil siswa. Sementara itu, nilai koefisien untuk alumni A012 dan A026 masing-masing menyusul di angka 0,820 dan 0,708. Secara umum, semakin besar nilai bobot kemiripan mendekati 1, maka semakin relevan pula kecocokan antara profil siswa dengan alumni tersebut.

E. Hasil Rekomendasi

Hasil Rekomendasi dari proses ini didapat dengan mengurutkan program studi berdasarkan nilai kemiripan atau *similarity score* tertinggi. Setelah nilai *Cosine Similarity* antara profil siswa dan seluruh profil alumni diperoleh, sistem mengurutkan nilai *similarity* dari yang tertinggi hingga terendah. Alumni dengan nilai *similarity* tertinggi dianggap memiliki karakteristik yang paling mendekati profil siswa. Program studi dari alumni tersebut kemudian dijadikan sebagai rekomendasi bagi siswa.

Tabel 7. Tabel Contoh Hasil Rekomendasi

ID Alumni	Profil	Similarity	Ranking
A002	Sistem Informasi	1,000	1
A012	Teknik Informatika	0,820	2
A026	Ilmu Komputer	0,708	3
A005	Teknik Industri	0,601	4
A006	Teknik Sipil	0,581	5

Berdasarkan data pada Tabel 7, dapat disimpulkan bahwa sistem mampu memberikan rekomendasi program studi berdasarkan tingkat kemiripan antara profil siswa dan profil alumni. Tahap *preprocessing* digunakan untuk membentuk profil alumni dan profil siswa dari atribut bidang minat, mata pelajaran pendukung, nilai mata pelajaran, serta kategori rata-rata rapor. Selanjutnya dilakukan pembobotan dengan memakai metode TF-IDF untuk menentukan tingkat kepentingan setiap *term* pada masing-masing profil.

Nilai *similarity* yang semakin tinggi menunjukkan bahwa karakteristik siswa semakin mendekati karakteristik alumni pada program studi tersebut. Oleh karena itu, urutan rekomendasi disusun berdasarkan nilai *similarity* tertinggi sehingga rekomendasi yang diberikan lebih relevan dengan minat dan kemampuan akademik siswa. Nilai *similarity* tertinggi diperoleh oleh alumni A002 karena seluruh atribut yang dimiliki identik dengan profil siswa, yaitu memiliki bidang minat Teknologi, mata pelajaran Matematika dan Informatika, nilai masing-masing 88 dan 87, serta kategori rata-rata rapor Cukup Tinggi. Oleh karena itu sistem merekomendasikan Program Studi Sistem Informasi sebagai rekomendasi utama.

F. Evaluasi Sistem dengan *Precision@3*

Setelah sistem menghasilkan rekomendasi program studi berdasarkan nilai *Cosine Similarity*, tahap selanjutnya adalah melakukan evaluasi terhadap hasil rekomendasi. Evaluasi dilakukan menggunakan metode *Precision@3* untuk mengukur ketepatan tiga rekomendasi program studi yang diberikan sistem. *Precision@3* digunakan karena sistem dirancang untuk menampilkan tiga rekomendasi program studi dengan nilai *similarity* tertinggi.

$$Precision@3 = \frac{3}{3} = 1$$

Berdasarkan hasil evaluasi, diperoleh nilai *Precision@3* sebesar 1,000 atau 100%. Hasil tersebut menunjukkan bahwa seluruh rekomendasi yang diberikan sistem pada data uji termasuk dalam kategori relevan terhadap karakteristik profil siswa. Dengan demikian, metode *Content-Based Filtering* yang dikombinasikan dengan pembobotan TF-IDF dan pengukuran *Cosine Similarity* mampu menghasilkan rekomendasi program studi yang sesuai berdasarkan data alumni yang digunakan.

Meskipun hasil pengujian menunjukkan nilai *Precision@3* yang sangat baik pada data uji ini, performa sistem tetap dipengaruhi oleh kualitas dan keberagaman data alumni yang digunakan. Semakin banyak dan semakin beragam data alumni yang tersedia, maka rekomendasi yang dihasilkan berpotensi menjadi lebih representatif dan akurat untuk berbagai karakteristik siswa.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, metode *Content-Based Filtering* dapat digunakan untuk menghasilkan rekomendasi program studi berdasarkan data alumni SMA Negeri 1 Nguter tahun 2025. Data yang digunakan meliputi bidang minat, program studi alumni, nilai mata pelajaran pendukung, dan rata-rata rapor. Proses rekomendasi dilakukan dengan membentuk profil alumni dan profil siswa menjadi fitur teks, kemudian diproses menggunakan TF-IDF dan dihitung tingkat kemiripannya menggunakan *Cosine Similarity*. Sistem menghasilkan tiga rekomendasi program

studi dengan nilai kemiripan tertinggi. Berdasarkan contoh pengujian pada bidang minat Teknologi, rekomendasi yang dihasilkan memperoleh nilai *Precision@3* sebesar 100%, sehingga sistem dinilai mampu memberikan rekomendasi yang relevan dengan bidang minat dan karakteristik akademik siswa. Saran untuk penelitian selanjutnya adalah menambahkan jumlah data alumni, menggunakan data prestasi siswa, serta mengembangkan sistem dalam bentuk aplikasi web agar dapat digunakan secara langsung oleh siswa dan guru Bimbingan Konseling.

UCAPAN TERIMA KASIH

Penelitian ini dapat terlaksana atas dukungan dan bantuan dari berbagai pihak. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada SMA Negeri 1 Nguter atas kerja sama, dukungan, serta izin yang diberikan dalam pengambilan data profil minat dan rekam jejak akademik siswa. Penulis juga mengucapkan terima kasih kepada Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta yang telah memfasilitasi, membimbing, dan memberikan dukungan penuh kepada penulis hingga publikasi ilmiah ini dapat diselesaikan dengan baik.

REFERENSI

- [1] A. Hasna, "Permasalahan yang di Hadapi Mahasiswa Salah Jurusan dari Sudut Pandang Psikologi," *Bandung Conference Series: Psychology Science*, vol. 4, no. 1, 2024.
- [2] M. Aucancela, A. Gonzalez Briones, and P. Chamoso, "Educational Recommender Systems: A Systematic Literature Review," in *The Barcelona Conference on Education 2023: Official Conference Proceedings*, 2023, pp. 933-951.
- [3] A. A. Huda, R. Fajarudin, and 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, pp. 1679-1686, 2022.
- [4] 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.
- [5] A. A. Zuhri, R. Kusumawati, M. A. Yaqin, A. F. Anwar, and A. F. A. Pahlevi, "Development of Academic Community Recommendation System Using Content-Based Filtering at UIN Malang Informatics Engineering Study Program," *J-INTECH: Journal of Information and Technology*, vol. 13, no. 1, pp. 157-166, 2025.
- [6] S. W. A. Wahyu and A. S. Purnomo, "Penerapan Content Based Filtering dalam Sistem Rekomendasi Pemilihan Jurusan Sekolah Menengah Kejuruan Berdasarkan Faktor Minat Bakat dan Nilai Akademik Siswa: Studi Kasus Sleman, Yogyakarta," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, pp. 1362-1369, 2025.

- [7] S. Oyadila, D. Abdullah, and A. Razi, "Implementasi Content-Based Filtering dengan TF-IDF dan Cosine Similarity untuk Sistem Rekomendasi Destinasi Wisata di Aceh Tengah," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1329-1339, 2025.
- [8] R. Priskila, N. N. K. Sari, and P. B. A. A. Putra, "Implementasi Content-Based Filtering Menggunakan TF-IDF and Cosine Similarity untuk Sistem Rekomendasi Resep Masakan," *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 18, no. 1, pp. 43-51, 2024.
- [9] J. J. Sihombing, Arnita, S. I. Al Idrus, and D. Y. Niska, "Implementation of Text Summarization on Indonesian Scientific Articles Using TextRank Algorithm with TF-IDF Web-Based," *Journal of Soft Computing Exploration*, vol. 5, no. 3, pp. 310-319, 2024.
- [10] I. A. W. Nandita, S. A. Putri, M. R. Kautsar, and R. Rahmadani, "News Recommendation System Using Content-Based Filtering through RSS Customization Service," *Journal of Applied Informatics and Computing*, 2025.
- [11] E. Anugrayningtyas, "Penggunaan Metode Content Based Filtering dalam Sistem Rekomendasi Elektronik di Toko Pareto," *Jurnal Fasilkom*, 2024.
- [12] S. I. Adam and W. G. Mokodaser, "Implementasi Sistem Rekomendasi Produk E-Commerce Menggunakan Content-Based Filtering Berbasis Cosine Similarity," *Simtek: Jurnal Sistem Informasi dan Teknik Komputer*, vol. 10, no. 2, pp. 418-424, 2025.
- [13] N. O. Idris and F. Pontooyo, "Sistem Rekomendasi Produk Makeup Berbasis Content-Based Filtering dengan TF-IDF dan Cosine Similarity," *KETIK: Jurnal Informatika*, 2025.