

ARTICLE

Pengembangan Sistem Rekomendasi Pembimbing Tugas Akhir Menggunakan Teknik Content Based Filtering

Development of a Final Project Supervisor Recommendation System Using Content-Based Filtering Technique

Femi Dwi Astuti¹ dan Widyastuti Andriyani^{*,2}

¹Informatika, Fakultas Teknologi Informasi, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia

²Magister Teknologi Informasi, Fakultas Teknologi Informasi, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia

*Penulis Korespondensi: widya@utdi.ac.id

(Disubmit 13-12-24; Diterima 09-05-25; Dipublikasikan online pada 20-06-25)

Abstrak

Pemilihan dosen pembimbing untuk tugas akhir merupakan tahapan penting bagi mahasiswa di jenjang pendidikan tinggi. Proses ini membutuhkan kecocokan antara bidang keahlian dosen dan topik penelitian mahasiswa, dengan tetap memperhatikan kendala seperti kapasitas pembimbing yang tersedia. Meninjau profil dosen dan rekam jejak penelitian secara manual dapat menjadi proses yang lambat dan kurang efisien. Untuk mengatasi tantangan tersebut, penelitian ini merancang sebuah sistem rekomendasi pembimbing tugas akhir dengan pendekatan Content-Based Filtering. Sistem ini memanfaatkan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk menilai relevansi istilah dalam abstrak penelitian, serta Cosine Similarity untuk mengukur kedekatan antara topik mahasiswa dan bidang penelitian dosen. Berdasarkan hasil pengujian, sistem mampu memberikan rekomendasi dosen pembimbing secara akurat berdasarkan kesamaan tertinggi dari judul proyek akhir yang diajukan, sehingga mempercepat dan meningkatkan ketepatan proses pemilihan. Sistem ini menjadi solusi praktis bagi mahasiswa dalam menentukan pembimbing yang sesuai dengan kebutuhan akademiknya, serta berpotensi meningkatkan kualitas proses bimbingan. Ke depan, penelitian lebih lanjut dapat mengembangkan integrasi dengan pendekatan penyaringan kolaboratif atau model rekomendasi hibrida guna menyempurnakan hasil rekomendasi yang dihasilkan.

Kata kunci: Penyaringan Berbasis Publikasi Karya Ilmiah; Sistem Rekomendasi; TF-IDF; Cosine Similarity; Pembimbing Proyek Akhir

Abstract

The selection of a supervisor for a final project is an important stage for students in higher education. This process requires a match between the lecturer's field of expertise and the student's research topic, while still considering constraints such as the available supervisor capacity. Manually reviewing lecturer profiles and research track records can be a slow and inefficient process. To overcome these challenges, this study designed a final project supervisor recommendation system with a Content-Based Filtering approach. This system utilizes the Term Frequency-Inverse Document Frequency (TF-IDF) method to assess the relevance of terms in research abstracts, as well as Cosine Similarity to measure the closeness between the student's topic and the lecturer's research field. Based on the test results, the system is able to provide accurate supervisor recommendations based on the highest similarity of the submitted final project title, thereby accelerating and increasing the accuracy of the selection process. This system is a practical solution for students in determining supervisors who are in accordance with their academic needs, and has the potential to improve the quality of the guidance process. In the future, further research can develop integration with a collaborative filtering approach or a

hybrid recommendation model to improve the recommendation results produced.

KeyWords: Publication Based Filtering of Scientific Papers; Recommendation System; TF-IDF; Cosine Similarity; Final Project Supervisor

1. Pendahuluan

Proyek akhir merupakan elemen esensial dalam pendidikan tinggi dan sering menjadi syarat utama untuk memperoleh gelar sarjana, magister, atau doktor. Proyek ini mencerminkan kemampuan mahasiswa dalam melakukan penelitian secara mandiri, mengolah data, serta menyampaikan hasil secara sistematis dan logis. Keberhasilan dalam pelaksanaan proyek akhir sangat bergantung pada kualitas bimbingan dan kompetensi akademik dosen pembimbing, yang memainkan peran krusial dalam menentukan arah serta mutu penelitian [1, 2]. Namun, memilih pembimbing yang sesuai sering menjadi tantangan tersendiri bagi mahasiswa. Di banyak perguruan tinggi, terdapat sejumlah dosen dengan keahlian yang saling tumpang tindih, sehingga mahasiswa kesulitan dalam menentukan pembimbing yang paling relevan dengan topik penelitian mereka. Selain itu, institusi pendidikan biasanya membatasi jumlah mahasiswa yang dapat dibimbing oleh satu dosen dalam satu periode akademik, yang membuat rumit proses pemilihan pembimbing [3, 4].

Saat ini, mahasiswa umumnya harus meninjau secara manual profil dan publikasi dosen yang tersedia, sebuah proses yang tidak hanya memakan waktu tetapi juga kurang efisien, terutama di institusi berskala besar. Untuk menjawab tantangan tersebut, penelitian ini mengembangkan sistem rekomendasi pembimbing akademik berbasis pendekatan *Content-Based Filtering* (CBF) [5, 6]. Sistem rekomendasi bagian dari kecerdasan buatan dirancang untuk memberikan saran yang dipersonalisasi dengan menganalisis sejumlah parameter tertentu. Dalam hal ini, metode CBF merekomendasikan berdasarkan karakteristik internal dari objek serta tingkat kesesuaiannya dengan preferensi pengguna. Sistem yang dibangun memanfaatkan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk analisis teks dan *Cosine Similarity* untuk mengukur tingkat kesamaan antara topik yang diajukan oleh mahasiswa dan profil penelitian para dosen [7, 8, 9].

Melalui penerapan sistem ini, mahasiswa dapat memperoleh saran pembimbing yang lebih akurat dan efisien, sehingga proses pemilihan menjadi lebih cepat dan tepat dibandingkan dengan cara manual. Artikel ini memaparkan proses pengembangan serta evaluasi sistem yang menggunakan data publikasi sebagai basisnya. Penelitian ini bertujuan memberikan kontribusi pada pengembangan sistem rekomendasi berbasis kecerdasan buatan di lingkungan akademik, sekaligus menawarkan solusi praktis untuk meningkatkan efektivitas proses bimbingan tugas akhir. Ke depan, integrasi metode lain seperti *collaborative filtering* juga dapat diujicoba guna menyempurnakan performa dan akurasi sistem rekomendasi.

2. Metode

2.1 TF-IDF

Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik dalam *text mining* dan *information retrieval* yang digunakan untuk mengevaluasi seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap koleksi dokumen lainnya. Metode ini bekerja dengan menggabungkan dua ukuran utama: frekuensi istilah (TF) dan kebalikan frekuensi dokumen (IDF) [10, 11]. Frekuensi istilah (TF) mengukur seberapa sering suatu kata muncul dalam sebuah dokumen, dengan asumsi bahwa kata yang sering muncul lebih penting bagi dokumen tersebut. TF saja tidak cukup untuk menilai kepentingan kata karena beberapa kata umum, seperti "dan", "atau", dan "adalah", mungkin sering muncul di seluruh dokumen, tetapi kurang bermakna. Oleh karena itu, IDF digunakan untuk mengurangi bobot kata-kata yang sering muncul di berbagai dokumen dengan menghitung logaritma rasio total dokumen terhadap jumlah dokumen yang mengandung kata tersebut. Kombinasi TF dan IDF memberikan bobot yang lebih seimbang, di mana kata-kata yang sangat relevan dengan dokumen tertentu tetapi jarang muncul di dokumen lain akan mendapatkan bobot yang lebih tinggi. Metode TF-IDF banyak digunakan dalam aplikasi seperti analisis teks, sistem pencarian informasi, dan *algoritme machine learning* untuk memproses data berbasis teks, karena efisiensinya dalam menangkap relevansi kata-kata terhadap dokumen dan keseluruhan koleksi teks.

2.2 Cosine Similarity

Cosine Similarity adalah metode dalam analisis data teks dan pengolahan vektor untuk mengukur kesamaan antara dua vektor, yang sering digunakan untuk membandingkan dokumen dalam ruang vektor. Metode ini menghitung sudut kosinus antara dua vektor di ruang berdimensi tinggi, yang merepresentasikan dokumen atau teks berdasarkan fitur seperti bobot TF-IDF dari setiap kata dalam teks tersebut [12].

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

Dimana

- A dan B : vektor teks atau dokumen
 $A \cdot B$: perkalian dot product antara dua vektor.
 $\|A\|$ dan $\|B\|$: panjang (norma) masing-masing vektor.

Nilai hasilnya berkisar antara -1 hingga 1, di mana:

- 1 berarti kedua vektor memiliki orientasi yang sama atau identik.
 0 berarti kedua vektor ortogonal (tidak memiliki kemiripan).
 -1 menunjukkan kedua vektor berlawanan arah (sangat tidak mirip).

Dalam konteks teks, dokumen yang lebih mirip secara semantik akan memiliki nilai *Cosine Similarity* yang lebih tinggi (mendekati 1). Metode ini sering digunakan dalam sistem pencarian informasi, analisis kluster, dan *recommendation systems* untuk mengukur relevansi antara dokumen, paragraf, atau kalimat. Keunggulan utama *Cosine Similarity* adalah kemampuannya menangani dokumen dengan panjang berbeda, karena metode ini hanya mempertimbangkan arah vektor, bukan magnitudonya.

2.3 Content-Based Filtering (CBF)

Content-Based Filtering (CBF) adalah teknik rekomendasi yang berfokus pada analisis atribut konten untuk memberikan rekomendasi yang relevan kepada pengguna. Dalam pendekatan ini, sistem merekomendasikan item berdasarkan kemiripan antara karakteristik item yang telah dinilai positif oleh pengguna sebelumnya dan item-item lainnya yang tersedia [12, 13, 14]. Prinsip dasar CBF adalah membangun profil pengguna dari preferensi mereka terhadap atribut spesifik item. Misalnya, dalam sistem rekomendasi film, atribut seperti genre, sutradara, atau aktor dapat digunakan untuk menentukan kesamaan antara film yang disukai pengguna dengan film lainnya [15].

$$\text{Similarity}(U, I_i) = \frac{\sum_{k=1}^n U_k \cdot I_{i,k}}{\sqrt{\sum_{k=1}^n U_k^2} \cdot \sqrt{\sum_{k=1}^n I_{i,k}^2}} \quad (2)$$

di mana:

- U_k : nilai fitur k pada profil pengguna
 $I_{i,k}$: nilai fitur k pada item i
 n : jumlah total fitur.

Langkah-Langkah CBF:

1. Ekstraksi Fitur: Atribut dari item (misalnya deskripsi, kategori, atau fitur spesifik) dikonversi menjadi representasi numerik, seperti vektor menggunakan metode seperti TF-IDF.
2. Pembuatan Profil Pengguna: Profil pengguna dibangun berdasarkan analisis item yang telah dinilai atau diakses sebelumnya. Profil ini merepresentasikan preferensi pengguna terhadap fitur atau atribut tertentu.
3. Penghitungan Kemiripan: Kemiripan antara item baru dan item dalam profil pengguna dihitung menggunakan metode seperti *Cosine Similarity* atau *Euclidean Distance*.
4. Pemberian Rekomendasi: Sistem merekomendasikan item dengan tingkat kemiripan tertinggi kepada pengguna.

(a) Kelebihan

- i. Personalisasi Tinggi: Rekomendasi didasarkan pada preferensi unik setiap pengguna.
- ii. Tidak Membutuhkan Data dari Pengguna Lain: Tidak memerlukan data pengguna lain (tidak terpengaruh oleh cold start pengguna baru).
- iii. Mudah Dipahami: Pendekatan yang intuitif karena langsung memanfaatkan atribut item.

(b) Kekurangan

- i. Cold Start untuk Item Baru: Tidak dapat merekomendasikan item yang belum memiliki deskripsi atau atribut jelas (masalah new item).
- ii. Kurang Variasi: Cenderung merekomendasikan item yang sangat mirip dengan item yang sudah dinilai, sehingga mengurangi eksplorasi (serendipity).
- iii. Keterbatasan Fitur: Bergantung pada ketersediaan dan kualitas atribut konten.

CBF banyak diterapkan dalam domain seperti rekomendasi buku, film, produk *e-commerce*, hingga layanan digital, dengan mengandalkan representasi konten yang baik untuk menghasilkan rekomendasi yang relevan dan personal seperti bobot TF-IDF dari setiap kata dalam teks tersebut.

2.4 Bidang Konsentrasi Dosen

Bidang konsentrasi dosen merupakan area keahlian atau fokus akademik yang mendalam, yang menjadi landasan bagi seorang dosen dalam menjalankan tugas tridharma perguruan tinggi, yaitu pengajaran, penelitian, dan pengabdian kepada masyarakat. Konsentrasi ini biasanya terkait dengan latar belakang pendidikan, pengalaman penelitian, serta minat pengembangan keilmuan dosen. Misalnya, dosen dengan konsentrasi dalam bidang kecerdasan buatan akan fokus pada pengajaran mata kuliah seperti pembelajaran mesin, sistem pakar, atau pengolahan data besar, sembari melakukan penelitian yang relevan di bidang tersebut. Konsentrasi ini tidak hanya memengaruhi kualitas pembelajaran, tetapi juga kontribusi dosen terhadap pengembangan ilmu pengetahuan, baik melalui publikasi ilmiah, penerapan teknologi, maupun kolaborasi dengan industri dan masyarakat. Pemilihan bidang konsentrasi juga berperan penting dalam menciptakan keunggulan akademik institusi, karena setiap dosen dapat menjadi spesialis dalam bidangnya masing-masing, yang pada akhirnya meningkatkan reputasi dan daya saing perguruan tinggi, terlihat pada Table 1.

a) Dataset

Dataset yang diperlukan:

- Data mahasiswa: Topik atau abstrak tugas akhir yang diusulkan. "Mahasiswa_ID": 1, "Topik": "Penerapan *Machine Learning* untuk Prediksi Cuaca"
- Data dosen: Bidang konsentrasi, riwayat penelitian, atau kata kunci dari publikasi dosen. "Dosen_ID": 101, "Konsentrasi": "*Machine Learning*, *Data Science*, Prediksi Cuaca"

b) Preprocessing Data

- *Tokenisasi*: Memecah teks menjadi kata-kata individu.
- *Stemming/Lemmatization*: Mengurangi kata ke bentuk dasarnya.
- *Stopword Removal*: Menghapus kata umum seperti "dan", "atau", "adalah". Vectorization (TF-IDF): Merepresentasikan teks dalam bentuk vektor numerik berdasarkan frekuensi istilah (TF) dan kebalikannya (IDF).

c) Perhitungan *Cosine Similarity*

Menggunakan formula *Cosine Similarity* untuk mengukur kemiripan antara vektor topik mahasiswa dan vektor bidang konsentrasi dosen:

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

Tabel 1. Bidang Konsentrasi Dosen

Bidang Konsentrasi	Deskripsi
Kecerdasan Buatan (<i>Artificial Intelligence</i>)	Pengembangan dan penerapan algoritma cerdas seperti pembelajaran mesin, <i>deep learning</i> , dan sistem pakar.
Sistem Informasi	Analisis, desain, implementasi, dan manajemen sistem informasi untuk mendukung pengambilan keputusan.
<i>Data Science</i> dan Analitik Data	Pengolahan, analisis, dan visualisasi data untuk mendapatkan wawasan dan mendukung pengambilan keputusan.
Rekayasa Perangkat Lunak (<i>Software Engineering</i>)	Metode dan alat untuk desain, pengembangan, pengujian, dan pemeliharaan perangkat lunak.
Keamanan Siber (<i>Cyber Security</i>)	Perlindungan data, jaringan, dan sistem komputer dari ancaman dan serangan siber.
Jaringan Komputer dan Komunikasi Data	Teknologi jaringan, protokol komunikasi, dan infrastruktur jaringan modern.
Pemrosesan Citra dan Komputer Grafik	Analisis gambar digital, pengolahan video, dan aplikasi seperti <i>augmented reality</i> dan visualisasi.
Komputasi Terdistribusi dan <i>Cloud Computing</i>	Sistem terdistribusi, komputasi berbasis <i>cloud</i> , dan teknologi virtualisasi.
<i>Internet of Things</i> (IoT)	Integrasi perangkat pintar yang terhubung dengan internet untuk solusi inovatif di berbagai bidang.
Sistem Tertanam (<i>Embedded Systems</i>)	Desain dan implementasi perangkat keras dan perangkat lunak untuk sistem tertanam seperti robotik.
Kecerdasan Komputasional (<i>Computational Intelligence</i>)	Algoritma berbasis inspirasi biologis seperti algoritma genetika, <i>swarm intelligence</i> , dan <i>fuzzy systems</i> .
<i>Human-Computer Interaction</i> (HCI)	Desain antarmuka pengguna dan pengalaman pengguna dalam berinteraksi dengan teknologi komputer.
Robotika dan Otomasi	Desain, pengendalian, dan pengoperasian robot untuk berbagai aplikasi.
Penambangan Data (<i>Data Mining</i>)	Teknik untuk mengekstrak pola atau pengetahuan dari kumpulan data yang besar.
Komputasi Kuantum (<i>Quantum Computing</i>)	Pengembangan algoritma komputasi yang memanfaatkan prinsip mekanika kuantum.
Komputasi Mobile dan Aplikasi Mobile	Pengembangan aplikasi dan sistem yang berjalan di perangkat mobile.
Teknologi <i>Blockchain</i>	Aplikasi berbasis blockchain untuk keamanan, keuangan, dan sistem terdesentralisasi lainnya.
Pengembangan Game dan Simulasi	Desain, pengembangan, dan implementasi game serta simulasi interaktif.

Dimana:

A : Vektor TF-IDF dari topik mahasiswa.

B : Vektor TF-IDF dari bidang konsentrasi dosen

Hasil *Cosine Similarity* akan memberikan nilai antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan kesamaan tinggi.

d) Langkah Implementasi

- Representasi Data dengan TF-IDF
 - (a) Gabungkan teks dari mahasiswa dan dosen menjadi corpus.
 - (b) Lakukan vectorization menggunakan metode TF-IDF.
 - (c) Simpan vektor mahasiswa (MMM) dan vektor dosen (DDD).
- Hitung Cosine Similarity

- (a) Untuk setiap mahasiswa, hitung nilai *Cosine Similarity* antara MMM dan setiap DDD.
 - (b) Urutkan hasil berdasarkan nilai kesamaan tertinggi.
- Rekomendasikan Dosen Pembimbing Rekomendasikan dosen dengan nilai *Cosine Similarity* tertinggi untuk setiap mahasiswa.

3. Hasil dan Pembahasan

Membangun sistem yang merekomendasikan pembimbing tugas akhir berdasarkan kesesuaian topik penelitian mahasiswa dengan bidang konsentrasi atau publikasi dosen, seperti terlihat pada Tabel 2.

Tabel 2. Dataset mahasiswa berdasarkan topik atau abstrak tugas akhir yang diusulkan

Mahasiswa_ID	Nama Mahasiswa	Topik/Abstrak Tugas Akhir
201	Mhs_1	"Penerapan Machine Learning untuk Prediksi Harga Saham: Studi Kasus Perusahaan Teknologi di Indonesia."
202	Mhs_2	"Sistem Informasi Manajemen untuk Pengelolaan Data Pegawai Berbasis Web pada Organisasi Non-Profit."
203	Mhs_3	"Analisis dan Pencegahan Serangan Siber menggunakan Teknik Kriptografi pada Sistem Keuangan Digital."
204	Mhs_4	"Pengembangan Sistem Internet of Things (IoT) untuk Monitoring Kualitas Udara di Perkotaan."
205	Mhs_5	"Desain dan Implementasi Aplikasi Mobile untuk Pemesanan Tiket Bioskop Berbasis Android."
206	Mhs_6	"Analisis Kinerja Jaringan Komputer dalam Implementasi Sistem Video Streaming dengan Protokol RTSP."
207	Mhs_7	"Peningkatan Akurasi Deteksi Objek pada Citra Satelit menggunakan Teknik Deep Learning."
208	Mhs_8	"Visualisasi Data Interaktif untuk Analisis Tren Penjualan di Pasar E-commerce."
209	Mhs_9	"Pengembangan Robot Penjelajah untuk Operasi Pencarian dan Penyelamatan dengan Algoritma Swarm Intelligence."
210	Mhs_10	"Implementasi Teknologi Blockchain untuk Pengelolaan Sertifikat Digital di Institusi Pendidikan."
211	Mhs_11	"Desain dan Pengembangan Game Edukasi untuk Anak Usia Dini Berbasis Android."
212	Mhs_12	"Penggunaan Komputasi Awan untuk Penyimpanan dan Analisis Data IoT pada Smart Farming."
213	Mhs_13	"Pengaruh Desain Antarmuka Pengguna pada Tingkat Kepuasan Pengguna dalam Aplikasi E-commerce."
214	Mhs_14	"Eksplorasi Algoritma Kuantum untuk Pemrosesan Data Besar dalam Analisis Genetik."
215	Mhs_15	"Implementasi Teknologi Blockchain untuk Sistem Voting yang Transparan dan Aman."

Data ini mencakup:

1. Mahasiswa_ID: Identifikasi unik untuk setiap mahasiswa.
2. Nama Mahasiswa: Nama lengkap mahasiswa.
3. Topik/Abstrak Tugas Akhir: Deskripsi singkat tentang topik atau abstrak tugas akhir yang diusulkan.

Data ini digunakan sebagai masukan untuk menghitung kesesuaian (*Cosine Similarity*) dengan dataset dosen, sehingga sistem rekomendasi pembimbing dapat memberikan hasil yang relevan, seperti terlihat pada Tabel 3.

Data ini mencakup informasi:

Tabel 3. rekomendasi pembimbing

Dosen_ID	Nama Dosen	Bidang Konsentrasi	Kata Kunci
101	Dsn_1	Kecerdasan Buatan, Pembelajaran Mesin	Artificial Intelligence, Machine Learning, Deep Learning
102	Dsn_2	Sistem Informasi, Manajemen Data	Information Systems, Database Management, Data Integration
103	Dsn_3	Keamanan Siber, Kriptografi, Forensik Digital	Cyber Security, Cryptography, Digital Forensics
104	Dsn_4	Internet of Things, Sistem Tertanam	IoT, Embedded Systems, Smart Devices
105	Dsn_5	Rekayasa Perangkat Lunak, Pengembangan Aplikasi	Software Engineering, Application Development
106	Dsn_6	Jaringan Komputer, Komunikasi Data	Computer Networks, Data Communication, Network Security
107	Dsn_7	Pemrosesan Citra, Komputer Grafik	Image Processing, Computer Graphics, Visual Computing
108	Dsn_8	Data Science, Analitik Data, Penambangan Data	Data Science, Data Analytics, Data Mining
109	Dsn_9	Robotika, Kecerdasan Komputasional	Robotics, Computational Intelligence, Swarm Intelligence
110	Dsn_10	Blockchain, Teknologi Keuangan Digital	Blockchain, FinTech, Decentralized Systems
111	Dsn_11	Pengembangan Game dan Simulasi	Game Development, Simulation, Interactive Systems
112	Dsn_12	Komputasi Terdistribusi, Komputasi Awan	Distributed Computing, Cloud Computing, Virtualization
113	Dsn_13	Human-Computer Interaction (HCI), Antarmuka Pengguna	HCI, User Interface Design, User Experience
114	Dsn_14	Komputasi Kuantum, Pengembangan Algoritma Kuantum	Quantum Computing, Quantum Algorithms, Quantum Systems
115	Dsn_15	Teknologi Blockchain, Keamanan Data Digital	Blockchain Technology, Digital Data Security, Smart Contracts

1. Dosen_ID: Identifikasi unik untuk setiap dosen.
2. Nama Dosen: Nama lengkap dosen.
3. Bidang Konsentrasi: Bidang fokus utama dosen dalam mengajar dan penelitian.
4. Kata Kunci: Daftar istilah kunci yang relevan dengan bidang konsentrasi untuk digunakan dalam analisis kemiripan (*Cosine Similarity*).

3.1 Tahap Tokenisasi

Tokenisasi alternatif untuk teks abstrak mahasiswa:

- a) **Penerapan Machine Learning untuk Prediksi Harga Saham** ['Penerapan', 'Machine', 'Learning', 'untuk', 'Prediksi', 'Harga', 'Saham']
- b) **Sistem Informasi Manajemen untuk Pengelolaan Data Pegawai** ['Sistem', 'Informasi', 'Manajemen', 'untuk', 'Pengelolaan', 'Data', 'Pegawai']
- c) **Analisis dan Pencegahan Serangan Siber menggunakan Teknik Kriptografi** ['Analisis', 'dan', 'Pencegahan', 'Serangan', 'Siber', 'menggunakan', 'Teknik', 'Kriptografi']
- d) **Pengembangan Sistem Internet of Things untuk Monitoring Kualitas Udara** ['Pengembangan', 'Sistem', 'Internet', 'of', 'Things', 'untuk', 'Monitoring', 'Kualitas', 'Udara']
- e) **Desain dan Implementasi Aplikasi Mobile untuk Pemesanan Tiket Bioskop** ['Desain', 'dan', 'Implementasi', 'Aplikasi', 'Mobile', 'untuk', 'Pemesanan', 'Tiket', 'Bioskop']
- f) **Analisis Kinerja Jaringan Komputer dalam Implementasi Sistem Video Streaming** ['Analisis', 'Kinerja', 'Jaringan', 'Komputer', 'dalam', 'Implementasi', 'Sistem', 'Video', 'Streaming']

Hasil ini memecah teks menjadi kata-kata individu untuk setiap abstrak tugas akhir yang diusulkan. Tokenisasi ini dapat digunakan dalam langkah selanjutnya seperti TF-IDF atau Cosine Similarity

3.2 Proses penghapusan stopwords dari teks tokenisasi

- Penerapan Machine Learning untuk Prediksi Harga Saham** ['Penerapan', 'Machine', 'Learning', 'Prediksi', 'Harga', 'Saham']
- Sistem Informasi Manajemen untuk Pengelolaan Data Pegawai** ['Sistem', 'Informasi', 'Manajemen', 'Pengelolaan', 'Data', 'Pegawai']
- Analisis dan Pencegahan Serangan Siber menggunakan Teknik Kriptografi** ['Analisis', 'Pencegahan', 'Serangan', 'Siber', 'menggunakan', 'Teknik', 'Kriptografi']
- Pengembangan Sistem Internet of Things untuk Monitoring Kualitas Udara** ['Pengembangan', 'Sistem', 'Internet', 'of', 'Things', 'Monitoring', 'Kualitas', 'Udara']
- Desain dan Implementasi Aplikasi Mobile untuk Pemesanan Tiket Bioskop** ['Desain', 'Implementasi', 'Aplikasi', 'Mobile', 'Pemesanan', 'Tiket', 'Bioskop']
- Analisis Kinerja Jaringan Komputer dalam Implementasi Sistem Video Streaming** ['Analisis', 'Kinerja', 'Jaringan', 'Komputer', 'Implementasi', 'Sistem', 'Video', 'Streaming']

Stopwords seperti "dan," "untuk," dan "dalam" telah dihapus untuk mengurangi kata-kata tidak penting dalam analisis teks. Proses ini membantu meningkatkan akurasi dalam analisis berbasis teks seperti TF-IDF atau *Cosine Similarity*.

3.3 Vectorization (TF-IDF)

Tabel 4. *Cosine Similarity* Topik Mahasiswa

	Mhs_1	Mhs_2	Mhs_3	Mhs_4	Mhs_5	Mhs_6
Mhs_1	1	0	0	0	0	0
Mhs_2	0	1	0	0.074	0	0.078
Mhs_3	0	0	1	0	0	0.099
Mhs_4	0	0.074	0	1	0	0.067
Mhs_5	0	0	0	0	1	0.099
Mhs_6	0	0.078	0.099	0.067	0.099	1

Hasil perhitungan *Cosine Similarity* antar topik mahasiswa telah disajikan dalam bentuk tabel seperti terlihat pada Tabel 4. Setiap nilai menunjukkan tingkat kemiripan antara pasangan topik mahasiswa berdasarkan bobot TF-IDF. Nilai mendekati 1.0 menunjukkan kemiripan yang tinggi, sedangkan nilai mendekati 0.0 menunjukkan tidak ada kemiripan.

Tabel 5. Rekomendasi Dosen Pembimbing Berdasarkan *Cosine Similarity*

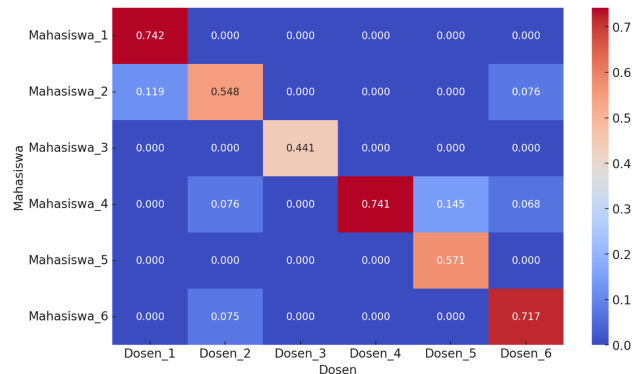
No	Dsn_1	Dsn_2	Dsn_3	Dsn_4	Dsn_5	Dsn_6
Mhs_1	0.742	0	0	0	0	0
Mhs_2	0.119	0.548	0	0	0	0.076
Mhs_3	0	0	0.441	0	0	0
Mhs_4	0	0.076	0	0.741	0.144	0.068
Mhs_5	0	0	0	0	0.571	0
Mhs_6	0	0.075	0	0	0	0.717

Rekomendasi dosen pembimbing berdasarkan nilai *Cosine Similarity* tertinggi:

- Mhs_1 → Dsn_ID_101
- Mhs_2 → Dsn_ID_102
- Mhs_3 → Dsn_ID_103
- Mhs_4 → Dsn_ID_104
- Mhs_5 → Dsn_ID_105

6. Mhs_6 → Dsn_ID_106

Hasil ini memberikan pasangan mahasiswa dan dosen dengan bidang konsentrasi yang paling relevan, sesuai dengan kemiripan topik penelitian tugas akhir mahasiswa seperti terlihat pada Gambar 1.



Gambar 1. Rekomendasi Dosen Pembimbing Berdasarkan *Cosine Similarity*

- Mahasiswa_1 memiliki nilai *Cosine Similarity* tertinggi dengan Dosen_1 (nilai 0.742). Ini menunjukkan bahwa Dosen_1 adalah pembimbing yang paling sesuai untuk Mahasiswa_1.
- Mahasiswa_4 memiliki nilai kesesuaian tertinggi dengan Dosen_4 (nilai 0.741). Ini menandakan kecocokan yang baik di antara keduanya.
- Mahasiswa_2 memiliki kesesuaian tertinggi dengan Dosen_2 (nilai 0.548), tetapi juga memiliki kesesuaian moderat dengan Dosen_6 (nilai 0.076), meskipun lebih rendah.

Sehingga mengidentifikasi dosen yang paling sesuai untuk setiap mahasiswa berdasarkan nilai *Cosine Similarity* tertinggi. Dengan visualisasi ini, pengambilan keputusan dapat dilakukan dengan lebih mudah dan efisien, karena pasangan dengan kesesuaian terbaik terlihat jelas melalui warna dan angka.

4. Simpulan

Penelitian ini berhasil merancang sistem rekomendasi pembimbing tugas akhir dengan pendekatan *Content-Based Filtering* (CBF) yang mengimplementasikan metode TF-IDF dan *Cosine Similarity*. Berdasarkan hasil pengujian, sistem mampu memberikan rekomendasi pembimbing secara akurat dengan mengidentifikasi kesamaan tertinggi, sehingga meningkatkan efektivitas dan ketepatan dalam proses pemilihan dosen pembimbing. Temuan ini memberikan kontribusi nyata dalam mendukung manajemen bimbingan akademik serta memperkuat proses pengambilan keputusan di lingkungan perguruan tinggi.

Untuk meningkatkan ketepatan dan fleksibilitas sistem rekomendasi yang telah dirancang, penelitian di masa depan dapat mempertimbangkan penggabungan metode penyaringan kolaboratif atau pendekatan hibrida yang memanfaatkan atribut dari pengguna dan item secara bersamaan. Selain itu, memperkaya dataset dengan informasi terkait performa bimbingan dosen dan umpan balik dari mahasiswa akan memperdalam analisis kecocokan. Pengembangan selanjutnya juga dapat mencakup penambahan fitur, seperti memperhitungkan beban kerja dosen, preferensi mahasiswa, serta jadwal bimbingan, guna meningkatkan relevansi hasil rekomendasi. Dari sisi teknologi, penerapan teknik *deep learning* atau pemrosesan bahasa alami (NLP) yang lebih mutakhir dapat digunakan untuk menganalisis abstrak tugas akhir maupun profil dosen secara lebih mendalam. Terakhir, sistem ini sebaiknya diuji pada cakupan yang lebih luas, seperti di institusi lain atau menggunakan dataset yang lebih besar, untuk memastikan kemampuan generalisasi dan keandalan model dalam berbagai kondisi akademik. Dengan upaya-upaya ini, sistem rekomendasi berpeluang menjadi alat yang lebih optimal dan berkontribusi nyata dalam mendukung proses akademik di perguruan tinggi.

Pustaka

- [1] C. M. Evans, "Student outcomes from high-quality project-based learning: A case study for PBLWorks," Cent. Assessment, Dover, NH, 2019, [Online]. Available: https://www.pblworks.org/sites/default/files/2020-01/PBLWorks_HQPBL_Teacher_Case_Study_Report_FINAL.pdf
- [2] M. Todd, P. Bannister, and S. Clegg, "Independent inquiry and the undergraduate dissertation: Perceptions and experiences of final-year social science students," *Assess. Eval. High. Educ.*, vol. 29, no. 3, pp. 335–355, 2004, doi: 10.1080/0260293042000188285.
- [3] LEONARD A. JASON and D. S. GLENWICK, *METHODOLOGICAL APPROACHES TO COMMUNITY-BASED RESEARCH*. 2016.
- [4] D. Van Damme and D. Zahner, Does Higher Education Teach Students to Think Critically? 2022. doi: 10.1787/cc9fa6aa-en.
- [5] R. Shatnawi, Q. Althebyan, B. Ghaleb, and M. Al-Maolegi, "A student advising system using association rule mining," *Int. J. Web-Based Learn. Teach. Technol.*, vol. 16, no. 3, pp. 65–78, 2021, doi: 10.4018/IJWLTT.20210501.0a5.
- [6] R. Rismanto, A. R. Syulistyo, and B. P. C. Agusta, "Research supervisor recommendation system based on topic conformity," *Int. J. Mod. Educ. Comput. Sci.*, vol. 12, no. 1, pp. 26–34, 2020, doi: 10.5815/ijme-cs.2020.01.04.
- [7] A. H. J. P. Juni Permana and Agung Toto Wibowo, "Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity," *Int. J. Inf. Commun. Technol.*, vol. 9, no. 2, pp. 1–14, 2023, doi: 10.21108/ijoict.v9i2.747.
- [8] S. Zhang, *Restaurant Recommendation System Based on TF-IDF Vectorization: Integrating Content-Based and Collaborative Filtering Approaches*, no. Dai 2023. Atlantis Press International BV, 2024. doi: 10.2991/978-94-6463-370-2_62.
- [9] J. Ni, Y. Cai, G. Tang, and Y. Xie, "Collaborative filtering recommendation algorithm based on TF-IDF and user characteristics," *Appl. Sci.*, vol. 11, no. 20, 2021, doi: 10.3390/app11209554.
- [10] W. Zhang, T. Yoshida, and X. Tang, "A comparative study of TF*IDF, LSI and multi-words for text classification," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765, 2011, doi: 10.1016/j.eswa.2010.08.066.
- [11] R. K. Dey and A. K. Das, "Modified term frequency-inverse document frequency based deep hybrid framework for sentiment analysis," *Multimed. Tools Appl.*, vol. 82, no. 21, pp. 32967–32990, 2023, doi: 10.1007/s11042-023-14653-1.
- [12] R. Saptono, H. Prasetyo, and A. Irawan, "Combination of cosine similarity method and conditional probability for plagiarism detection in the thesis documents vector space model," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 2–4, pp. 139–143, 2018.
- [13] M. T. Sharad Phalle and P. S. Bhushan, "Content Based Filtering And Collaborative Filtering: A Comparative Study," *J. Adv. Zool.*, vol. 45, pp. 96–100, 2024, doi: 10.53555/jaz.v45is4.4158.
- [14] C. Channarong, C. Paosirikul, S. Maneeroj and A. Takasu, "HybridBERT4Rec: A Hybrid (Content-Based Filtering and Collaborative Filtering) Recommender System Based on BERT," in *IEEE Access*, vol. 10, pp. 56193–56206, 2022, doi: 10.1109/ACCESS.2022.3177610.
- [15] C. T. Madongo and T. Zhongjun, "A movie box office revenue prediction model based on deep multi-modal features," *Multimedia Tools and Applications*, vol. 82, no. 21, pp. 31981–32009, Sep. 2023, doi: 10.1007/s11042-023-14456-4.